Leveraging Social Media and Machine Learning for enhanced Communication and Understanding between Organizations and Stakeholders

By

Piyumini Rasangika Wijenayake

Bachelor of Science Special degree in Computer Science (Honours)

A thesis submitted in total fulfilment

of the requirements for the degree of

Doctor of Philosophy

Research Centre for Data Analytics and Cognition

La Trobe Business School

College of Arts, Social Sciences and Commerce

La Trobe University

Victoria, Australia

June 2021

To wonderful Eshali

Acknowledgements

It would not have been possible to successfully complete the research presented in this thesis without the generous support and guidance from various individuals and organizations.

First and foremost, I wish to thank the funding bodies. This work was supported by an Australian Government Research Training Program Scholarship and a La Trobe University Research Studentship Scholarship sponsored by the La Trobe ICT Unit.

I am extremely grateful to my esteemed principal supervisor, Prof. Damminda Alahakoon for his continuous and unreserved support, guidance and inspiration extended throughout my PhD studies, without which this thesis would not have been possible. Thank you for bestowing your unparallel expertise, creativity and tutelage with impeccable grace and patience. I would also like to extend my heartfelt gratitude to my co-supervisor, Prof. Daswin de Silva for his continuous guidance, support and encouragement to strive for better outcomes.

My gratitude extends to all my colleagues and collaborators including all of the PhD students and staff at the Research Centre for Data Analytics and Cognition (CDAC), the La Trobe Business School, the La Trobe Business Intelligence unit and my international research collaborators, for the varied support and encouragement extended and for their insightful comments and feedback provided throughout my candidature.

I am indebted to my parents who instilled in me the value of science. I wish to thank my mother and my daughter for their continuous support and understanding, and my family and friends for their unwavering support and belief in me.

Last but not least, I offer merits to everyone who has supported, nurtured and enriched my life in numerous ways throughout my journey for me to be here today at this milestone.

Finally, I acknowledge and pay homage to knowledge and wisdom that has been handed down through the generations.

Table of Contents

List of Figures

List of Tables

Abstract

In the modern digital environment with the rise of adoption, diversification and specialization of social media, social media platforms have become a real-time, high velocity data bank of human opinions, emotions and experiences. Accordingly, organizations are increasingly depending on social media management techniques to analyse social media content to understand and serve stakeholders better and obtain a competitive advantage. However, social media data have not yet been fully investigated in leveraging for organizations to understand their stakeholders on a deeper level.

This thesis explores novel opportunities and techniques presented through leveraging social media, to capture, analyse and understand human communication and interactions from an organizational perspective.

The research is carried out using digital traces of human emotions, opinions and experiences available in social media. Machine learning techniques such as deep learning, natural language processing and word embedding are utilized together with a widely accepted social media framework. Established theories from psychology, marketing, business and communication such as human personality traits, brand personality and online social roles are utilized to build novel models and techniques.

The contributions of the thesis include the introduction of a novel holistic model of a digital stakeholder, a framework to utilize this model to monitor stakeholder perception towards an organization across a range of organizational aspects, a framework to model , monitor and benchmark organizational brand (to understand the organization and competitors from a stakeholder's perspective) and a framework to automate the

extraction and monitoring of varied roles and levels played by digital stakeholders (to understand the stakeholder from an organization's perspective) in near real-time.

The proposed models, techniques and frameworks have been trialled and the efficacy proven to generate strategical organizational insights in near real-time using a case study organization and a social media dataset of over 1.2 million textual conversations from public online discussion forums.

Statement of Authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Piyumini Rasangika Wijenayake

18 June 2021

List of Publications and Awards

The list of publications and awards arising from this thesis:

Journal Articles:

- Deep LSTM for generating Brand Personalities using Social Media: A case study from Higher Education Institutions. Published in The International Journal of Computer and Communication Engineering(IJCCE) and *presented at ICICA 2020-* International Conference on Intelligent Computing and Applications, Brisbane, Australia, http://www.ijcce.org/list-86-1.html*)*

Wijenayake, P., Alahakoon, D., De Silva, D., & Kirigeeganage, S. (2021). Deep LSTM for Generating Brand Personalities Using Social Media: A Case Study from Higher Education Institutions. *International Journal of Computer and Communication Engineering*, *10*(1), 17–27. https://doi.org/10.17706/IJCCE.2021.10.1.17-27

- Ethos Pathos Logos for Contemporary Organizations: Social Media Data Driven Brand Management for organizational decision-making and benchmarking. *(In review:* Journal of Industrial Marketing Management*)*

Book chapters:

- Leveraging Big Data for Organizational Performance Management & Control. In Routledge companion for Organizational Performance Management & Control (Chapter 8) (https://www.routledge.com/The-Routledge-Companion-to-Performance-Management-and-Control/Harris/p/book/9781138913547)

Alahakoon, D., & Wijenayake, P. (2018). Leveraging big data for organizational performance management and control. In *The Routledge Companion to Performance Management and Control*. https://doi.org/10.4324/9781315691374

Conference Proceedings:

- Automated detection of social roles in online communities using deep learning (Presented & Published at ICBDSC 2020- International Conference on Big Data and Smart Computing, Sydney, Australia, https://dl.acm.org/doi/abs/10.1145/3378936.3378973)

Wijenayake, P., Alahakoon, D., De Silva, D., & Kirigeeganage, S. (2020). Automated detection of social roles in online communities using deep learning. *ACM International Conference Proceeding Series*, 63–68. https://doi.org/10.1145/3378936.3378973

Awards:

- Three Minute Thesis (3MT): First Place Award 2018: La Trobe Business School. Topic: "*Care for an Artificial hug?"*
- 3MT: First Place Award 2019: La Trobe Business School. Topic: "*How about an Artificial Hug*?"
- 3MT People's choice Award 2019: La Trobe University, College of Arts, Social Sciences and Commerce.

1. Introduction

"The wettest, weirdest environment is human interaction. Whatever we build gets misunderstood, corroded and chronic, and it happens quickly and in unpredictable ways. That's one reason why the web is so fascinating-it's a collision between the analytic world of code and the wet world of people."

~Seth Godin

Human interaction has been an intriguing component of organization since the beginning of human endeavours and has witnessed a remarkable evolution from the natural - wet world to the contemporary digital universe we live in at present. In this modern digital environment, social media has prompted a lifestyle revolution, completely shifting the way individuals and organizations interact and communicate. Today, social media has widely accepted usage and come in varied forms catering to diverse needs and cohorts of the society. For example, Facebook is a platform catering to online relationships between individuals through exchange of messages, photos etc. and LinkedIn caters to careers, professionals/employees and organizations/employers and Instagram connects celebrities to fans whereas twitter provides live updates of news/events.

With this rise of adoption, diversification and specialization of social media, these social media platforms have become a real time, high velocity data bank of human opinions, emotions and experiences. Hence, today organizations are increasingly depending on social media management techniques to analyse social media content and to professionalize their social media engagement (Benthaus et al., 2016). Consequently, it is now vital more than ever before for organizations to have access to technological tools that aid in understanding and monitoring social media as a channel of human opinions, emotions and experiences.

1.1) The evolution from face-to-face interactions into digital interconnection/ interactivity

Humans were carrying out trade very early on in existence, with geography playing a key role. By 3000 BC, humans were using trade routes for interregional exchange of commodities, and by 1000 BC, merchants were becoming a part of societies or the earliest organizations. They conceived ideas about how to treat the customer. The farmer, the craftsman, the baker, and the local perfumier had to meet the needs of the customer personally, with face-to-face interaction and nascent marketing techniques (Pawlewicz, n.d.).

Societies and organizations have taken a lengthy journey since then. Interactions among individuals as well as that between organizations and individuals have witnessed a momentous evolution. Exclusively face-to-face interactions were replaced sequentially by the use of electric telephone, email, interactive voice-responses, customer relationship management (CRM) software, call-centres, help desks, online messages

(OLM), online-help desks, live chat/live messaging, social media and finally by Artificial Intelligence (AI) applications (Pawlewicz, n.d.).

Nowadays, people interact digitally to maintain relationships, for work, education and entertainment with diminishing influence of geographical and physical boundaries. Today, social media has changed the way we live. It has become widespread, unavoidable, and extremely powerful with impact on human autonomy. The current covid-19 pandemic has been an accelerator for this move of human interaction into a complete digital existence. Today, in facing the covid-19 pandemic, we see entire organizations, communities and countries carrying out their interactions, work, education, and day-to-day activities almost entirely online without any face-to-face interaction. This demonstrate that humanity has the capacity to go on digitally. Hence, digital living has the potential to become the normal practice or the new "common sense" in the coming decades.

Today, over 40% of the world population have access to the internet at home with an estimated 4.1 billion individuals accessing the internet in 2019 (*Internet live stats*, n.d.). Furthermore, majority of users are connecting through mobile-cellular devices making the computer no longer a necessity to connect to the digital world (International Telecommunication Union, 2019). According to the 2020 Global digital overview, the number of active social media users have passed 3.8 billion at the beginning of 2020 with the number increasing over 9% since 2019(Kemp, 2020). Furthermore, according to ITU statistics 93% of the entire world population lives within reach of a mobile/internet service making the possibility of these numbers to skyrocket in near future (International Telecommunication Union, 2019). If past trends are any indication, in future active social media usage is destined to be the norm for majority of the world population. This is an aspect that organizations need to be prepared for.

Hence, this thesis is an attempt to cater to this new digital world order that is inevitably coming by exploring opportunities and techniques presented through leveraging social media, to capture, analyse and understand human communication and interactions from an organizational perspective. This is attempted using digital traces of human emotions, opinions and experiences available in social media to enable organizations to understand and serve their stakeholders in a more proactive manner.

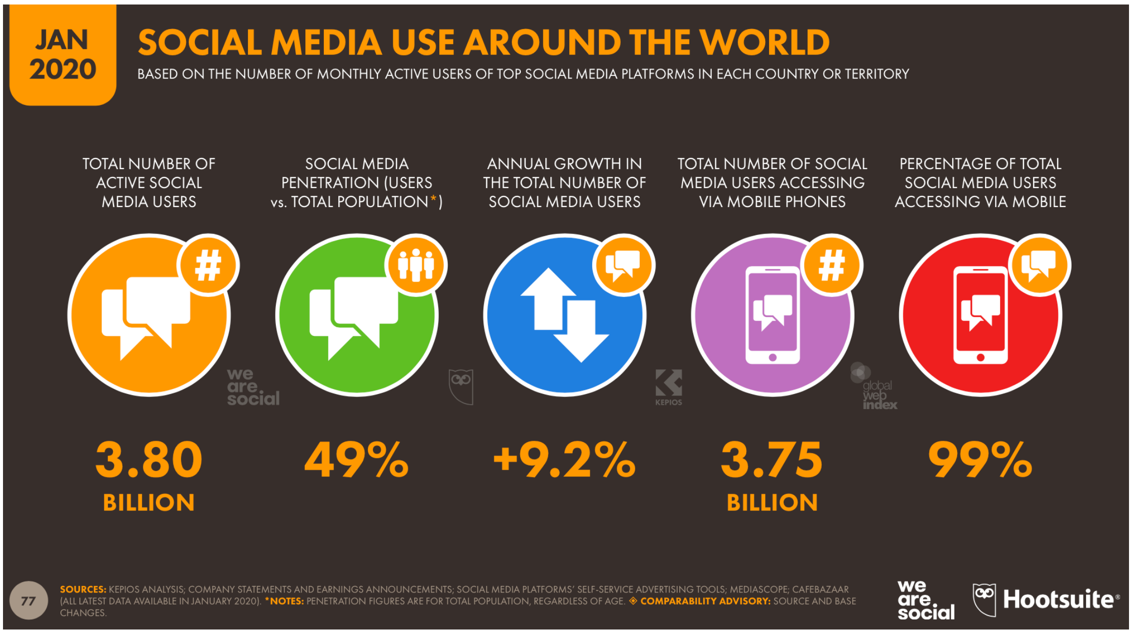


Image 1.1: A summary of global social media users around the world (Image courtesy: https://datareport.com/reports/digital-2020-global-digital-overview)

1.2) Big Data Environment, Machine Learning (analytics) and Contemporary organization

Currently humans live in an environment where information is generated at high volume, velocity and variety in relation to every aspect of human life, known as the big data environment. Big data is a term used to capture the changes brought about by the explosion in the quantity and diversity of high frequency digital data generated from an increasing number of diverse data sources. Almost all activities now generate big data that can be categorized into three main categories as a) public, b) private data and c) exhaust (Newell et al., 2012);(Benthaus et al., 2016);(Pournarakis et al., 2017);(George et al., 2014).

Public data is mostly collected by governments, governmental departments, and local authorities such as data relating to transportation, property and assets, energy use and health care. Private data refers to private information held by private firms, non-profit organizations, and individuals and include consumer transactions, radio-frequency identification tags (RFID) used by supply chains, movement of company goods and resources, web browsing, and mobile phone usage. Data exhaust or ambient data is passively collected, such as Internet searches and telephone hotlines. Although with limited value at the original data-collection point, these can be used to infer people's needs, desires and intentions. Another data source is community data such as consumer reviews on products, voting buttons, and twitter feeds also contribute to the diversity and volume of big data and is a versatile source to distil meaning to infer patterns in social structure and self-quantification data. Data self-collected by individuals through the quantification of personal actions and behaviors such as through wristbands that monitor exercise and movement is becoming another significant source (George et al., 2014).

From a public benefit aspect, this data hold the potential for decision makers to track development progress, improve social protection, and understand where existing policies and programs require adjustment (Ammu & Irfanuddin, 2013). But the manipulation and analysis of this data requires the use of powerful computational techniques to unveil trends and patterns within and between these extremely large socioeconomic datasets. New insights derived from such value extraction can be used to complement the traditionally static official statistics, surveys, and archival data sources (George G, 2014). For private organizations, such data makes it possible to better understand their customers and stakeholders, adapting their internal processes, workflows and decision making thus providing more customized services and products with better communication (Alahakoon & Wijenayake, 2018).

Moreover, over the last few decades, the information technology systems assisting businesses in managing data have seen several paradigm changes. They have moved from decision support, to executive support, to online analytical processing, to business intelligence, to analytics and now to big data (Davenport T, 2014, as cited in Alahakoon & Wijenayake, 2018).

Currently organizations are collecting data in a multitude of different forms and sources, such as formal documents, log files, databases, social media data etc. This social media data constitutes not only of structured data, but semi-structured as well as a large volume of unstructured big data, such as text.

With the recent boom of social media, stakeholders have begun to express themselves more frequently , in-depth and in-detail via social media platforms (Kietzmann et al., 2012);(Wijenayake et al., 2020). Such user generated data are rich in stakeholder opinions, emotions and experiences that are freely expressed and updated often (Wijenayake et al., 2021). This large flow of data in social media holds a huge business advantage for an organization to understand the cognitive, affective, emotional, social and physical responses of their stakeholders toward the organization. Consequently, researchers and commercial organizations are now increasingly using this social media data to inform customer analytics, human resources, recruitment etc.(Alahakoon & Wijenayake, 2018). According to Alahakoon and Wijenayake, organizations that adopt these new big data sources and technologies will be able to make real-time business decisions and thrive, while those that are unable to embrace and make use of this shift will find themselves at a competitive disadvantage (Wijenayake et al., 2021)). However, research shows social media data has not yet been fully investigated in leveraging for organizations to understand their stakeholders on a deeper level. As per research organizations' ability to harness the opportunities presented by big data available in social media is still in it's infancy (Alahakoon & Wijenayake, 2018).

As with any new development, technology or concept, to be able to reap the benefits from social media-big data it is essential that various challenges and issues associated with integrating and adapting this technology are identified and addressed (Katal et al., 2013).

Advanced analytics made up of a suite of data mining, statistics, machine learning and artificial intelligence algorithms and tools in combination with data storage and management technologies provide the capability of reaping the benefits of this social media data (Alahakoon & Wijenayake, 2018). According to (Mithas et al., 2013) advanced analytics techniques can help organizations successfully handle complex decisions by providing new insights as well as create a virtuous cycle by generating a demand for better techniques, tools, and approaches for leveraging social media data and business analytics.

Over the years, organizations have effectively used analytics to reap the benefits offered by data driven decision making and it has been demonstrated that companies that use analytics have achieved higher performance (Brynjolfsson et al., 2011). Data-driven decisions are better decisions and the use of big data enables managers to decide on the basis of evidence rather than intuition. Since they are the enabling technologies and the environment, it can be said that big data and analytics technologies have the potential to revolutionize organizational management and decision making (McAfee et al., 2012). Furthermore, recent surveys have shown a widespread belief in organizations that big data analytics offers value with half of the respondents saying that improvement of information and analytics was a top priority in their organizations (Alahakoon & Wijenayake, 2018).

However, the traditional information management and analytics have solely been based on numbers and mostly focused on internal data. Big data found in social media differs notably in this regard where most of the data is externally generated and are in unstructured formats such as text. (Alahakoon & Wijenayake, 2018). The key difficulty in analysing social media data is due to transferring opinions, ideas and emotions expressed in unstructured text into a structured form that can be evaluated, quantified and reported. This is a highly challenging process that is still under research. Moreover, the extent to which digital technology and the associated social media analytics can positively impact society and organizations in the long term, is yet unknown, leaving much potential for research in social media analytics for understanding stakeholder interactions and communication for organizations.

1.3) Digital Stakeholders of contemporary organizations

In traditional business organizations one of the key factors in keeping a customer is the quality of the customer experience. In order to create and maintain a positive customer experience, it is essential to understand the changing opinions, emotions and needs of stakeholders and their experience with the organization.

Statistics show consumer stakeholders who have an emotional connection with a brand have a 306% higher lifetime value (Morgan, 2019). Furthermore, stakeholders are willing to pay more for a great experience with companies understanding their needs and expectations (Morgan, 2019). Also, 79% of U.S. consumer stakeholders only consider brands/organizations that show they care and 70% of the buying experience is based on how the stakeholder feels they are treated (Morgan, 2019).

Customer experience is defined by Meyer and Schwager as the customer's cognitive, affective, emotional, social, and physical responses to an organization (Meyer and Schwager, 2007) as cited in (Asbjørn Følstad & Knut Kvale, 2018)). We believe this is an all-inclusive depiction that can be used for an organizational stakeholder in general.

In this digital environment, the actors who impact an organization are not just the customers or consumers. For example, for a higher education organization, parents and other peers such as career experts and alumni play a vital role in the decision of whether a customer (in this case an enrolled student) will be created or retained. Thus, all these actors become influential stakeholders for the organization. Therefore, it is important for an organization to understand who their stakeholders are (as opposed to simply who their customers/consumers are) and what responses these varied individuals have towards the organization.

Hence, it is imperative for organizations to understand these said dimensions (*cognitive*, *affective*, *emotional*, *social* and *physical)* of their stakeholders' responses to their organization to improve the stakeholder experience to maintain the organizational image in the modern competitive environment and retain the customers/consumers.

According to stakeholder theory, businesses should create value for customers, suppliers, employees, communities, and financiers (or shareholders), and pay careful attention to how these stakeholder relationships/interactions are managed to ensure creation of value for the stakeholder (Freeman, 2011). These theories have suggested the importance of businesses being managed "solely for the benefit of stakeholders" for over decades (Freeman, 2011).

This view holds immense significance today more than ever before. The modern digital environment in which the contemporary organization operate create endless possibilities to interact and build relationships online. These relationships between organizations and stakeholders, if not managed effectively has the potential to build or break an organization. To manage these relationships, stakeholders' digital presence need to be understood. This is where social media analytics presents organizations with an immense advantage.

Although research and business literature possess many theories and definitions for stakeholders and customers, there is a lack of an essential definition for a digital stakeholder. Hence, we define the cognitive, affective, emotional, social and physical understanding of a stakeholder generated through the above-mentioned digital environment as a digital stakeholder. Hence, we will be exploring the potential of social media analytics to understand this digital stakeholder in depth.

1.4) Motivation for the thesis

With the significance of human interactions and relationships, understanding of human behaviour, opinions/discussions of interests, moods, sentiments and emotions are very important for a range of reasons and domains such as government, healthcare, education, business and varied organizations. The interest in capturing such information about people is to provide products and services more suitable and customized to their needs and expectations for organizations to "create value for stakeholders".

This has been carried out mainly using techniques such as observations, questionnaires and interviews (A detailed literature review is provided in chapter 4 of the thesis). In the current digital world with the increased connectivity, interactions and globalization, the number of stakeholders as well as the range of products have increased drastically making this task extremely complex. Handling this complexity is important for organizations to keep satisfying the stakeholder. Hence, organizations have been looking at innovative approaches of doing this.

Currently with the prominence of social media, replacing physical socialization and face-to-face interaction the relationship dynamics of individuals has changed. For example, an individual may have ten friends in the physical world but 10,000 friends or connections on Facebook, Twitter or Instagram. These social media platforms are sources abundant in highly granular levels of information of varied nature related to individual stakeholders. For example platforms such as Facebook, Instagram and Myspace are rich in personal details and interactions of an individual and LinkedIn contains professional information related individuals and organizations whereas Twitter consists of highly volatile updates or news, and online discussion forums are rich in word-of-mouth related to specific topics with identity being anonymous. This data is not only freely expressed and publicly available but also updated often providing a rich timeline of events, thoughts and experiences of a global community. Organizations have started realizing the value of these sources. According to recent statistics by Forbes, 89% of all companies have already adopted a digital-first business strategy or plan to do so and 87% of companies think digital will disrupt their industry, but only 44% are prepared for a potential digital disruption.(Morgan, 2019).

Although social media analytics have been in operation for several years (as discussed in chapter 4), it's been continuously presenting challenges with the nature of big data as stated above. Therefore, Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP) etc. have been utilized to harness the value of this data. Although there have been studies in using these tools, no comprehensive investigations and studies have been reported going deeply into the individuality of the social media users, together from their social aspect and the organizational aspect as stakeholders.

In summary, the social media is available as a platform for social interaction in digital form, the data is available, and the technology is present and promising, but organizations are still behind in adoption of this data to understand the stakeholders and manage organizations to serve their purpose in satisfying and creating value for the stakeholder. Hence, this research was carried out to understand and establish a digital stakeholder, understand and monitor the image of the organization from the stakeholders' perspective and to understand the stakeholder from the organization's perspective including varied roles and levels of influence played by stakeholders using social media big data and analytics techniques for organizations to gain knowledge and insights to manage the organization's products, services ,processes and focus areas to better serve its purpose of existence.

1.5) Objectives of the thesis

1) Establish the significance, opportunities and value of the digital environment of organizational and human interactions, and emotions that are captured in social media.

2) Establish the significance of “digital stakeholders” and explore how these digital stakeholders can be understood better using analytics and the big data environment; Develop a holistic, generic model for a digital stakeholder from an organization perspective, which can be used by organizations to understand their stakeholders on a deeper level that will enable a pro-active approach to serve their stakeholders in a timely manner.

3) Explore and demonstrate how heavily underutilized social media sources can simply and effectively be utilized by organizations to understand the perceived brand Image (perceived by the eye of the stakeholder) and stakeholder needs of an organization to serve their stakeholders better; Develop a platform capability to demonstrate this by capturing the varied customer perceptions of an organization and it’s processes using public forum posts (textual big data), text mining, machine learning and NLP techniques.

4) Explore and develop analytics capabilities to understand and monitor digital stakeholders in online communities external to the organization, different roles and levels of significance and their impact on organizations. Explore the possibility to establish a technique to classify online stakeholders and automate the observation of different levels of interaction in an online community using machine learning and literature on “online social roles”.

1.6) Research Questions

Subsequent to the objectives, this thesis addresses the following research questions:

1) To what depth can knowledge and value for organizations be captured from the digital environment and social media to understand communication and interactions between organizations and individuals?

2) How can a holistic and generic framework of a digital stakeholder be developed from an organizational perspective using organization's internally generated data and the social media data sources available in the digital environment?

3) How can an organization utilize digital stakeholder opinions present in social media to develop a model of the external viewpoint of themselves and their organizational brand, to generate insights to facilitate a better stakeholder experience?

4) What key roles do digital stakeholders play in online communities from an opinion providing and information creation and diffusion perspective and how can a generic framework of automation be provided to observe these online stakeholder roles?

1.7) Robustness of the approaches used

This section presents a brief discourse on the robustness of the approaches used in this thesis. This thesis focuses on leveraging social media to develop novel frameworks, models and techniques to enhance communication and understanding between organizations and stakeholders. The data sources used and demonstrated on for this research are social media sources, and thus not traditional data sources that have been used by organizations. This research thesis reveals and demonstrates novel opportunities by augmenting with social media as a non-traditional data source for organizations to understand and serve their stakeholders. While there could be a general hesitancy in utilizing such non-traditional big data sources to understand and measure stakeholder perceptions, there is a trail of evidence and indication in the applicability, value and robustness of such approaches.

A review conducted on social media analytics and business research by (Choi et al., 2020) recommends online voice of customers (VoC) expressed in social media as an effective contemporary data source for researchers to conduct business intelligence research while acknowledging a dearth in studies that demonstrate leveraging social media from a business intelligence research perspective. This provides further justification toward the stringent need to conduct this research endeavour presented in this thesis.

According to (Choi et al., 2020), easy access and flexibility have made social media the most powerful and effective window for customers to voice their opinions during the last decade. As per (Xiao et al., 2016) and (Choi et al., 2020), approximately 77% of customers read other customers' online product reviews, 75% of them trust the reviews rather than individual direct recommendations, and more than 80% of them obtain helpful advice or information through social media-based channels. In addition, As established by (Balbi et al., 2018) and (Choi et al., 2020) majority of individuals express their opinions about a product or service without any censorship when posting or commenting on social media channels as opposed to completing surveys, questionnaires and interviews for organizations to conduct evaluations for their brand performance and customer satisfaction, where an humane need to restrict expressing true emotions, opinions and experiences could easily arise.

According to (Rose et al., 2011), subjective opinions of customers provided online can determine whether other potential customers will purchase a product or not. Furthermore, as established by (Choi et al., 2020) modern customers are not passive buyers of products/services, but active participants who control the future development direction of products/services and its survival, either directly or indirectly through their powerful voices that are embedded in social media.

Further justification for using social media to generate organizational insights presented in this thesis such as perceived organizational brand, algorithms for organizational benchmarking and a social media empowered digital stakeholder can be exemplified through an examination into the intrinsic properties of social media data such as volume, velocity and variety.

The massive volume of conversations available on social media provides opportunities for high levels of accuracy in analysing opinions, experiences and emotions as opposed to the limited number of input that can be obtained by implementing traditional methods used for evaluating customer opinions, such as questionnaires and interviews. High velocity of social media data provides the advantage of recency of information to near real time. This is simply not possible with the stated traditional methods and sources which necessitate a significant expenditure in terms of cost, time and human resources, limiting the frequency at which evaluations can be conducted. Furthermore, the inconceivable variety of information embedded in social media conversations present opportunities to uncover and understand stakeholder perceptions on novel areas that have not been known to the organization. In contrast, traditional methods of questionnaire and interviews provide a pre-conceived template (structuring at point of input) for stakeholders to express themselves which can limit the range and accuracy of expressed opinions, emotions and experiences. The unstructured nature of social media provide a vast repertoire of stakeholder perceptions for analysis as structuring occur at the point of processing and analysis rather than at the point of input, even though it necessitates the use of advanced analytics techniques such as AI and ML.

Moreover, there exist recent research studies that have effectively utilized AI and ML to extract emotions and value from text conversations found in social media demonstrating the practical applicability of such data for value creation and decision making.

For example, a study by (De Silva et al., 2018) presented an AI based analytics framework to support "*social media empowered*" patients and recommends social media to be used as a vital resource for *patient-centred* care. (T. R. Bandaragoda et al., 2017) successfully utilized user generated micro blog discussions to detect events by leveraging ML and emotion detection. A research conducted by (Adikari et al., 2020) highlighted how conversations from online cancer support groups can be effectively used to observe psychological morbidity of patients and provide support using AI, ML and NLP techniques (Choi et al., 2020).

In sum, it is evident that the use of digital traces of stakeholder opinions, emotions and experiences available through social media is a viable source for developing frameworks, models and techniques to enhance communication and understanding between stakeholders and organizations.

1.8) Summary of contributions and the thesis outline

This section presents a summary of the research contributions presented in this thesis. The thesis consists of the introduction followed by a literature review (chapter 2), four contributing chapters (chapters 3, 4, 5 and 6) and the thesis conclusion (chapter 7). The figure 1.1 presents a visual diagrammatical illustration of the chapter contributions and the relationships between the seven chapters of the thesis.

The next chapter, the literature review presents a recollection of the fundamental concepts and existing work related to the context of the research questions presented above. It presents a review and summarization of existing research, theories and techniques of machine learning that can be used for the identification and understanding of individuals and organizations that identify each other through the digital world with a special focus on social media.

Chapter 3 defines and establishes the concept of “digital stakeholder” from an organizational and social media perspective through an exploration into the nature of human cognition, the contemporary organization and social media. Furthermore, the chapter introduce and demonstrate a novel holistic and generic model of a digital stakeholder and a novel framework to apply this model to an organization domain and monitor the digital stakeholder’s journey through an organization by augmenting organizational data sources with social media.

Chapter 4 explores techniques for understanding the digital stakeholder's perception about an organization (Perceived Organization) from social media. A novel framework and techniques are proposed and demonstrated for understanding the Perceived Organization from a social media perspective that structure and transform social media into lenses or aspects of the organization. Also, novel techniques to remodel raw unstructured social media data to generate knowledge of these organizational aspects using digital traces of stakeholder emotions and opinions are proposed. Furthermore, the chapter presents a detailed unpacking of organizational features and aspects represented and discussed over social media in the form of public online discussion forums. Techniques to leverage social media and monitor the organizational perception from varied lenses of products, processes and focus areas as well as positioning the organization among its competitors are presented.

Chapter 5 takes a step further and demonstrate how these heavily underutilized, unstructured, high volume, high velocity and publicly available social media sources can simply and effectively be utilized to understand the ethos, pathos and logos of an organization's brand Image, and realize stakeholder needs of the organization to serve their stakeholders better. This chapter proposes a technique with alternative methods through low to high use of machine processing and automation to suit the technological capacity, human resources and culture of different organizations. Furthermore, a demonstration of how this method enables organizations to extract their competitors' brand image, competitors' organizational processors and issues and allows benchmarking for a competitive advantage is presented. Demonstrations of how the proposed technique can effectively monitor brand consistency and the harmony of brand architecture at different granularities, across a spectrum of elements within an organization is presented.

Chapter 6 presents an exploration into understanding the dynamic online social roles played by this established digital stakeholders in social media communities external to an organization from an opinion providing and information creation and diffusion perspective that impacts the image and brand of organizations. The chapter presents and demonstrates a methodological approach for automated detection of social roles in online communities by utilizing machine learning technologies. Having established in previous chapters; the value and importance of understanding stakeholder conversations in social media to comprehend stakeholder opinions and emotions related to an organization; *online social roles* are identified and established as a highly useful and versatile tool for organizations to take a step further in knowledge discovery to understand 'who' and 'why' is responsible for creating these online conversations (that are explored in chapters 4 and 5 to generate organizational insights) and drawing-in and diffusing information related to the organization from the real world into the digital world and influencing the organization's image and brand.

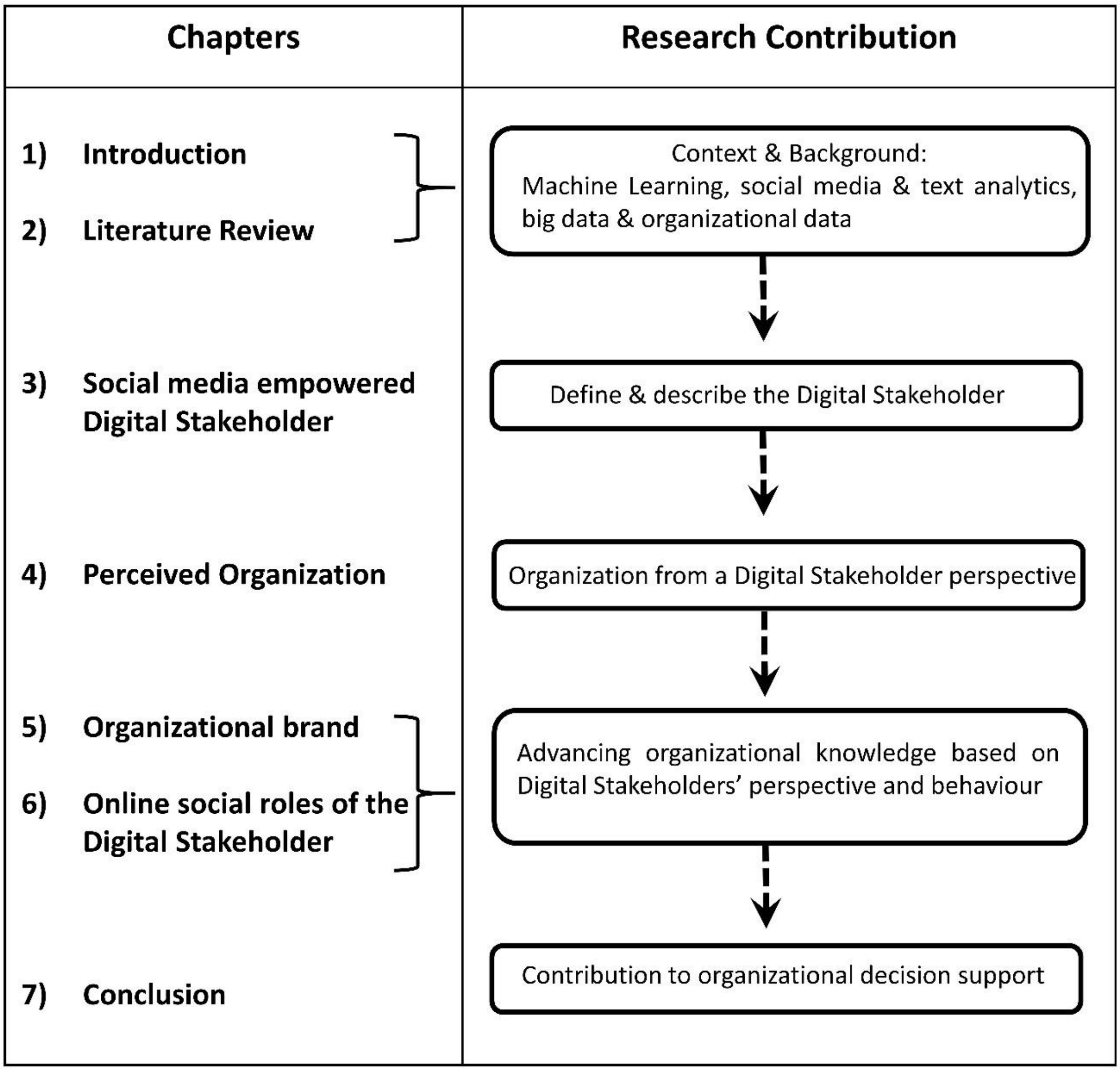


Figure 1.1: Thesis diagram outlining the research contributions and relationships between the thesis chapters. The diagram illustrates how the chapters are interconnected, with research contributions being shared throughout the discourse of the thesis. The diagram depicts how technology and the digital environment is leveraged to understand the individual as a 'Digital Stakeholder' and the 'Organization' and through that gain organizational knowledge and insights to support organizational decision making and strategy.

2. A Review of Related Research, Theories and Techniques

"If I have seen further, it is by standing on the shoulders of giants."

~ Isaac Newton

This chapter is devoted to the recollection of the fundamental concepts and existing work related to the context of the research questions presented in this thesis. The chapter reviews, surveys and summarizes existing research, theories and techniques of machine learning for the identification and comprehension of individuals and organizations who identify each other through the digital world; specifically, social media. The following diagram (figure 2.1) presents a diagrammatical illustration of the topics discussed in this chapter and their relationships to the contribution chapters of this thesis.

The figure 2.1 presents an outline of the topics covered by the key pillars of literature investigated for the research endeavour presented in this thesis. The key pillars of research literature this thesis builds on are: "Individual as a digital stakeholder for organizations" (individuality of stakeholders), "organization and it's image in the eye of an individual" (organizations' brand and impression), "big data, social media and analytics” (the digital environment and the opportunities presented for organizations) and “an overview of machine learning techniques”(technologies available to leverage the opportunities presented by the digital environment). From here on, this chapter will present the literature review categorized according to these key pillars of literature as presented diagrammatically in the figure 2.1.

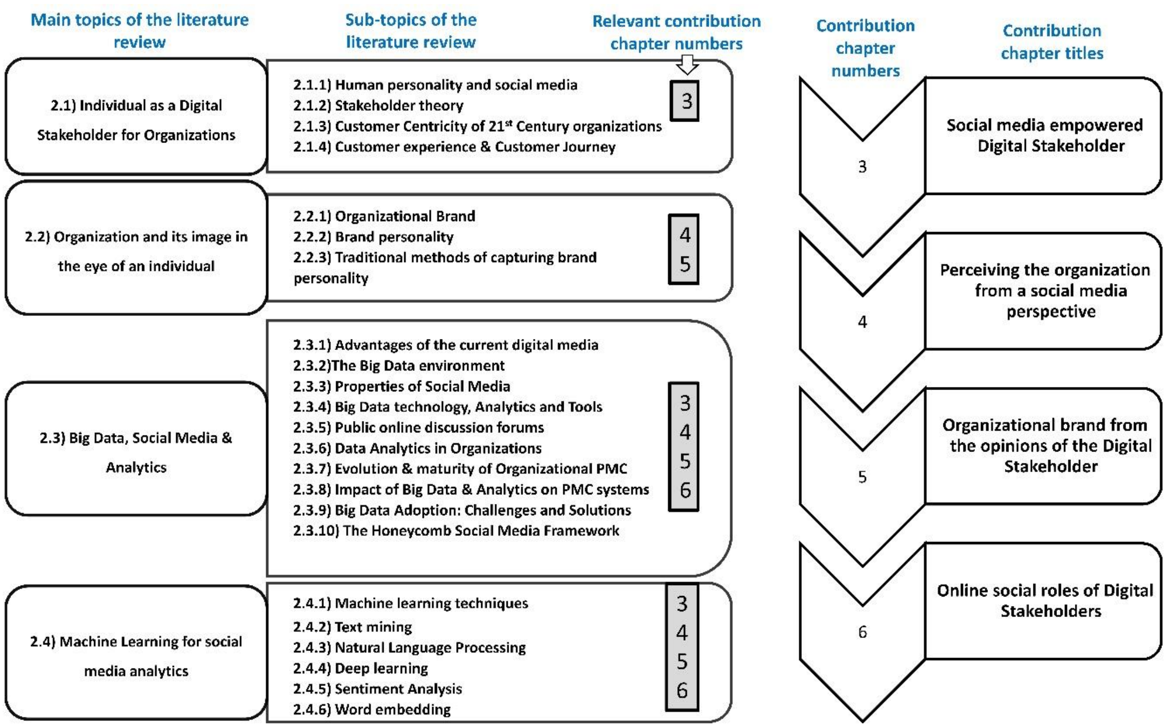


Figure 2.1: The topics discussed in chapter 2 and their connection to the contribution chapters in the thesis. The two columns on the left side present the major topics and the subtopics introduced in the literature review. The two columns on the right side present the chapter numbers and the chapter titles of the contribution chapters of the thesis. The numbers presented in the shaded squares represent the chapter numbers of the contribution chapters that discusses the topics in chapter 2, the literature review.

The figure 2.1 presents an outline of the topics covered by the key pillars of literature investigated for the research endeavour presented in this thesis. The key pillars of research literature this thesis builds on are: “Individual as a digital stakeholder for organizations” (individuality of stakeholders), “organization and it’s image in the eye of an individual” (organizations’ brand and impression), “big data, social media and analytics” (the digital environment and the opportunities presented for organizations) and “an overview of machine learning techniques”(technologies available to leverage the opportunities presented by the digital environment). From here on, this chapter will present the literature review categorized according to these key pillars of literature as presented diagrammatically in the figure 2.1.

2.1) Individual as a Digital Stakeholder for Organizations.

This section presents a collection of summarized discussions from literature on the topics of *human personality and social media*, *stakeholder theory, customer centricity of 21st century organizations* and the *importance of customer experience and customer journey*. These topics are discussed further in chapter 3 to answer the research question 2: “*How can a holistic and generic framework of a digital stakeholder be developed from an organizational perspective using organizational data and the digital environment?”*

2.1.1) Human personality and social media

According to (Farnadi et al., 2016) and (Ozer & Benet-Martinez, 2006), research in psychology suggest that behaviour and preferences of individuals can be explained to a great extent by underlying psychological constructs or personality traits. Previous research in the field of psychology as well as human computer interaction (HCI) have emphasized the importance of recognizing and classifying users' personality traits and preferences to support building adaptive and personalized systems in order to provide rich and improved user experiences (Farnadi et al., 2016).

This means that knowledge of an individual's personality allows us to make predictions about preferences across contexts and environments, and to enhance recommendation systems (Lambiotte & Kosinski, 2014; Pu et al., 2011). Furthermore, an individual's personality can affect the decision making process and has been shown to affect preferences for products and services, brands, websites (Kosinski et al., 2013), and for content such as movies, TV shows, and books (Cantador et al., 2013; Farnadi et al., 2016).

The most widely accepted model of personality is the Big Five or Five Factor Model. This model presents five traits; Openness, Conscientiousness, Extroversion, Agreeableness, and Emotional Stability (often conversely referred to as Neuroticism) (Costa Jr & McCrae, 2008; Farnadi et al., 2016). A detailed description of each of the five traits are presented in table 2.1 from (Farnadi et al., 2016).

Trait	Description
Openness	Openness is related to imagination, creativity, curiosity, tolerance, political liberalism, and appreciation for culture. People scoring high on Openness like change, appreciate new and unusual ideas, and have a good sense of aesthetics.
Conscientiousness	Contrast to a spontaneous one. People scoring high on Conscientiousness are more likely to be well organized, reliable, and consistent. They enjoy planning, seek achievements, and pursue long-term goals. Non-conscientious individuals are generally more easy-going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans.
Extroversion	Extroversion measures a tendency to seek stimulation in the external world, the company of others, and to express positive emotions. People scoring high on Extroversion tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative; they do not mind being at the centre of attention and make new friends more easily. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation.
Agreeableness	Agreeableness relates to a focus on maintaining positive social relations, being friendly, compassionate, and cooperative. People scoring high on Agreeableness tend to trust others and adapt to their needs. Disagreeable people are more focused on themselves, less likely to compromise, and may be less gullible. They also tend to be less bound by social expectations and conventions, and more assertive.
Emotional Stability	Emotional Stability, reversely referred to as Neuroticism, measures the tendency to experience mood swings and emotions such as guilt, anger, anxiety, and depression. People scoring low on Emotional Stability (high Neuroticism) are more likely to experience stress and nervousness, while people scoring high on Emotional Stability (low Neuroticism) tend to be calmer and self-confident.

Table 2.1: Overview of the Big Five personality model, by (Farnadi et al., 2016).

The traditional methods used to measure personality requires participants to answer a series of questions (typically, from 20 to 360) evaluating their behaviour and preferences (e.g., (Costa Jr & McCrae, 2008; Farnadi et al., 2016; John et al., 1999)). Farnadi et al, demonstrate how this approach is time consuming and impractical, especially in the context of on-line services; as on-line users might be unwilling to spend time filling-in a questionnaire to personalize their search results or product recommendations.

However, it has been shown with recent research studies that the digital footprint of users can be used to automatically infer their personality. For example, studies conducted by (Kosinski et al., 2013) and (Youyou et al., 2015) demonstrated that automated personality judgments based on Facebook *Likes* are more accurate than the personality judgements made by users' friends or their closest of kin. Furthermore,(Chung & Park, 2017) presented how similar predictions can be made based on language usage in social media. Additionally, a range of other methods have been proposed using different prediction mechanisms, feature spaces, and focusing on different on-line environments ((Celli & Rossi, 2012; Quercia et al., 2011) as presented in (Farnadi et al., 2016)).

Research has been conducted on the use of multivariate regression for personality prediction on Facebook (Iacobelli & Culotta, 2013) and YouTube (Farnadi et al., 2016). Most recent research studies such as (Marouf et al., 2020) ,(Vora et al., 2020) and (Kunte & Panicker, 2020) reveal the use of feature selection algorithms , machine learning, ensembles and deep learning to predict personality traits from social media.

According to (Farnadi et al., 2016) and (Vora et al., 2020), Social media provide a unique opportunity to capture various aspects of user behaviour. This is because, in addition to users' structured information contained in their profiles, e.g., demographics, users produce large amounts of data about themselves in a variety of ways including textual (e.g., status updates, blog posts, comments) or audio-visual content (e.g., uploaded photos and videos) (Farnadi et al., 2016).

Research shows that latent variables such as personalities, emotions and moods—which, typically, are not explicitly given by users—can be extracted from this user generated content ((Back et al., 2010; Farnadi et al., 2016, 2013; J. Golbeck et al., 2011) as presented in (Farnadi et al., 2016; Marouf et al., 2020; Vora et al., 2020)).

In sum, research into automatic personality prediction using social media data is a highly promising area which is gaining increased research attention due to its potential in many computational applications (Farnadi et al., 2016; Marouf et al., 2020; Vora et al., 2020).

2.1.2) Stakeholder theory

Over the last thirty years a notable view emerged from a group of scholars in a diverse range of disciplines, from finance to philosophy called the stakeholder theory. This elaborate that businesses should create value for customers, suppliers, employees, communities, and financiers (or shareholders), and pay careful attention to how these stakeholder relationships/interactions are managed to ensure creation of value for the stakeholder (Freeman, 2011). These theories suggest the importance of businesses being managed "solely for the benefit of stakeholders"(Freeman, 2011).

This view holds immense significance today more than ever before. The modern digital environment in which the contemporary organization operate create endless possibilities to interact and build relationships online. These relationships between organizations and stakeholders, if not managed effectively has the potential to build or break an organization. To manage these relationships, stakeholders' digital presence needs to be understood.

In 1992, Hill and Jones developed the "agency stakeholder model" where the organization was viewed as a nexus of implicit contracts between stakeholders and market processes, and these contracts were considered to be driven by organizational structures that evolve to monitor and enforce the terms of implicit contracts (Hill, C. W., & Jones, 1992); (Friedman & Miles, 2002). Mitchell et al. developed the awareness that stakeholders become salient to managers to the extent that those managers perceive stakeholders as possessing power, legitimacy and urgency (Mitchell et al., 1997). However, this research delved into only a brief discussion on the dynamics of the stakeholder–manager relations. A model which is capable of explaining the dynamics of stakeholder-organization relationship over time was introduced by Friedman and Miles (Friedman & Miles, 2002). This model combines stakeholder theory with realist theory of social change and differentiation, and enables an analysis of the stakeholder-organization relationship (Friedman & Miles, 2002).

2.1.3) Customer centricity of 21st century Organizations

According to Peter Drucker, the father of business consulting**,** "Business exist for the pure purpose of creating and keeping a customer", and "Profit, is not the explanation, cause, or rationale of business behaviour and business decisions, but the test of their validity” (Drucker, 1974). An organization’s purpose must lie outside of the business itself, in society since business organization is an organ of society (Drucker, 1974).

We are currently witnessing an era where organizations are shifting from a product-centric paradigm into a customer-centric paradigm. Organizations that have succeeded in navigating this paradigm shift have proven customer centricity to be an essential condition for an organization to succeed in the 21 century marketplace (Shah et al., 2006). These organizations have managed to secure valuable rewards in the form of customer loyalty as well as financial performance. This is due to customer centricity empowering a competitive advantage that has been proven to be sustainable and not easily refuted by competition (Shah et al., 2006). Historically, organizations were disposed to be product-centric and hence were more internally oriented with major focus on manufacturing superior products rather than on being oriented toward the purchasers and users of those products((Levitt 1960) as cited in (Shah et al., 2006)).

2.1.4) Importance of Customer experience and Customer Journey

One of the key factors in keeping a customer is the quality of the customer experience. Customer experience is presented as a key entity for competitive advantage in the service sector (Asbjørn Følstad & Knut Kvale, 2018). As per Meyer and Schwager, customer experience is continuously affected by every communication/interaction between the customer and the organization, driving a significant need to manage this experience through design across customer touchpoints and services (Meyer et al., 2007).

Customer experience is defined by Meyer and Schwager as the customer's cognitive, affective, emotional, social, and physical responses to an organization ((Meyer et al., 2007) as cited in (Asbjørn Følstad & Knut Kvale, 2018)). Hence, it is imperative for organizations to understand these dimensions (cognitive, affective, emotional, social and physical) of their stakeholders' responses to their organization to improve the customer experience and retain the customer.

Furthermore, a concept closely linked to customer experience throughout literature is customer journey. Throughout literature, customer journey perspective had been utilized to support the management and design for customer experience (Asbjørn Følstad & Knut Kvale, 2018). As per Følstad and Kvale customer journeys have been understood to concern more than just the observable series of steps or touchpoints which a customer goes through as part of service provision; customer journeys have also taken into consideration the emotional and cognitive responses of the customer (Asbjørn Følstad & Knut Kvale, 2018). Past research studies demonstrate how mapped customer journeys have been used to describe what customers think and feel in addition to their observable behaviour at each step and touchpoint of interest to the organization (Asbjørn Følstad & Knut Kvale, 2018).

2.2) Organization and its image in the eye of an individual

2.2.1) Organizational Brand

"For decades, marketers have sought the Holy Grail of *brand loyalty*. Just as the legendary grail of Arthurian quest held the promise of extended life and renewal, marketers attribute to *brand loyalty* and its sister icon *customer retention*; the promise of long-term profitability and market share. Unfortunately, marketing's knights-errant face a daunting problem: They have not fully understood what the grail looks like or where it can be found. As a result, marketers have devised strategies and designed programs to build loyalty with limited information about their real impact or ultimate consequences“ (Bhattacharya, Rao, and Glynn 1995; Reicheld and Sasser 1990; Dowling and Uncles 1997; Fournier, Dobscha, and Mick 1998 as cited in Mcalexander et al., 2002).

As illustrated through the above quote, throughout literature, decades of research have established that *Brand* or the consumers’ cognitive associations for an organization as a strategic asset, as well as a source of sustainable competitive advantage (D. A. Aaker, 2012; Brown & Dacin, 1997; Dowling, 1986; Hall, 1993). Maintaining this brand image has been considered an important strategic task and organizations spend a great sum of money each year on corporate advertising, corporate philanthropy, sponsorships, cause-related marketing, and public image studies (The Conference Board 1994; Kinnear and Root 1995; Schumann, Hathcote, and West 1991; Smith and Stodghill 1994 as cited in (Brown & Dacin, 1997)). Thus, it is widely accepted that brand management, brand orientation, corporate vision, strategic direction of the brand and implementing brand strategy are critical responsibilities of an organization’s senior management (de Chernatony and Dall'Olmo Riley and Harris, 1998; Wong & Merrilees, 2007; Mosmans, 1996; Urde, 1994, 1999 as cited in (Hankinson, 2012)).

2.2.2) Brand personality

A key construct that has been given considerable attention throughout consumer behaviour and marketing research is the concept of *Brand Personality* that imbue dimensions or traits to conceptualize *Brand* in a similar way that *human personality* has been conceptualized with the likes of the Big Five Personality traits discussed earlier(J. L. Aaker, 1997). The term Brand Personality, was first introduced by Martineau in 1958 (Martineau, 1958) and refers to a set of human characteristics or stakeholder perceptions associated with a brand (A. Xu et al., 2016).

Varied research has been conducted into developing brand personality frameworks and scales and into determining the number and nature of brand personality dimensions or traits. The foundation for vast majority of brand personality scales used throughout literature is Aaker's(1997) seminal dimensions of brand personality which have evolved and have been extended extensively over time by academics and practitioners into many specific areas and industries (J. L. Aaker, 1997). Since 1997, most of the marketing literature has adopted Likert scale surveys(Likert, 1932) based on the Aaker's scale to evaluate brand personalities (A. Xu et al., 2016).

As defined in business and marketing literature, the concept of Brand Personality can be used to differentiate products and companies (Clifton & Simmons, 2003) and brand personality scales have been demonstrated to be a reliable, effective and generalizable tool for evaluating *Brand Personality* (A. Xu et al., 2016).

2.2.3) Traditional methods of modelling/capturing organizational brand personality

Traditional techniques, most widely used throughout literature for the extraction of this brand personality, are questionnaires and interviews conducted online and offline (Matzler et al., 2016; Sung et al., 2015; So et al., 2017; Gordon et al., 2016; Zenker et al., 2017; Chung and Park, 2017; Hultman et al., 2015).

Some of the more contemporary additions are online product reviews and social media (Saboo et al., 2016; Garanti & Kissi, 2019; Gensler et al., 2015; Unnisa et al., 2016; Hudson et al., 2016 ; Callarisa et al., 2012; Carpentier et al., 2019).

A study by Callarisa et al. (2012) used social media data to analyse the different components of brand equity and their relations in the hotel industry using 11,912 online customer reviews from the TripAdvisor site based on 653 hotels (Callarisa et al., 2012). Another study conducted by Xu et al.(2016) used social media to predict the perceived brand personality and the association with its three contributing factors embodied in social media (*User Imagery*, *Employee Imagery*, and *Official Announcement)* using a combination of survey data and social media data (A. Xu et al., 2016). A total of 10,950 survey responses were collected, for 219 brands from 3060 participants. The study managed to define the relative importance of three main factors contributing to perceived brand personality for each of the five dimensions: sincerity, excitement, competence, sophistication and ruggedness. As a result, an illustration of how modelling brand personality can help users find brands suiting their personal characteristics and help companies manage brand perceptions was proposed. In 2017, (Pournarakis et al., 2017) introduced a computational model for mining consumer perceptions in social media using a novel genetic algorithm to improve clustering of data to produce insights into brand awareness and brand meaning. The model was used to analyse over 280,000 tweets related to the Uber transportation network and to reveal negative and positive consumer perceptions for specific subjects related to Uber. A few recent studies such as (Carpentier et al., 2019) and (Garanti & Kissi, 2019) have attempted to observe organizations' social media brand personality through how consumers chose to engage with the organization's social media page. However, these studies have not attempted to use raw social media text comments/posts by consumers to generate the brand personalities perceived by consumers, but rather study how individuals are engaging (or prefer to engage) with the organization created and managed social media pages.

Lu et al. performed text analysis on posts in Facebook fan pages to discover factors that affect a post to attract consumer attention. Keywords representing brand personalities were analysed for this task from 14 Facebook fan pages of e-commerce companies together with an online questionnaire (Lu et al., 2019).

Robertson et al. explored the role played by employees in creating employer brand positioning online by performing content analysis in DICTION (a computer aided text analysis program) on 6300 employee reviews from the social media platform Glassdoor (Robertson et al., 2019).

A study by (Hu et al., 2019) demonstrate a text analytics framework for brand personality prediction that uses social media data. This study used text analysis and Elastic-Net regression analysis and utilized self-descriptions of consumers who followed the sample of brands on Twitter, employee reviews of the brands' firms from Glassdoor and brands' official tweets. This study has successfully established the value of using social media analytics for predicting and measuring brand personality with high accuracy. The study demonstrates that profiles of individuals who choose to associate with brands on social media is an important predictor of brand personality through *User Imagery*; and provides evidence for a consumer identity-brand personality link. The study also identifies a link between an organization's internal corporate environment as perceived by employees through *employee imagery* and brand personality.

While these recent studies have utilized text analytics to observe a brand personality scale through social media, these presented methods do not provide a mechanism to predict brand personality when consumers and consumer social media profiles are unknown. Glassdoor possess voluntary online reviews in contrast to freely expressed word of mouth present in public online discussion forums (discussed in section 2.3.5). Also, the existing studies do not present a generic method that can be used to predict brand personality from public online discussion forums which carry immense value that is heavily underutilized by research and business (further discussed in section 2.3.5 below). This reveals a clear contemporary business need for a generalizable method of social media analytics to predict and monitor brand personality in near real time through high velocity social media conversations.

The traditional methods that have been widely used and accepted, while being effective have required time, resources, money and experts in domain knowledge to design, execute and analyse. Furthermore, the quantity of data (sample size) that could be collected and processed has been limited due to the time and resource consuming nature of data collection and analysis. This confines the quality of results. Furthermore, there had been no studies or opportunities to extract this information for competitor brands using the accepted traditional methods.

When considering branding in the higher education industry, many recent successful efforts have been made using the aforementioned traditional techniques.

The paper by (Khanna et al., 2014) have come up with a touch-point wheel for developing a higher education brand based on 276 questionnaires from students and alumni of management schools across Mumbai developed after in-depth interviews and a literature review. However, the sample came mostly from students studying in just one management school in Mumbai. The study by Naidoo and Hollebeek on higher education brand alliances, examined prospective students' purchase intentions for duel degrees; a specific brand alliance type arrayed in the higher education industry. Eight particular dual degrees were studied by interviewing 24 prospective students. Independent researchers were used to generate key themes, analyse and interpret the data. Hence, the investigation was limited to the exploration of only a specific set of variables derived from the initial qualitative phase (Naidoo & Hollebeek, 2016).

It is evident, that these techniques while being accepted and successful are subject to limitations. One of the inherent limitations of survey-based methods is flexibility. Conducting surveys can be highly time-consuming and labour-intensive, thus limiting the sample size to hundreds or several thousand at most. In addition, it is expensive to assess brand personality frequently. Furthermore, survey-based methods suffer from non-response and sampling related discrepancies, and bear the risk of experimenter bias (A. Xu et al., 2016).

In summary, it can be observed that the past approaches used for extracting brand personality utilized primarily traditional instruments of questionnaires, interviews and surveys for data collection with the exception of a few recent studies using experimental social media analytics techniques. Collected data mainly consisted of numeric data and limited structured text data. Data processing and analysis was carried out mainly using manual/human intervention of domain experts utilizing statistical tools to support analysis. Social media data have been used in recent years with very limited usage, limited adoption of advanced analytics and algorithms and a dearth of generic methods for analysing vast high velocity social media resources. Organizations still struggle with the processing of large volumes of unstructured textual data being generated at high velocity, as well as with intelligent separation of tasks for automation and manual intervention. This is due to the lack of knowledge and expertise in the adoption of Text Mining (TM), Machine Learning (ML) and Natural Language Processing (NLP). However, a vast array of algorithms, tools and technologies are becoming widely available, such as topic extraction, neural networks, deep learning, word embedding and sentiment analysis, which can be utilized by organizations according to their acceptance, capability and availability of expertise and culture. Chapter 5 of this thesis proposes an approach which caters to the wide range of such organizations by introducing a generic, reproducible machine learning based methodology of three possible pathways with low to high use of technology and automated processing.

2.3) Big Data, Social Media and Analytics

Big data and analytics have not only made a significant impact on organizational performance management and control (PMC) but has been a key driver in the evolution of PMC systems. It is expected that this trend will continue and that the current data driven level of maturity will move on to the next generation of mobile device and web-based systems which also work with much higher granularity of data. New algorithms, tools and technology will develop to support these requirements with advanced analytics and machine learning extending into deep learning and cognitive computing based highly intelligent algorithms (Alahakoon & Wijenayake, 2018).

This section provides an overview of the opportunities presented by the current digital environment in which businesses operate and a brief introduction to the term big data with some background on how and why this is significant from a business and organizational perspective. The technologies and tools for harnessing the value of big data are presented and key data analytics techniques are described.

In the last few years, big data technologies have gained considerable attention due to its potential to transform data mining and business analytics practices and the possibility for a wide range of highly effective decision-making tools and services. This section explores big data analytics and its impact on organizational performance management and control systems. The concept of performance management analytics is discussed and a model of the components of performance management analytics is developed and proposed. The section also discusses the evolution of management control systems and the new analytics technologies and systems that have come up to fill the requirements of the new systems. The evolution of organizations' PMC systems is investigated from corporate planning capability aspect as well as reporting perspective. The current maturity model for PMC is extended with a new data driven level and a future mobile and internet of things (IOT) environment-based level of maturity. Main challenges in the adoption of big data in organizations are summarized (Alahakoon & Wijenayake, 2018).

2.3.1) The advantages of the current digital environment

"The information technology (IT) revolution in the latter part of the 20th century introduced extraordinary improvements in collecting, storing, analysing, and transmitting huge amounts of information. Firms realized that this presented a great opportunity to invest in IT for managing customer relationships. Customer relationship management (CRM) became a buzzword and companies started investing millions of dollars in CRM software packages, database marketing initiatives, and IT infrastructure to support technology-driven marketing"(Shah et al., 2006).

Moreover, over the last few decades, the information technology systems assisting businesses in managing data have seen several paradigm changes. They have moved from decision support, to executive support, to online analytical processing, to business intelligence, to analytics and now to big data (Davenport T, 2014, as cited in (Alahakoon & Wijenayake, 2018)).

Currently organizations are collecting data in a multitude of different forms and sources, such as formal documents, log files, databases, social media data etc. This data constitutes not only of structured data, but semi-structured as well as a large volume of unstructured big data, such as data from social media, forums (public opinions) and location awareness (movement patterns of individuals). This large flow of data holds a huge business advantage for an organization to understand the cognitive, affective, emotional, social and physical responses of their customers toward the organization.

2.3.2) The Big Data Environment

Big data is a term used to capture the changes brought about by the explosion in the quantity and diversity of high frequency digital data generated from an increasing number of diverse data sources (Ammu & Irfanuddin, 2013). It could also be thought of as the new data and information intensive environment we live in sometimes called the information age (Alahakoon & Wijenayake, 2018). Almost all activities now generate data and (George et al., 2014) has categorized such data into three main categories as ;

a) public,

b) private data and

c) exhaust.

Public data is mostly collected by governments, governmental departments, and local authorities such as data relating to transportation, property and assets, energy use and health care. Private data refers to private information held by private firms, non-profit organizations, and individuals and include consumer transactions, radio-frequency identification tags (RFID) used by supply chains, movement of company goods and resources, web browsing, and mobile phone usage. Data exhaust or ambient data is passively collected, such as internet searches and telephone hotlines. Although with limited value at the original data-collection point, these can be used to infer people's needs, desires and intentions.

Another data source is community data such as consumer reviews on products, voting buttons, and twitter feeds which also contribute to the diversity and volume of big data and is a versatile source to distil meaning to infer patterns in social structure and self-quantification data.

Data self-collected by individuals through the quantification of personal actions and behaviors such as through wristbands that monitor exercise and movement is becoming another significant source (Alahakoon & Wijenayake, 2018; George et al., 2014).

From a public benefit aspect, this data hold the potential for decision makers to track development progress, improve social protection, and understand where existing policies and programs require adjustment (Ammu & Irfanuddin, 2013). But, the manipulation and analysis of this data requires the use of powerful computational techniques to unveil trends and patterns within and between these extremely large socioeconomic datasets. New insights derived from such value extraction can be used to complement the traditionally static official statistics, surveys, and archival data sources (George et al., 2014). For private organizations, such data makes it possible to better understand their customers and stakeholders, adapting their internal processes, workflows and decision making thus providing more customized services and products with better communication (Alahakoon & Wijenayake, 2018).

However, the term "big data" has misleadingly focused attention to the volume or the size of the data. It is important to understand the properties and characteristics which make up big data to realize that big data is much more than a large volume of data.

2.3.3) Properties of Social Media as a source of Big Data

A widely known definition of big data refers to the 3 Vs; volume, velocity and variety as the key properties and these properties apply to social media data (Alahakoon & Wijenayake, 2018).

Volume

Volume means that there is a growing quantity of data. At present, data exists in petabytes and is expected to increase to zeta bytes in near future. The current social networking sites generate terabytes of data on a daily basis (Alahakoon & Wijenayake, 2018).

Velocity

Velocity refers to the speed of the data being generated from various sources. This characteristic is not limited to the speed of incoming data but also speed at which the data flows (Katal et al., 2013). Speed of data creation has become even more important than the volume in situations where the ability to capture and learn near real time from data is a crucial factor. Real-time or nearly real-time information makes it possible for an organization to be much more agile than its competitors (Alahakoon & Wijenayake, 2018; McAfee et al., 2012).

Variety

Highly diverse types of data are generated by organizations and individuals. It includes traditional data but also the semi structured and unstructured data such as messages, updates, images, readings from sensors, GPS signals, and also from sources such as web pages, web log files, social media sites, e-mail, documents, cell phones, sensors etc. All this data is diverse and multi-modal, made up of a mix of numeric, categorical, text, images as well as multimedia. (Katal et al., 2013; McAfee et al., 2012) .

In addition to the three 'V's, further 'V's Veracity, Value and Variability have been added to expand the definition of big data.

2.3.4) Big Data Technology, Analytics and Tools

Due to the (above mentioned) properties of big data, structured databases which stored majority of corporate information until recently are ill suited to store and process big data. Traditional systems are also not sufficiently capable in performing the analytics on data which is constantly in motion (McAfee et al., 2012).

At the same time, the steadily declining costs of all the elements of computing such as storage, memory, processing, bandwidth, has resulted in previously expensive data-intensive approaches becoming quite viable and practical. As more and more business activities are digitized, new sources of information and ever-cheaper equipment combine to bring on a new era, where large amounts of digital information exist and are accessible on virtually any topic of interest to a business.

Today, many businesses are managed using operations support or business support systems (OSS/BSS) reliant on traditional database, data warehouse, and business intelligence tool sets. These technologies are typically applied to the data in each organizational silo and are configured to create reports and dashboards aimed at solving the business problems of the individual divisions. As the traditional tools are not scalable and cost effective for extremely large data sets, very often data is left unanalyzed, and data from multiple organizational units not correlated. To address these limitations technologies designed to handle data on a massive scale have emerged and are being called "big data technologies" (Alahakoon & Wijenayake, 2018; Schläfke et al., 2013).

Unlike the structured data that can be handled repeatedly through a Rational Database Management System (RDBMS), semi-structured and unstructured data which make up a large percentage of big data may call for *ad hoc* and one-time extraction, parsing, processing, indexing, and analytics requiring scalable and distributed environments. MapReduce has been hailed as a revolutionary new platform for such large scale, massively parallel data access. Built on MapReduce, Hadoop provides a Java based software framework for distributed processing of data intensive transformation and analytics and the top three commercial database suppliers: Oracle, IBM, and Microsoft, have all adopted Hadoop in their technology platforms (Chen et al., 2012).

Hadoop and MapReduce based systems have become another viable option for big data analytics in addition to the commercial systems developed for RDBMS, column-based Database Management Systems (DBMS), in-memory DBMS, and parallel DBMS (Chen et al., 2012).

The open source Apache Hadoop has also gained significant traction as a big data technology including Chukwa for data collection, HBase for distributed data storage, Hive for data summarization and *ad hoc* querying, and Mahout for data mining. Data analytics tools and algorithms utilize these big data platforms to manage and handle the requirements of big data (Alahakoon & Wijenayake, 2018).

Data analytics refers to the business intelligence and analytics (BI&A) technologies that are founded upon data mining, machine learning and statistical analysis. Most of these techniques rely on the mature commercial technologies of relational DBMS, data warehousing, ETL (Extract Transform Load), OLAP (Online Analytical Processing), and BPM (Business Process Management). Since the late 1980s, various data mining algorithms have been developed by researchers from the artificial intelligence, algorithm, and database communities. Some of the most widely accepted and used data mining algorithms are C4.5, k-means, SVM (support vector machine), Apriori, EM (expectation maximization), PageRank, AdaBoost, kNN (k-nearest neighbors), Naïve Bayes, and CART (Alahakoon & Wijenayake, 2018).

These algorithms cover classification, clustering, regression, association analysis, and network analysis. Most of these popular data mining algorithms have been incorporated in commercial and open source data mining systems. Other advances such as neural networks for classification/prediction and clustering and genetic algorithms for optimization and machine learning have all contributed to data mining in diverse applications (Alahakoon & Wijenayake, 2018; Chen et al., 2012).

Techniques such as Bayesian networks, Hidden Markov models, support vector machines, reinforcement learning, and ensemble models, have been applied to data, text, and web analytics applications. Other new data analytics techniques explore and leverage unique data characteristics, from sequential/temporal mining and spatial mining, to data mining for high-speed data streams and sensor data. Many of these methods are data-driven, relying on various anonymization techniques, while others are process-driven, defining how data can be accessed and used. The big data technologies and platforms such as MapReduce and Hadoop support the application of these algorithms on big data (Alahakoon & Wijenayake, 2018).

2.3.5) Public online discussion forums

This section presents the business value held by public online discussion forums as a heavily underutilized category of social media platforms by utilizing a case study.

The literature review revealed that rich, public big data sources such as public online discussion forums have not been leveraged to enhance organizational performance. Furthermore, when we looked into the case study domain; higher education sector, higher education institutes have not yet engaged with this type of social media. These sources possess much potential to empower agility of an organization and calls for traditional business processes to be adapted to the big data environment with machine learning techniques to enable smarter and conscious organizations. When considering the case study domain: higher education sector, the study revealed that several public forums are rich sources of public opinions and comparisons related to organizations (higher education institutes), products (courses) as well as organizational processes (university processes) in Australia.

Big data sources such as public forums are rich in information that can be used as feedback to continuously enhance organizational processes. For example, the selected forums are virtual spaces where customer groups post discussions, questions and answers, share experiences and opinions, compare products, brands and organizations.

When considering the higher education sector, customer groups of interest are potential students, current students, alumni, parents, staff, experts etc.

Public forums such as Whirlpool, ATAR notes and Bored of Studies have hundreds of thousands of posts from these stakeholder groups discussing resources and processes related to universities (higher education institutes) in Australia. They share requirements, opinions and experiences about courses and universities, discuss university facilities, well-functioning and defective processes they have come across etc.

I met an international student completing their Masters at Swinburne who was offered Masters at UoM, Monash, La Trobe and Swinburne, and chose Swinburne because they offered the lowest fees. It made me curious about the international fees for undergraduate places, and whether they can be used as long-term indicator for unregulated local fees in 2015. Only time can confirm/deny. I compared the 2014 fees for biomedical science. University of Melbourne Bachelor of Biomedicine Annual fee: $32,384-$34,600 AUD Monash University Bachelor of Biomedical Science Annual fee: $ 34,700 AUD RMIT University Bachelor of Biomedical Science Annual fee: $31,680.00 AUD Swinburne University Bachelor of Science (Bioscience) Annual fee: $23,950 AUD Victoria University Bachelor of Science (Biomedical Sciences) Annual fee: $21,160 AUD

Figure 2.2: A forum post from “Whirlpool” public online discussion forum

It's because the Go8 universities in Australia (along with several other unis like La Trobe) follow this calendar: https://www.universitiesaus tralia.edu.au/australias-universities/teaching-calendar Quote Yeah feeling you bro... Unfortunately we don't get a mid sem break! RMIT have their "mid semester break" this week lel

Figure 2.3: A forum post from “Whirlpool” public online discussion forum

There are many postgraduate physiotherapy schools around Australia. In Victoria, you would be looking at the Doctor of Physiotherapy at Melbourne University and the Masters of Physiotherapy at La Trobe. You need an anatomy and physiology subject at 2nd year to be eligible for Melbourne. For La Trobe, the same would be needed, but a Biomechanics subject would be required on top of that (can't complete that subject at Melbourne as they don't offer it. Doing a single subject at La Trobe to satisfy that prerequisite might be an option). There is a guaranteed entry pathway for Melbourne University - 94+ ATAR and a 75+ WAM would get you a full fee place. CSP places would be based on interview and the WAM (a competitive grade would be at least 70+). Due to the limited options I have in Victoria, I've also looked at other interstate options. University of Sydney, Macquarie University in NSW UQ in Brisbane, or Griffith (but they have an interview and many other annoying pre-requisites that would be more easily satisfied in an Exercise Science degree). Also Bond University, but they have an interview and are quite expensive, as you can imagine for a private uni. Lastly, there is Curtin in Perth. There you go, they are all Masters level degrees, that don't require a Bachelor degree in physio. They are designed for people like us who have completed an undergraduate science degree. All the best with your applications.

Figure 2.4: A forum post from "ATAR Notes" public online discussion forum

I spoke with latrobe today and they told me that if I have not been interviewed that I have been not been successful. They are in the process of notifying people about unsuccessful applications. I was told this by the psychology postgraduate admin coordinator. She said that they had 600 applications. May I ask who you spoke to at Latrobe as they seem to be giving different information to different people?

Figure 2.5: A forum post from "Whirlpool" public online discussion forum

Figures 1, 2, 3 and 4 demonstrate the type of information that is available in these forum posts. Hundreds of thousands of such posts holds a potential gold mine for big data and analytics tools and techniques.

A textual dataset was mined from public online discussion forums in order to extract the following highly valuable insights, which will enhance university processes and the overall student management system.

- Public opinions of the university and its various schools and courses over time and market segments.
- Students' impression and experiences related to university processes such as enrolment, exams, teaching, on campus accommodation, student services, library etc.

- The university’s standpoint compared to other universities - in terms of demand for various courses, smooth functioning of processes and resources.

This information extraction was achieved using machine learning and advanced analytics techniques. NLP techniques such as SentiStrength, SentiWordNet (Sentiment extraction algorithms and tools) was incorporated into the text dataset and information and relationships related to organizations, products and organizational processes in the case study domain: higher education was extracted.

Novel models, frameworks and techniques related to extraction and understanding of organizations and their processes are proposed and experimental results and evidence presented in the proceeding contribution chapters 3, 4, 5 and 6 using La Trobe University business processes as the case study.

2.3.6) Data Analytics in Organizations

The term “intelligence” has been used in Artificial Intelligence (AI) by researchers since the 1950s. However, “Business intelligence” (BI) became a popular term in the business and IT communities only in the 1990s. In the late 2000s, “business analytics” was introduced to represent the key analytical component in BI (T. Davenport, 2014). More recently the terms big data and “big data analytics” have been used to describe the aforementioned massive data sets and the analytical techniques used on them in applications (Alahakoon & Wijenayake, 2018; Russom, 2011).

According to Hal Varian, Chief Economist at Google and emeritus professor at the University of California and Berkeley, the opportunities associated with data and analysis in different organizations have generated a significant interest in business intelligence and analytics (BI&A). BI&A is often referred to as the techniques, technologies, systems, practices, methodologies, and applications that analyse critical business data to help an enterprise better understand its business and market and make timely business decisions. In addition to the underlying data processing and analytical technologies, BI&A includes business-centric practices and methodologies that can be applied to various high-impact applications such as e-commerce, market intelligence, e-government, healthcare, and security (Alahakoon & Wijenayake, 2018; Russom, 2011).

Most of the literature on data analytics focus on the application of analytics tools and techniques on various data sets. These highlight the data related issues as well as the technology and algorithmic aspects in data analytics. Comparatively very little has been published on how analytics could support management and decision making within an organization (Alahakoon & Wijenayake, 2018).

Organizations typically store their critical data arising from various transactions in relational databases. Such data is mostly structured data. However, there is a large amount of unstructured data generated from blogs, images, email, social media, scientific experiments, various surveys, etc. as mentioned above, that contain useful information. Mining these kinds of data is becoming extremely useful for making intelligent business decisions (Alahakoon & Wijenayake, 2018; Dhar & Mazumdar, 2014).

Performance Management Systems (PMS) and Management Control Systems (MCS) support management control and decision making in organizations. According to Ferreira et.al (Ferreiraa A, 2009) an effective and complete PMS should:

- Clarify the vision and mission of an organization and focus attention of managers and employees on this vision and mission;
- Identify the key success factors and clarify how these can be brought to the notice of managers and employees;
- Illustrate how the organizational structure affects PMC design and use;
- Highlight processes and activities which are required for the implementation of organizational strategies and plans;
- Identify the key performance measures;
- Identify the appropriate performance targets for the key performance measures;
- Identify the already existing performance evaluation processes;
- Set the rewards for target achievement; and
- Highlight the information flows that can support the performance management (Alahakoon & Wijenayake, 2018).

A PMS is focused on the measuring of performance, but this alone does not provide competitive edge to an organization (Schläfke et al., 2013). It is imperative that an organization understands the business dynamics and review strengths, weaknesses as well as opportunities and threats. A further important factor is to ensure that right data is available and also the skills and capability to transform such data into appropriate information.

As previously discussed, advanced analytics tools and techniques can be used to generate insights from diverse data sources. Therefore, by incorporating into PMS, these have the potential to extend the domain of performance management to provide an improved understanding of business dynamics and lead to better decision making. Use of analytics can extend performance management from purely focussing on financial performance drivers to the inclusion of non-financial aspects by bringing in techniques such as mathematics, statistics, econometrics, data mining, machine learning, information technology, and tools for data gathering and analysis (Alahakoon & Wijenayake, 2018).

Schläfke has called such extensive use of analytical techniques with PMS as performance management analytics (PMA) (Schläfke et al., 2013). PMA involves the extensive use of data and analytical methods to understand relevant business dynamics, to effectively control key performance drivers, and to actively increase organizational performance.

In order to effectively incorporate business analytics, organizations must satisfy a number of key requirements such as availability of data, appropriate IT infrastructure and related skills including business data analysis skills. An analytical management system's value is extremely high in an organization that already has an advanced IT infrastructure such as an enterprise resource planning (ERP) system, data warehouse, data mining systems, or well-established customer relationship management. Based on a model of applications of PMA proposed by (Schläfke et al., 2013), we propose a model for the key components of PMA as shown in figure 2.6. In this model performance management analytics is considered as the intersection between (a) organizational business processes and information systems (b) decision support systems and (c) analytical, data mining techniques and tools (Alahakoon & Wijenayake, 2018).

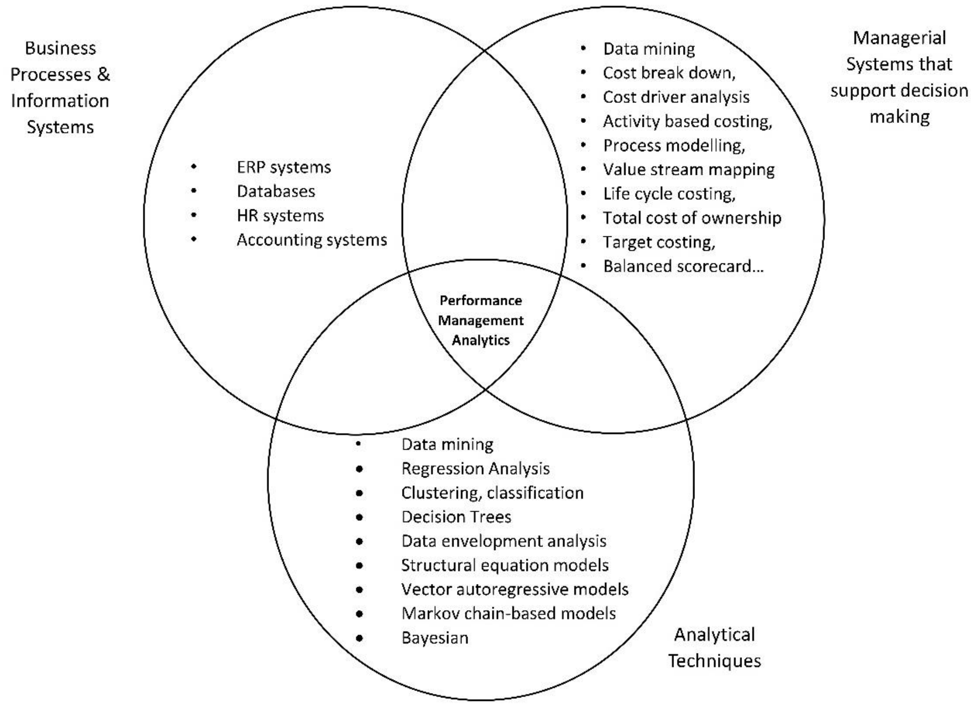


Figure 2.6: Key components of Organizational Performance Management Analytics (PMA) proposed by (Alahakoon & Wijenayake, 2018).(diagram courtesy of (Alahakoon & Wijenayake, 2018))

The design and implementation of an analytical performance management system require a thorough understanding of the organization and it’s business model, it’s performance indicators, it’s key success factors, information and data sources. It is also very important to be clear about the techniques which alert management about data, transactions and events therefore triggering suitable actions. It is important to build such a system with a solid foundation such as a good theoretical framework (Alahakoon & Wijenayake, 2018).

Schläfke et. al (Schläfke et al., 2013) have proposed a multi-layer performance management framework which will provide decision makers with additional information which could address the limitations of traditional PMS and enable the use of analytics. With the use of the four layers for capturing business drivers in inputs, processes, outputs and outcomes, coupling of performance drivers, identifying levers of control and designing such controls and finally internal and external communication of the performance drivers, Performance Management and Control (PMC) is achieved.

PMC systems thus gathers and uses information to evaluate the performance of multiple organizational resources such as human, physical, financial and also organization as a whole in terms of the organizational strategies. This also include control systems such as tools for steering an organization toward its strategic objectives to realise competitive advantage (Alahakoon & Wijenayake, 2018).

2.3.7) Evolution and Maturity of Organizational Performance Management and Control Systems

The information rich big data environment has resulted in the generation of new types of data different and diverse from the traditional data used in organizations. Since this non-traditional data accounts for 80% of enterprise data and continue to grow rapidly, enterprises have realised the need for a strategy to combine both traditional and non-traditional data together to support the organizational PMC processes. As discussed in the previous section, the new discipline known as performance management analytics (PMA) has arisen in the attempt to cater to this need (Alahakoon & Wijenayake, 2018). However, these attempts are still in infancy in their path towards achieving PMA on an integrated enterprise wide data management platform (Dhar & Mazumdar, 2014). To appreciate the need and value of a PMA and its role in an organization, it is necessary to have a good understanding of the diverse types of data that contribute to the big data environments in organizations. These are considered as the information assets of an organization. It is also important to understand the organizational processes which generate and capture this data.

Organizations rely on a growing set of applications to communicate with and provide services/products to today's demanding consumer and business communities:

- They are collecting, storing, and analyzing more granular information about more products, people, and transactions than ever before.
- They rely on email, social media, collaboration tools, and mobile devices to communicate and conduct business with customers and business partners.
- They are creating, receiving, and collecting machine and sensor-generated messages, sometimes at very high volumes, and are driving operations and business processes from that message data

  (Alahakoon & Wijenayake, 2018).

The growth in the number, size, and importance of information assets is not limited to just large government agencies, large enterprises, or Internet web sites. A wide variety of organizations, ranging from small and medium-sized businesses (SMBs) to large enterprise and government agencies, are dealing with a flood of data as they and their customers (Villars et al., 2011):

- Digitize business records and personal content (including the generation of ever-larger numbers of photos, movies, and medical images) driven by continued advancements in device features, resolution, and processor power
- Instrument devices (e.g., set-top boxes, game systems, smartphones, smart meters), buildings, cities, and even entire regions to monitor changes in load, temperatures, locations, traffic patterns, and behaviors
- Address governance, privacy, and regulatory compliance requirements that complicate the retention of business information

  (Alahakoon & Wijenayake, 2018)

To cater to this data intensive environment a new breed of analytics systems as well as algorithms and tools have evolved and been created. Figure 2.7 shows the evolution of the business intelligence and analytics systems in organizations in the first column, the data types in column two and algorithms and techniques in column three (Alahakoon & Wijenayake, 2018).

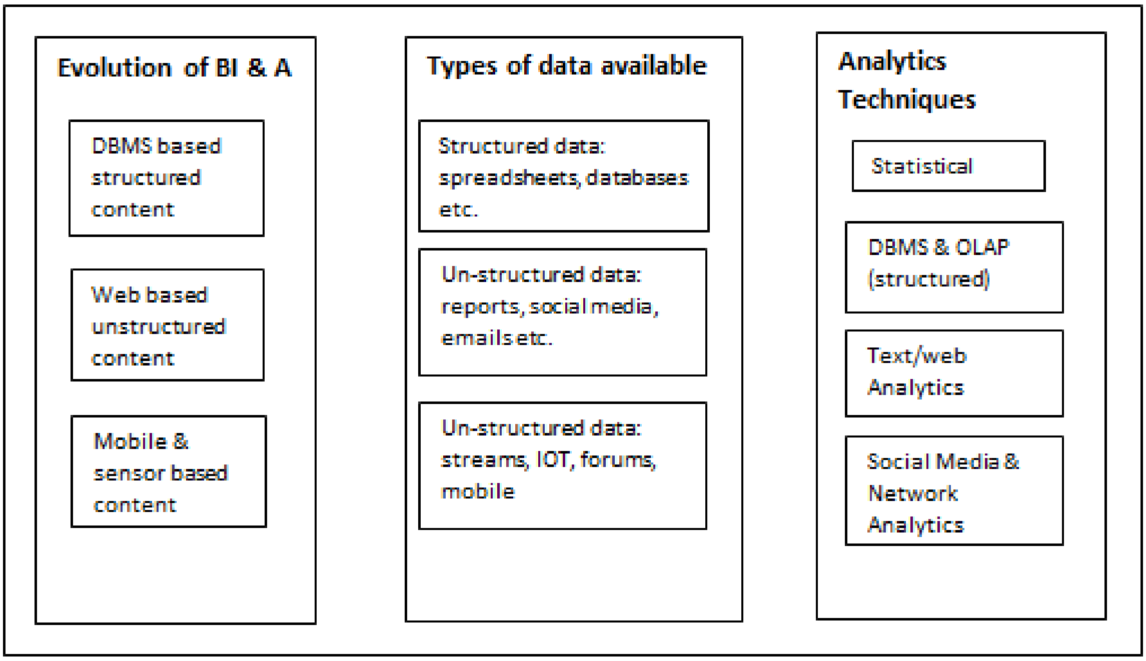


Figure 2.7: Evolution of analytics, types of data and techniques (Image courtesy of (Alahakoon & Wijenayake, 2018))

Business intelligence in the past was mainly derived from data base management based structured data stores. Reporting and querying (such as SQL) enabled the retrieval of the required information and transforming into end user requested formats. As described above, the past decade has evolved into the big data environment with a flood of unstructured data as well as the wide use of mobile devices. To cater to the need of this environment new analytics tools and techniques have evolved such as text and web analytics and social media and network analytics (Alahakoon & Wijenayake, 2018).

A strategy to understand and also capture the impact of this change and evolution in the data intensive environment on performance management and control is to observe the levels of maturity in such systems in organizations. (Patas, 2015) has discussed such maturity levels highlighting the levels of maturity reached based on IT enabled capabilities. The different levels of maturity have been detailed from the perspectives of corporate planning and corporate reporting. Table 2.2 is based on and extends the maturity levels from Patas J (Alahakoon & Wijenayake, 2018; Patas, 2015).

Maturity Level descriptor	**Corporate planning capability**	**Corporate reporting capability**
Basic	Financially oriented and manually done using spreadsheets	Highly manual, Paper based reports oriented towards financial measures and external requirements
Guided	Supported by basic IT systems	Included reporting and analysis services for corporate and business units
Integrated	Well organized financially oriented planning systems	Well designed, automated reporting with risk and compliance measures and advanced analytics.
Strategy-driven	Supported by Advanced IT- for organizational vision and strategy	Reporting with emphasis on strategy measures, analysis and instruments.
IT- advanced	Supported by modern IT for process optimization, planning integration and enhanced quality	Reporting utilizes modern IT enabled systems with mobile devices and dashboards
Data Driven and big data (Current and near future)	**Supported by advanced analytics tools and technologies incorporating structured and unstructured data from internal as well as external environments**	**Further extending reporting capabilities in IT-advanced stage with insights derived from data using descriptive and predictive analytics. Better visualizations and visual analytics**
Mobile, web based and IOT era (Next generation)	**Ability to work with highly granular information, supported on web based and mobile devices and cloud based technology and applications**	**Wide use of mobile and web based systems, ability to drill down to highly granular levels due to availability of such detailed data. Personalised reporting and ability to link with wider external systems and data sources.**

Table 2.2: Evolution of corporate planning and reporting capabilities for Management Control systems. (Table courtesy of (Alahakoon & Wijenayake, 2018))

The two new maturity levels are proposed to cater to the current and future of PMC systems. In the data driven stage, the previous IT enabled systems are extended and harnessed with advanced analytics techniques as well as the capabilities of handling unstructured and large internal and external data sources. For reporting, more derived insights rather than direct retrieved data and also much improved visualization capabilities are used.

In the future, level of maturity, the mobile environment and the wider usage and availability of IOT systems and data will be the key defining factor. This will generate and make available highly granular data and the systems and algorithms will have to cater to these requirements. The reporting also reflects these changes and also will have to cater to the technical needs of linking with external data sources for enriching internal data.

It can be seen that the data driven level is mostly an evolution due to the availability of new data and the advancement of data handling and analytics technologies. The next generation level is mainly due to the acceptance in society of the mobile life styles and also the advent of IOT (Alahakoon & Wijenayake, 2018).

2.3.8) Impact of Big Data and Analytics on Performance Management Control Systems

According to (McAfee et al., 2012) the use of big data and analytics is not only applicable to companies *born digital* such as Google and Amazon, but has the potential to transform traditional businesses as well and provide competitive advantages. Due to analytics enhanced performance management and control, managers now have the capacity to make better predictions and smarter decisions, carry out more-effective interventions based on data and information rather than intuition (Alahakoon & Wijenayake, 2018).

Big data and analytics can impact the PMC of an organization in diverse ways as discussed in (T. H. Davenport & Dyché, 2013). For the ease of presentation and understanding we have grouped these under three main key categories as:

(a) Impact on organizational systems and processes

(b) Impact in improving existing tasks and workflows

(c) Impact on organizational decision-making culture

Impact on organizational systems and processes

There are several ways big data can have an impact on changing organizational systems and processes. By making information more accessible to stakeholders in a timely manner the organization creates higher transparency generating high value. Using automated algorithms to support human decision making also impacts systems and processes. In some cases, decisions will not necessarily be automated but augmented by analysing huge, entire datasets using big data techniques and technologies rather than just smaller samples. New innovative business models, products and services are enabled due to big data. Insights gleaned from the use of products or services can be used for improvements for the innovation of new products (Alahakoon & Wijenayake, 2018).

Impact in improving existing tasks and workflows

Use of big data can substantially improve the time required to perform a computing task, or new product and service offerings, reduce the cycle time for complex and large-scale analytical calculations (T. H. Davenport & Dyché, 2013). Big data can also be a valuable tool for analysis of individual or team behavior, using sensors or badges to track individuals as they work together, move around their workspace, or spend time interacting with others or allocated to specific tasks ((George 2014) as cited in (Alahakoon & Wijenayake, 2018)) .

Organizations can collect more accurate and detailed performance data (in real or near real time) on everything from product inventories to personnel sick days and use such data to analyze variability in performance and to understand its root causes; can enable leaders to manage performance to higher levels ((Manyika J, 2011) as cited in (Alahakoon & Wijenayake, 2018)).

Big data also allows organizations to create highly specific segmentations and to tailor products and services precisely to meet those needs. This approach is well known in marketing and risk management. Consumer goods and service companies that regularly use segmentation for many years have begun to deploy ever more sophisticated big data techniques such as the real-time micro-segmentation of customers to target promotions and advertising ((Manyika J, 2011) as cited in (Alahakoon & Wijenayake, 2018)).

Creating a new culture of decision making

Managerial challenges are greater than the technical challenges when implementing a big data strategy (McAfee et al., 2012). When there is a scarcity of data, data is expensive to obtain, or not available in digital form, it makes sense to let well-placed people make decisions, which they do on the basis of experience they've built up and patterns and relationships they've observed and internalized. According (McAfee et al., 2012) executives interested in leading a big data transition should start with two simple techniques. *First, they can get in the habit of asking* "What do the data say?" when faced with an important decision and following up with more-specific questions such as "Where did the data come from?," "What kinds of analyses were conducted?," and "How confident are we in the results?" *Second, managers or decision makers can allow themselves to be overruled by the data*; few things are more powerful for changing a decision-making culture than seeing a senior executive concede when data have disproved a hunch. This is an essential step in changing the culture towards a data driven organization (Alahakoon & Wijenayake, 2018).

2.3.9) Big Data adoption: Challenges and Solutions

As described in the previous sections, although big data bring about features to handle complex requirements of today's organizations towards a data driven PMC, there are still many challenges in adoption of big data at enterprise level (Russom, 2011). In this section the key challenges for adoption of big data technologies and strategies to overcome them are discussed (Alahakoon & Wijenayake, 2018).

Management Challenges for Big Data

Organizations will not reap the full benefits of a transition to using big data unless they're able to manage change effectively. Key management challenges as per (McAfee et al., 2012) are summarized below:

Leadership: Companies succeed in the big data era not simply because they have more or better data, but because they have leadership teams that set clear goals, define what success looks like, and ask the right questions. The power of big data does not reduce the need for vision or human insight. On the contrary, it needs business leaders who can spot opportunity, understand how market develops, think creatively and propose truly novel offerings, articulate a compelling vision, persuade people to embrace it and work hard to realize it, and deal effectively with customers, employees, stockholders, and other stakeholders. Successful companies of the next decade will be the ones whose leaders can do all that while changing the way their organizations make many decisions (McAfee et al., 2012); (Alahakoon & Wijenayake, 2018).

Skill management: As data become cheaper, the complements which are required to generate value from data become more valuable. Some of the most crucial of these are data scientists and other professionals skilled at working with large quantities of information. Skilled IT professionals are also essential to work beside the data scientists to handle the challenges of working with very large data sets. These skills should not be limited to technical ones but also should extend to research, analytical, interpretive and creative skills. Furthermore, these skills need to be developed in individuals hence requires training programs to be held by the organizations. People with these skill sets are in short supply and great in demand (Alahakoon & Wijenayake, 2018; McAfee et al., 2012; Villars et al., 2011).

Technology: The tools available to handle the volume, velocity, and variety of big data have improved greatly in recent years. These technologies are quite affordable to even smaller organizations and most of the software is now open source. However, these technologies do require a specialized skill set that is new to most IT departments, which will need to work hard to integrate all the relevant internal and external sources of data. As such it is essential that strong focus with high level support for keeping up with the technology is set up as necessary component of a big data strategy (Alahakoon & Wijenayake, 2018; McAfee et al., 2012).

Other Big Data challenges faced by Organizations

Privacy and Security

An extremely important issue with big data which is sensitive and includes conceptual, technical as well as legal significance. It is now possible to link the personal information of a person with external large data sets leading to the inference of new facts about that person. It is possible that these kinds of facts about the person are confidential and the person might not want the data owner to know about them. Information regarding the users (people) is collected and used in order to add value to the business of the organization. This is done by creating insights in their lives which they are unaware of thus building up the capacity to advise, guide and support (Alahakoon & Wijenayake, 2018; Katal et al., 2013).

Data Access and Sharing of Information

If data is to be used to make accurate decisions in time, it becomes necessary that it should be available in accurate, complete and timely manner. This makes the data management and governance process a bit complex adding the necessity to make data open and make it available to government agencies in standardized manner with standardized APIs, metadata and formats. This leads to better decision making, business intelligence and productivity improvements (Katal et al., 2013). Careful thought has to be put into the setting up of organizational data policy (Alahakoon & Wijenayake, 2018).

Quality of Data

Availability of larger data sets over longer periods of time for analysis will provide better and more accurate results. Such data when used for decision making or for predictive analysis in business will definitely lead to better results. Business Leaders will always want more and more data storage whereas the IT Leaders will take all technical aspects in mind before storing all the data. Big data focuses on quality capture and data storage rather than having very large irrelevant low-quality data so that better results and conclusions can be drawn. This further leads to various questions like how it can be ensured that which data is relevant, how much data would be enough for decision making and whether the stored data is accurate or not to draw conclusions from it etc. (Alahakoon & Wijenayake, 2018; Katal et al., 2013).

Development Scalability and Maintainability

Many issues such as Lack of IDEs, Testing, Deployment and Administration tools (which suit the big data scale) make the development phase of big data slow and also create difficulties in maintenance. Big data teams require a range of skills in developers ranging from application logic, data modeling as well as infrastructure administration proficiency. Availability of proper sets of tools in these areas could help enterprises developing big data applications faster and also maintaining the same. Integration of existing popular tools (like Eclipse) with big data technology perspective is also taking place (Alahakoon & Wijenayake, 2018; Dhar & Mazumdar, 2014).

Reusability

For the adoption to big data, proper data modeling and unified big data architecture across structured and unstructured data sources should be available. The massive volume, velocity and variety of the big data may result in the difficulty of a human to comprehend the complete picture of the project or operation which can result in the development of non-reusable solutions (Alahakoon & Wijenayake, 2018).

2.3.10) The Honeycomb social media framework

The rise of social media has resulted in a democratization of corporate communication. As per Kietzmann, "The power has been taken from those in marketing and public relations by the individuals and communities that create, share, and consume blogs, tweets, Facebook entries, movies, pictures, and so forth. Communication about brands happen, with or without permission of the firms in question“ (Kietzmann et al., 2011).

Even though it is clear that social media hold immense power, many executives are reluctant or unable to develop strategies and allocate resources to engage effectively with social media resulting in organizations mismanaging the opportunities and threats presented by this big data environment (Kietzmann et al., 2011). As per Kietzmann, one of the key reasons behind this ineptness is a lack of understanding regarding what social media are and the varied forms they take (Kietzmann et al., 2011). To address this knowledge gap a honeycomb framework of seven social media building blocks were proposed by (Kietzmann et al., 2011). These blocks can be utilized together or individually by organizations to make sense of the social media ecology, and to understand their audience and their engagement needs (Kietzmann et al., 2011).

The present big data environment is home to a rich and diverse ecology of social media sites, which vary in terms of their scope and functionality (Kietzmann et al., 2011). Therefore, before using any automation technique on social media data, first a decision has to be made on which functionalities of the social media should be considered to engage with for the task at hand.

In order to compare and contrast the functionalities and implications of different social media activities in online communities, the honeycomb social media framework (Kietzmann et al., 2011) was selected and has been applied throughout this thesis.

According to (Kietzmann et al., 2012) the social media honeycomb model provides an analytical lens for firms' specific ‘community needs’, and can educate the design or use of an appropriate social media platform. This framework introduces a honeycomb of seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation, and groups. The blocks allow to unpack and examine different facets of the social media user experience, and their implications for organizations. The building blocks help to understand how different levels of social media functionality can be configured (Kietzmann et al., 2011). According to this framework each social media platform is driven by primary, secondary and tertiary building blocks, which provide the foundation for important social media design decisions (Kietzmann et al., 2012).

The honeycomb framework for social media proposed by (Kietzmann et al., 2011) is presented in figure 2.8. According to Kietzmann et al., each block allows unpacking and examination of: "(1) a specific facet of social media user experience, and (2) its implications for firms. The blocks are neither mutually exclusive, nor do they all have to be present in a social media activity. They are constructs that allow to make sense of how different levels of social media functionality can be configured" (Kietzmann et al., 2011).

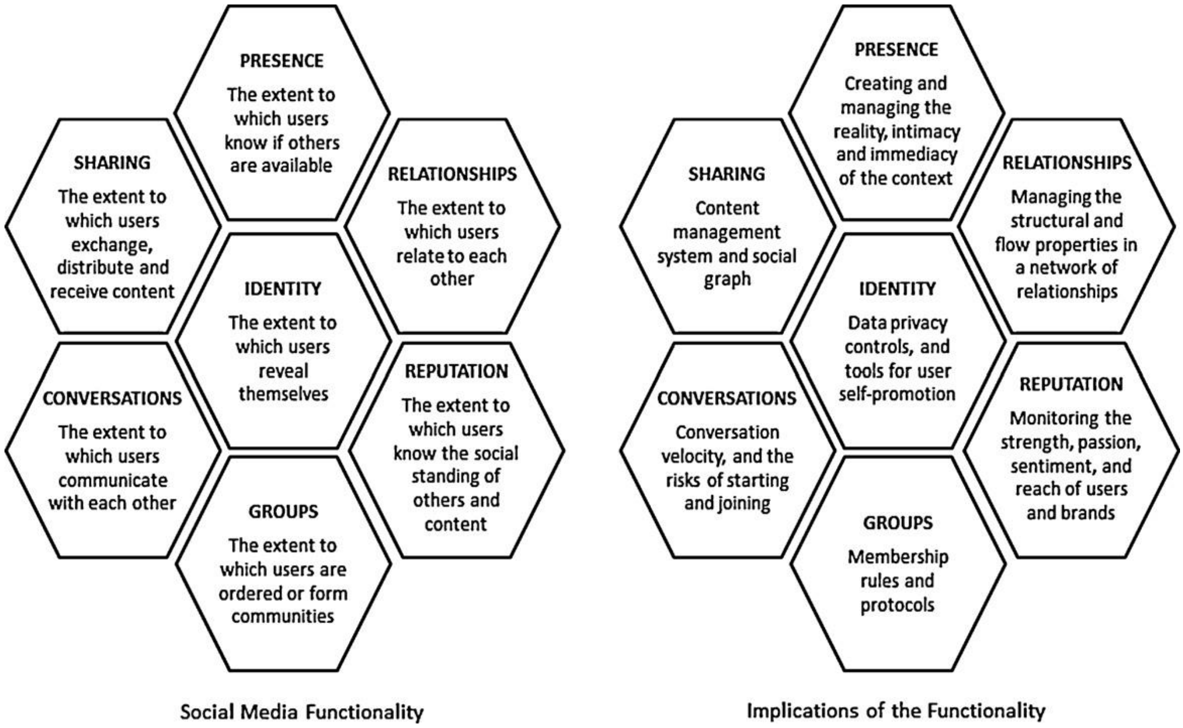


Figure 2.8: The honeycomb of social media by (Kietzmann et al., 2011). The social media framework of seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation and groups. (Image courtesy of (Kietzmann et al., 2011).)

After an intensive investigation into the functionality and data of the selected social media sites, the honeycomb framework was applied to identify the areas of the selected social media (online communities) to utilize for the extraction of the dataset to conduct experiments for the research presented in this thesis. Furthermore, this framework is discussed in more detail throughout the following contribution chapters where novel applications of this framework is proposed and demonstrated.

2.4) An overview of Machine Learning techniques

This section introduces *machine learning* in detail and explores the techniques that have been used (and currently being used) by business organizations to understand their stakeholders and stakeholders' impressions about organizations. The section focuses specifically on techniques of social media analytics relevant to the research questions of this thesis, used by organizations. This section will introduce some key terminologies in machine learning used in the thesis and the relevant techniques and literature.

2.4.1) Machine learning techniques

Machine learning is the study of computational methods for automation of knowledge acquisition from examples (Bose & Mahapatra, 2001). It's a discipline that has evolved to eliminate the laborious and expensive knowledge engineering process involved in developing knowledge-based systems. A commonly used strategy is discovery of patterns in training data sets (Bose & Mahapatra, 2001). These patterns are then used to classify and/or predict the behaviour of new examples (Bose & Mahapatra, 2001).

Today , machine learning has become one of the world's most rapidly growing technical fields lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science (Jordan & Mitchell, 2015). According to Jordan and Mitchell, recent progress in machine learning has been driven both by the development of new learning algorithms and theory, and by the ongoing explosion in the availability of online data and low-cost computation.

Currently, organizations are exploring big data to discover information they didn't know before which is extremely important in the current pandemic situation as studies such as (Russom & others, 2011) have shown how the previous economic recessions have forced deep changes into business organizations (Russom & others, 2011). In this highly volatile business environment machine learning is increasingly becoming a vital tool for organizations as it has the capacity to progress advanced analytics to study big data to understand the current state of the business and track still-evolving aspects such as customer behaviour (Russom & others, 2011).
According to (Schaeffer & Rodriguez Sanchez, 2020), companies are now storing practically all data on any client transaction making machine learning an essential tool to turn this data into information that can be used to drive business decisions.

The following paragraphs present a summarization of machine learning techniques that have been used by business organizations to support their decision-making process.

According to, Bose and Mahapatra, the behaviour of machine learning techniques vary with variations in data and operational characteristics of data mining applications. A review on machine learning techniques in business conducted by Bose and Mahapatra presented a summarized list of applications by functional area, identifying the techniques used and the type of data mining problem addressed. The data mining applications related to business were identified by reviewing a large number of IS and computer science journals and conference proceedings related to data mining (Bose & Mahapatra, 2001).

2.4.2) Text mining

As this thesis is an exploration into the utilization and adoption of social media data for organizational decision-making, unstructured text data was encountered to be the key challenge. Hence, this section attempts to summarize machine learning techniques utilized in text mining to generate knowledge from unstructured text data found in social media.

Research studies (Farnadi et al., 2016), (J. A. Golbeck, 2016) and (Sewwandi et al., 2017) establish how a person's language usage online can be used to predict characteristics such as intelligence and education level. This is further discussed in chapter 3, section 3.2.2.

According to (Kumar & Sangwan, 2019), the most popular machine learning techniques used for rumour detection using social media includes SVM, Naïve Bayes, and Decision Trees. Also, techniques such as k-Nearest Neighbor, Clustering, Bernoulli Naïve Bayes, Random Forests, Logistic Naïve Bayes, Natural Language Processing (NLP), social spam analytics and detection framework (SPADE), SEIZ, etc. have also been used (Kumar & Sangwan, 2019).

2.4.3) Natural Language Processing

Natural language processing (NLP) is the term used for the collection of computational techniques for automatic analysis and representation of human languages, motivated by theory (Chowdhary, 2020).

Yet, according to Chowdhary, the automatic analysis of text, at par with humans, requires a far deeper understanding of natural language by machines, which is still far from reality. Many examples of NLP, such as, online information retrieval, aggregation, and question-answering, have been mainly based on algorithms relying on the textual representation of web pages, as well NLP to some extent. Such algorithms are very good at retrieving texts (IR), splitting it into parts, checking the spellings, and performing word-level analysis, but not successful for analysis at sentence and paragraph level (Chowdhary, 2020).

Hence, when it comes to the question of interpreting sentences and extracting meaningful information, according to Chowdhary, the capabilities of these algorithms are still very limited (Chowdhary, 2020).

Available toolkits for NLP

There are varied tools to perform NLP tasks, some of which are integrated in powerful toolkits. Table 2.3 is an overview of commonly used toolkits presented by (Sun et al., 2017).

Available toolkits for NLP.

Toolkit	Language	Description
NLTK	Python	Natural Language Toolkit (NLTK) is an open source platform for performing NLP tasks including tokenization, stemming, POS tagging, parsing, and semantic reasoning. It provides interfaces for many corpora and lexicons which are useful for opinion mining and sentiment analysis. http://www.nltk.org/
OpenNLP	JAVA	The Apache OpenNLP is a JAVA library for the processing of natural language texts, which supports common tasks including tokenization, sentence segmentation, POS tagging, named entity recognition, parsing, and coreference resolution. https://opennlp.apache.org
CoreNLP	JAVA	Stanford CoreNLP is a framework which supports not only basic NLP task, such as POS tagging, named entity recognization, parsing, coreference resolution, but also advanced sentiment analysis . http://stanfordnlp.github.io/CoreNLP/
Gensim	Python	Gensim is an open source library for topic modeling which includes online Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Random Projection, Hierarchical Dirichlet Process and word2vec. All implemented algorithms support large scale corpora. LSA and LDA have distributed parallel versions. http://radimrehurek.com/gensim/
FudanNLP	JAVA	FudanNLP is an open source toolkit for Chinese NLP, which supports word segmentation, POS tagging, named entity recognition, dependency parsing, coreference resolution and so on. https://code.google.com/archive/p/fudannlp/
LTP	C++/Python	The Language Technology Platform (LTP) is an open source NLP system for Chinese, including lexical analysis (word segmentation, POS tagging, named entity recognition), syntactic parsing and semantic parsing (word sense disambiguation, semantic role labeling) modules. http://www.ltp-cloud.com/intro/en/
NiuParser	C++	NiuParser is a Chinese Syntactic and Semantic Analysis Toolkit, which supports word segmentation, POS tagging, named entity recognition, constituent parsing, dependency parsing and semantic role labeling. http://www.niuparser.com/index.en.html

Table 2.3: An overview of available toolkits for NLP presented in (Sun et al., 2017)

NLTK, OpenNLP and Stanford CoreNLP are extensively used general purpose NLP toolkits that support most of basic NLP tasks, such as POS tagging, named entity recognition, parsing where as NLTK provides industrial-strength NLP libraries which provides implementations of various techniques for each basic NLP task (Sun et al., 2017). Hence for the research presented in this thesis, mainly NLTK as well as Gensim and OpenNLP have been used.

2.4.4) Deep learning

According to Skansi, machine learning is divided into three main branches: supervised learning, unsupervised learning and reinforcement learning. Deep learning can be considered as a special approach in machine learning which covers all three branches and seeks also to extend them to address other problems in artificial intelligence which are not usually included in machine learning such as knowledge representation, reasoning, planning, etc. (Skansi, 2018). Deep learning is discussed further in chapter 6 of this thesis.

Deep learning is a representation learning method that employ multiple processing layers to learn hierarchical representations of data, and have produced state-of-the-art results in many domains (Young et al., 2018). Deep learning has revolutionized the field of natural language processing (NLP) and recently, a wide variety of model designs and methods have blossomed in the context of NLP (Young et al., 2018).

Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are the two main types of Deep Neural Network (DNN) architectures, widely used to handle various

NLP tasks (Yin et al., 2017). CNN is supposed to be good at extracting position invariant features and RNN at modelling units in sequence. The state-of-the-art on many NLP tasks often switches due to the battle of CNNs and RNNs.

CNNs are shown to be effective at extracting local and position-invariant features (Yin et al., 2017). According to Yin et al, several studies show that RNNs such as GRU and Long Short-Term Memory (LSTM) networks perform well in sequential learning tasks and overcome the problems of vanishing and explosion of gradients in traditional RNNs when learning long-term dependencies (Yin et al., 2017).

Need for Recurrent Neural Networks (RNNs)

In the context of this thesis RNNs draw significant interest as it performs sequential processing by modelling units in sequence and has the ability to capture the inherent sequential nature present in language, where units are characters, words or even sentences. Words in a language develop their semantical meaning based on the previous words in the sentence (Young et al., 2018).

RNNs are tailor-made to model context dependencies in language and similar sequence modelling tasks, which according to Young et al. turned to be a strong motivation for researchers to use RNNs over CNNs in these areas. Also, RNN's ability to model variable length of text, including very long sentences, paragraphs and even documents aids in its suitability for sequence modelling tasks (Young et al., 2018). Furthermore, one of the selling points for major works using RNNs has been that unlike CNNs, RNNs have flexible computational steps that provide better modelling competence and create the possibility to capture unbounded context (Young et al., 2018). In the context of NLP, RNNs are based on the Elman network (Elman, 1990) and originally consist of three layers (Young et al., 2018). Figure 2.9 from Young et al. illustrate a general RNN unfolded across time to accommodate an entire sequence. In the figure 2.9 , "x_t is taken as the input to the network at time step t and s_t represents the hidden state at the same time step" (Young et al., 2018). Calculation of s_t is based on the following equation (Young et al., 2018).

$$\mathbf{s_t} = f\,(U\mathbf{x_t} + W\mathbf{s_{t-1}})$$

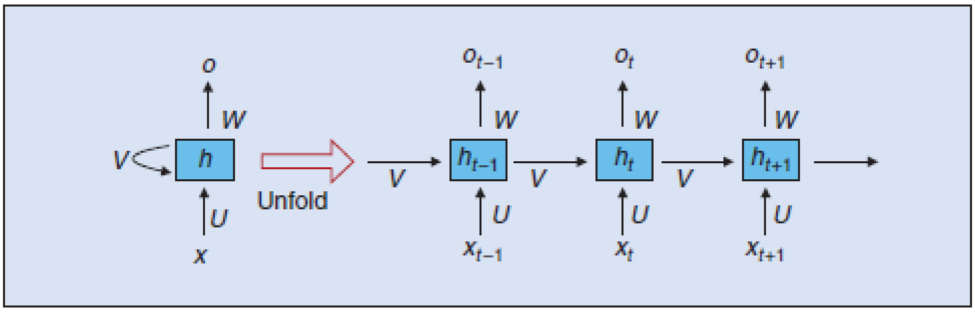


Figure 2.9: A simple RNN network (Image courtesy of (Young et al., 2018))

"s_t is calculated based on the current input and the previous time step's hidden state. The function f presents a non-linear transformation such as *tanh*, *ReLU* and *U,V,W* represent weights that are shared across time. In the context of NLP, x_t generally comprises of one-hot encodings or embeddings. They can also be abstract representations of textual content. o_t illustrates the output of the network which is also often subjected to non-linearity, especially when the network contains further layers downstream" (Young et al., 2018).

The hidden state of the RNN is considered the most important element as it is the network's memory component that accumulate information from other time steps (Young et al., 2018). In application, these simple RNNs suffer from the vanishing gradient problem making it difficult to learn and tune parameters of previous layers in the network. This shortcoming was overcome by networks such as LSTM (Young et al., 2018).

Long Short-Term Memory (LSTM)

When considering LSTMs (a special type of RNN), it has additional "forget" gates over the simple RNN, which allows errors to back-propagate through an unlimited number of time steps (Young et al., 2018). It consists of three gates: input, forget and output gates, which calculates the hidden state by taking a combination of these three gates as per the LSTM equations below (Young et al., 2018).

$$\mathbf{x} = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$

$$f_t = \sigma(W_f . \mathbf{x} + b_f)$$

$$i_t = \sigma(W_i . \mathbf{x} + b_i)$$

$$o_t = \sigma(W_o . \mathbf{x} + b_o)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c . \mathbf{x} + b_c)$$

$$h_t = o_t \odot \tanh(c_t) .$$

2.4.5) Sentiment Analysis

Sentiment analysis (also known as opinion mining) was first proposed in early this century and has gradually become an active research area (Sun et al., 2017). Sentiment analysis is the field of study that examine humans' opinions, sentiments, appraisals, evaluations, emotions and attitudes towards entities for example services, products, individuals, organizations, issues, topics, events, and the attributes (Soong et al., 2019).

Various practical applications of sentiment analysis, such as product pricing, competitive intelligence, market prediction, election forecasting, nation relationship analysis, and risk detection in banking systems have drawn extensive attentions from industrial communities (Sun et al., 2017). Furthermore, the boom of social media, e-commerce and online review sites, such as *Twitter, Amazon*, and *Yelp*, provides a large amount of corpora which are crucial resources for research and have motivated both academia as well as industry to promote the development of sentiment analysis (Sun et al., 2017).

Sentiment analysis of social media text is the science of extracting the sentiment information embedded in messages posted on social media websites (Z. Li et al., 2019). Sentiment analysis for social networks is more challenging because of the unique characteristics of social media data (Z. Li et al., 2019). The length limitation and the informal nature of the textual content present in social media exert varied difficulties and challenges (Z. Li et al., 2019).

Traditional textual sentiment analysis has been applied at three levels; document level, sentence level, and entity level (Z. Li et al., 2019). As per Li et al, existing sentiment analysis techniques for textual social media can be divided into five categories: Machine Learning, Lexicon-Based, Hybrid, Graph-Based and Others (Z. Li et al., 2019).

The benefits of sentiment analysis according to Soong et al. are; from the commercial perspective: recommendation and opinion about a product can be determined for the merchant and customer, and from the political perspective: political members can use to generate political information to win the election (Soong et al., 2019).

According to *(Soong et al., 2019),* Sentiment analysis can be further classified into two categories:

- *Lexicon Analysis*: “aims to calculate the polarity of a document from the semantic orientation of words or phrases within the documents. Nevertheless, applications refers to lexicon analysis and it does not reflect to consider the studied context”(Soong et al., 2019).

- *Machine Learning*: “encompasses building models derived from labelled training dataset (sentences or instances of texts) in order to find the document orientation. Studies that apply to this type of methods have been executed on an exact topic”(Soong et al., 2019).

The application of natural language processing by sentiment analysis/opinion mining is to collect and examine the sentiment words and opinions and then discovery of subjective attitudes in the large social data; has become a popular area in the field of NLP and data mining (Soong et al., 2019).

2.4.6) Word embedding

One of the recent interesting trends in NLP is word embedding. The aim of word embedding is to build a low dimensional vector representation of word from a corpus of text (Naili et al., 2017). According to Naili et al., the main advantage of word embedding is that it offers a more expressive and efficient representation by maintaining the contextual similarity of words and by building low dimensional vectors. The two methods most widely used and reported to be most efficient at learning vector representations of words are the word embedding models Word2Vec and GloVe (Naili et al., 2017).

Word embeddings follow the distributional hypothesis, according to which words with similar meanings tend to occur in similar context (Young et al., 2018). Hence, these vectors attempt to capture the characteristics of the neighbours of a word (Young et al., 2018). The main advantage of these distributional vectors is that they capture similarity between words, making measuring similarity between vectors possible using measures such as cosine similarity (Young et al., 2018).

Word embedding models are frequently used as the first data processing layer in a deep learning model. Typically, word embeddings are pre-trained by optimizing an auxiliary objective in a large unlabelled corpus, such as predicting a word based on its context where the learned word vectors can capture general syntactical and semantic information (Young et al., 2018). Hence according to Young, word embeddings have proven to be efficient in capturing context similarity, analogies and are fast and efficient in computing core NLP tasks due to its smaller dimensionality (Young et al., 2018).

2.5) Chapter summary

This chapter presented an assortment of concise discussions from the literature review conducted for the research presented in this thesis. The literature has been summarized and presented in light of addressing the research questions presented in the introduction chapter (section 1.6) and for the purpose of highlighting the research gap presented in section 1.4 in the introduction; the lack of comprehensive investigations and studies augmenting social media data with organizations and individual stakeholders delving deeply into the individuality of the social media users, together from their social aspect and the organizational aspect as stakeholders.

The literature review presented four main topic areas. Following is a summarization of these four major topics together with an indication of their relevance and continuation in the proceeding contribution chapters of the thesis.

1) *Individual as a digital stakeholder for organizations*; presented existing work, their implications and reliability on discovering human personality traits from social media and the importance of understanding stakeholders and their experiences for organizations in the contemporary business environment. This understanding will provide the foundation for the next contribution chapter (chapter 3) to create and establish a novel and comprehensive social media empowered Digital Stakeholder.
2) *Organization and it's image in the eye of an individual*; offered a glimpse into the existing research and theories on organizational impression management drawing on established concepts from marketing theory and exploring their

applications in current and past organizations. This discussion provides the foundation for the contribution chapters 4 and 5; to propose novel methods to perceive the organization from a social media perspective (continued in chapter 4) and to generate organizational brand from the opinions of the Digital Stakeholder using social media conversations (continued in chapter 5).

3) *Big data, social media and analytics*; presented a summary of existing literature and an extensive discussion and contribution on the current digital environment, it's untapped opportunities, properties of social media; specifically focusing on public online discussion forums which are highly promising and heavily underutilized by organizations for value creation. A summary and applications of frameworks, technologies and tools used by organizations for big data and social media analytics were presented. This compilation provides the indispensable technological foundation and credibility for the innovative social media based organizational solutions (in the form of novel models, techniques and frameworks) proposed in all four of the contribution chapters (chapters 3, 4, 5 and 6) utilizing the foundation laid by the honeycomb social media framework.
4) *An overview of machine learning techniques*; presented an overview of existing technologies and algorithms in the domain of machine learning for processing and extracting knowledge related to stakeholders and organizations from unstructured social media data with a specific focus of social media text data. This knowledge provides the corpora of tools to build and test the solutions which are based on social media analytics, proposed in all four of the contribution chapters.

These topics will continue their journey in discourse, laying groundwork and accumulating value and credibility in the proceeding contribution chapters; 3, 4, 5 and 6 to enable organizations to adapt underutilized social media data in the form of public online discussion forums into traditional organizational data sources to understand their stakeholders and their stakeholders' perception of the organization on a deeper level than was possible ever before; to serve their purpose of satisfying customers.

3. A Social Media empowered Digital Stakeholder

"Being human in the digital world is about building a digital world for humans."

~Andrew Keen

This chapter explores the extent to which a digital stakeholder can be established within the current digital environment to answer the research question: "*How can a holistic and generic framework of a digital stakeholder be developed from an organizational perspective using organizational data and the digital environment?*". The nature of human cognition and the contemporary organization is explored to establish the concept of the "digital stakeholder" from an organizational perspective. Social media features are explored to establish the same from a digital environment perspective. The chapter presents an intense investigation and research conducted to capture and model the changing preferences, sentiments, and behaviours of the digital stakeholder using a comprehensive; "holistic" conceptual model of the psychological and behavioural aspects of a human being which represents multiple facets (demographics, psychographics, personality traits, emotional tendencies, personal preferences etc.) of an individual to add deeper comprehension and meaning to the understanding of a stakeholder. The chapter introduces a novel, *holistic* and generic model of a digital stakeholder and a novel framework to apply this conceptual model to an organizational domain and generate organizational insights. This is achieved by augmenting with social media and incorporating knowledge captured through social media into information captured and stored in standard data bases and data warehouses of organizations.

An attempt is made to model this "digital stakeholder" as an inner person (psychological depiction) and an outer-person (behavioural depiction) using the combination of the proposed model (of the digital stakeholder) and the framework which maps the journey of the digital stakeholder through the organization, illustrating how this inner person and outer person relate to an organization. A novel matrix to map theory from psychology to features of social media is presented as a conceptual theoretical contribution.

Furthermore, the chapter presents how the proposed framework can be adopted to an organizational domain to construct an organizational representation (an Avatar) of a digital stakeholder and how this Avatar can be used to observe roles of a digital stakeholder (eg: a student, a customer, a client, a patient) in the organization on a deeper level than was possible earlier.

Additionally, the framework allows observing of the stakeholder's journey through the organization while facilitating monitoring and prediction of stakeholder behaviour and performance at different stages/milestones of the stakeholder life cycle, using internal data (organization generated and owned data) and external data (the digital environment: public data) available for the organization. Finally, a demonstration is made on how this generic model and framework can be adopted to a specific organization domain by using higher education domain as a case study and developing a holistic model of a university student as a digital stakeholder, and presenting the student journey using the introduced generic framework.

3.1) Towards establishing the digital stakeholder

As recognized in section 2.1 in the literature review, contemporary organizations are experiencing a paradigm shift from being product-centric to being customer-centric. In a customer centric business environment, the customer experience is key for survival of the organization (section 2.1.3 in the literature review). In order to continuously provide an enhanced customer experience, it is important to understand and observe the stakeholder as a holistic human being with changing needs, emotions and responses.

Customer experience is defined by Meyer and Schwager as the customer's cognitive, affective, emotional, social, and physical responses to an organization (Meyer et al., 2007). We selected this definition as the most appropriate theoretical backbone to create a definition for the digital stakeholder.

As established in section 2.3 in the literature review, the large flow of data of the current digital environment, specifically social media which captures many aspects of human thinking, emotions, behaviour and peculiarities holds the key for an organization to understand the cognitive, affective ,emotional, social and physical responses of their stakeholders toward the organization. In this digital environment the actors who impact a business organization are clearly not only the customers. For example, for universities; parents and other peers such as career experts and alumni play a vital role in the decision of whether a customer (in this example a student) will be created or retained. Hence, it is important for an organization to understand who their 'Stakeholders' are and what responses these varied individuals (who clearly have an impact on the organization) have towards the organization.

We define the cognitive, affective, emotional, social and physical understanding of a stakeholder generated through their digital footprint left behind in the aforementioned digital environment as a *digital stakeholder*.

Throughout literature attempts have been made by organizations and researchers to model varied aspects of stakeholders using digital footprints. However, as elaborated later on in this chapter, these studies have focused on modelling just a single aspect or a limited number of aspects of the individual's inner person or outer person that is considered relevant to the organization of interest. An attempt to design an all-inclusive holistic model bringing together all psychological and behavioural aspects of an individual has not been successfully attempted. Hence, to address this limitation in existing research, the following sections of this chapter presents the research endeavour carried out to model and understand this digital stakeholder in depth through a psychological and a behavioural depiction.

3.2) Digital stakeholder (Part 1)- Psychological depiction

The first phase of the research study was dedicated to understand and model the inner-person of this digital stakeholder.The inner-person is the composition of attributes related to a human's psychology, cognition and personality. These inner personality traits effect the way a person respond to the environment and it is important to establish these characteristics to understand ‘how’ an individual may behave and also to understand ‘why’ an individual does the things they do. A conceptual model was designed to capture a holistic view of this inner human being.

3.2.1) Exploration for holistic models of a human being.

A literature review was conducted in several phases to discover any existing all-inclusive/comprehensive holistic models of a person and/or the building blocks to construct such a model of a human being. By a ‘holistic model’, we refer to a model that is capable of capturing the varied complex dimensions that an individual comprises of; going beyond demographics and the intentionally declared facts/information and preferences.

An extensive investigation was carried out into existing research on psychology theory, human models, personality models, cognitive models, profiling and adaption of psychology principles and models in advanced analytics to predict human behaviour and personality traits. This was carried out in search of models, theories or frameworks that can be adapted to model the *digital stakeholder* using data analytics.

The study unveiled the existence of a range of theories and models to capture different aspects of human behaviour and personality. While there is a vast body of research, which provide substantiated, tested models to analyse different cognitive and behavioural facets of a human using data analytics, the study revealed a significant lack of an essential holistic, comprehensive model of a human being that can be adopted to model and monitor a *digital stakeholder* (a student, customer, employee etc. based on the domain) in near real-time to generate organizational insights. A holistic, comprehensive model of an individual will provide organizations with the ability to understand their stakeholders' needs and decisions on a deeper, more comprehensive level; empowering the organization to become a pro-active customer-centric organization. We believe the lack of such a model is a significant gap in research and business decision making that if addressed would generate immense opportunities.

3.2.2) A summary of outcomes of the investigation

The study revealed the existence of a vast body of research, which provide proven models to analyse different cognitive and behavioural aspects of a human using data analytics. When analysing these existing theories and research, it became evident that a person’s behaviour, personality and performance is based on and linked to a number of different factors. These discovered factors include demographics, linguistics, individual economics, Big Five personality and dark personality traits (see section 2.1.1 in literature review on human personality and social media).

Based on this comprehension, existing (tested and proven) models were summarized and categorized into five profiles of human psychology, cognition and behaviour as presented herein (later discussed in detail and graphically presented in figure 3.3) to generate the psychological depiction or inner person of the *digital stakeholder*. These profiles will enable organizations to understand their stakeholders on a more comprehensive level than is possible with standard, traditional organizational data sources and predictive models (such as organizational databases and data warehouses).

The proposed model captures a holistic representation of an individual. Understanding the psychological (inner person) profile of a *digital stakeholder* allows all types (descriptive, predictive, diagnostic, prescriptive) of analytics to be performed on a individualistic level to generate more informative insights, enabling an organization to understand more about “why” a stakeholder makes the decisions and choices they do. For example, analytics using organizational data sources can reveal the patterns and trends of choices and behaviour of stakeholders but does not provide a sufficient understanding of the stakeholder’s individual human traits to decipher the ‘reasons’ behind these patterns, behaviours and decisions. An understanding of individuals’ personal psychological traits (the inner person) would give organizations an unprecedented advantage to generate more meaningful insights through analytics that will support a more refined understanding of stakeholders and their changing needs and wants. This will enable an organization to install a proactive approach to serve their customers, keep customers satisfied, strengthen organizational processes (by understanding how each of them perform and serve stakeholders) and obtain the competitive advantage which is much needed to survive in the modern business environment. For example, with such a comprehensive understanding of the stakeholders, organizations will have more knowledge and tools at their disposal to answer questions such as; “Why someone (a student or an employee) failed” and “why (and if) a stakeholder chooses to leave the organization”.

The five profiles constructed for human psychology and behaviour modelling as a result of the investigation are as follows:

1) Psycho-demographic profile

2) Psycho-linguistic profile

3) Big Five personality traits profile

4) Dark Triad profile

5) Psycho-economic profile

The subsequent sections provide a detailed illustration of each of these profiles and how the five profiles were discovered and summarized from existing research and literature.

Human Psychology and behaviour Profiles

1) Psycho-demographic profile

The constructed psycho-demographic profile is presented diagrammatically in figure 3.1. Research publications of (Kosinski et al., 2013), (Farnadi et al., 2016), (Ryan & Xenos, 2011), (Lee & Liu, 2015) and (Amichai-Hamburger & Vinitzky, 2010) established how demographics, highly delicate personal characteristics as well as psychological/emotional states can be accurately captured using digital footprints of individuals. These studies demonstrate the ability to identify characteristics such as ethnicity, sexual orientation, personality traits, religious views, political views, intelligence, happiness, well-being, loneliness, shyness, use of addictive substances, parental separation, age and gender; even when the individual has not disclosed them.

These studies used a variety of contemporary computational personality recognition approaches. This included the use of a combination of approaches comprising of, a variety of machine learning algorithms, diverse feature sets, and assorted types of digital footprints. Data collection environments also demonstrate a narrow range of diversity in terms of the sources and types of social data. This included ground truth social media data from Facebook, Twitter and YouTube as well as online behaviour in microblogs.

However, there were no existing studies, which utilized data from other social media sites (other than the aforementioned three sites) or a combination of more than one type of social media. Nevertheless, there were studies such as (Farnadi, 2016) which have endeavoured to successfully review a range of existing studies combining all three of the above-mentioned social media sites.

Furthermore, studies published by (Youyou et al., 2015) and (Warshaw et al., 2015) provided ground-breaking empirical evidence of how these computer models are more accurate and valid in judging individuals' personalities than their closest relatives and acquaintances.

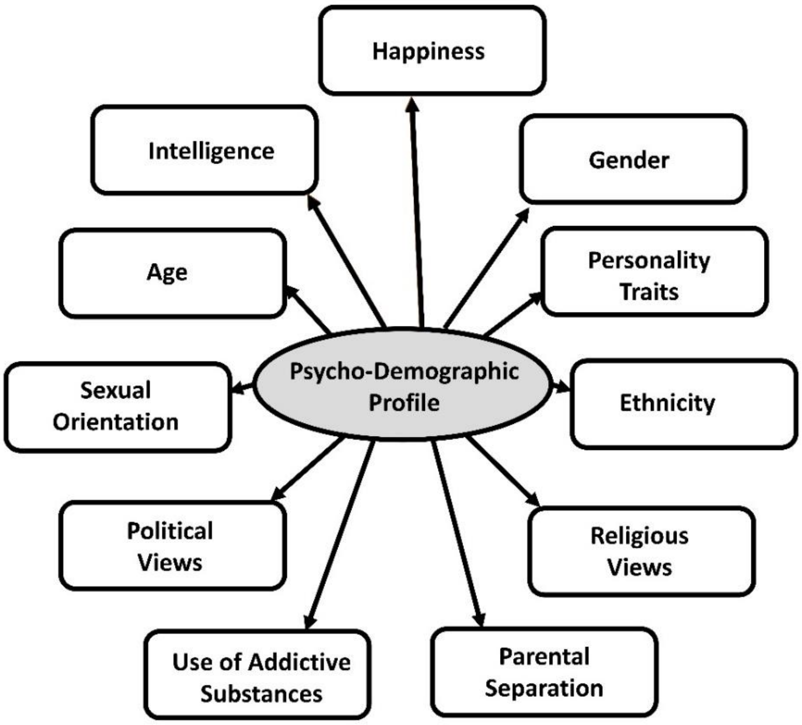


Figure 3.1: Constructed Psycho-demographic profile of a digital stakeholder that can be generated from the digital environment, based on existing research. The diagram illustrates the varied aspects of an individual's psycho-demographic profile that can be accurately discovered and constructed using social media data available in the digital

environment, even when the individual has not disclosed them. The components are mutually exclusive and together form the detailed psycho-demographic profile proposed based on studies by (Kosinski et al., 2013),(Farnadi et al., 2016), (Ryan & Xenos, 2011), (Lee & Liu, 2015) and (Amichai-Hamburger & Vinitzky, 2010). Future research work is expected to enable further enhancement and extension of this profile by introducing new components related to human psychology that are proven to be discoverable through the digital environment using novel machine learning techniques.

2) Psycho-linguistic profile

This model functions by capturing sentiments and emotions from language in the form of text by using text-mining techniques. According to (Farnadi et al., 2016), studies in psychology show that there exist a link between a person's linguistic features and personality traits. Research studies by (Farnadi et al., 2016), (J. A. Golbeck, 2016), (Preoţiuc-Pietro et al., 2015) and (Sewwandi et al., 2017) establish how a person's language usage online can be used to predict personality characteristics.

For example, (Preoţiuc-Pietro et al., 2015) used linguistic features extracted from user generated text to successfully predict an individual's intelligence and education level. Furthermore, research by Farnadi et al. discovered that demographic features such as age and gender have a significant correlation with Big Five personality scores across different types of social media data (three social media data sets had been used for this study) (figure 3.2).

After an analysis of these existing research studies related to user generated content, a psycho-linguistic profile was constructed and proposed as a model (illustrated in figure 3.2). This model functions by capturing sentiments and emotions from language in the form of text by using text-mining techniques such as Sentistregth. A detailed summary of the techniques that have been used (psycho-linguistic analysis tools and techniques) for the generation of psycholinguistic features is presented in the next section titled "Psycho-linguistic analysis tools and techniques" after the Big Five personality profiles section below.

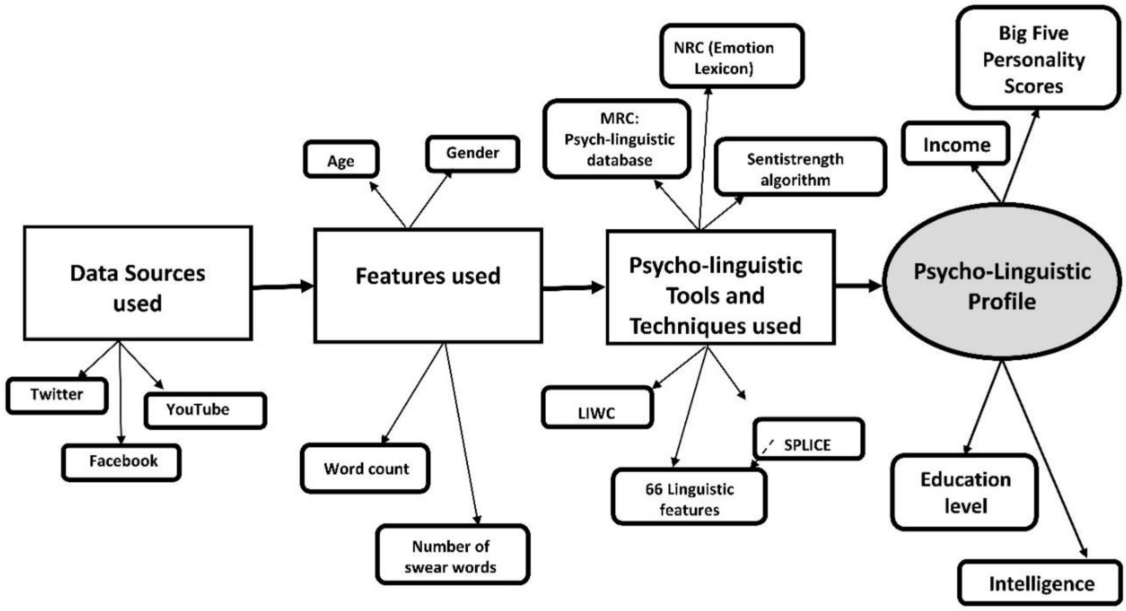


Figure 3.2: The process model for generating the Psycho-Linguistic profile of a *digital stakeholder* from the digital environment, based on existing research. The diagram illustrates the process of generating an individual's psycho-linguistic profile using social media data together with existing psycholinguistic analysis tools (Farnadi et al., 2016; J. A. Golbeck, 2016; Preoţiuc-Pietro et al., 2015; Sewwandi et al., 2017).

3) Big Five personality traits profile

Big Five is one of the most widely accepted models in anthropology to model human personality. Several investigations such as (Ryan & Xenos, 2011), (Lee & Liu, 2015), (Bai et al., 2014),(Ortigosa et al., 2014) and (Amichai-Hamburger & Vinitzky, 2010) demonstrated how the Big Five characteristics ; *openness, conscientiousness, extraversion, agreeableness* and *neuroticism* are predictable from social media content. (Bai et al., 2014) illustrates this further, but utilizes information from micro-blogs instead of social media. (Ferwerda et al., 2016) further adds on to this repertoire by modelling how voluntary non-disclosed information by users in social media (by a study using Facebook) can be used to predict personality (see table 2.1 in the literature review for a detailed description of the Big Five traits).

4) Dark Triad profile

Dark personality characteristics: *Psychopathy*, *Narcissism* and *Machiavellianism* together are known as the "*Dark Triad*". Publications by (Garcia & Sikström, 2014), (Goodboy & Martin, 2015) and (Buckels et al., 2014) demonstrate how these *Dark* personality traits can be plotted using online behaviour and social media usage. By identifying the *Dark Triad*, possibilities of aggressive, impulsive behaviour can be predicted for individuals.

By combining the results of these three studies, it became evident that associations can be derived between the Big Five traits, Cyber bullying behaviour, Online Trolling, gender and global internet habits. This demonstrate the potential and opportunities for interesting predictions and observations related to individuals to be performed.

5) Psycho-Economic profile

Research conducted by (Preoţiuc-Pietro et al., 2015) illustrated a predictive model of individual income based on user generated content and online behaviour in social media (using Twitter posts). This study has managed to draw a strong correlation between predicted and actual income. This information can provide an array of opportunities for organizations to understand the buying power, financial stress and socio-economic status of their stakeholders and enable the usage of this to categorize stakeholders and use in descriptive, prescriptive and predictive analytics to aid in organizational decision making related to the stakeholders.

In addition to the above classification, the investigation revealed a set of tested tools and models that can be used to derive the above elaborated psychological properties from textual data. These findings are represented as the *Psycho-linguistic analysis tools and techniques* section henceforward. These techniques can aid data scientists to extract these profiles from social media data.

Psycho-Linguistic Analysis Tools and Techniques

The following is a list of techniques and tools used for psycho-linguistic analysis in the studies used for the creation of the psycho-linguistic profile proposed above in figure 3.2.

1) Linguistic Inquiry & Word Count tool (LIWC).

LIWC is a text analysis tool used widely in psychology and sociology studies. This tool is capable of extracting meaningful features from text. This includes features related to word count, psychological routes (e.g., number of anger words such as *hate* and *annoyed*), linguistic magnitudes (e.g., the number of swear words), personal disquiets (e.g. the number of words referring to career such as *job* and *majors*), and relativity (e.g., the number of verbs in future tense) (Farnadi et al., 2016; Sewwandi et al., 2017).

2) NRC- an Emotion Lexicon.

This tool categorizes 14000 distinct English words which have been associated with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust), and two sentiments (negative, positive) used for personality prediction (Farnadi et al., 2016).

3) MRC- a psycholinguistic database

This database contains psychological and distributional information about words. The MRC database contains 150,837 entries with information about 26 properties (such as the number of syllables in the word, the number of letters, etc.). Extracted features include number of letters in the word, number of phonemes in the word, number of syllables in the word, concreteness, etc. MRC features exhibit that there is a substantial relationship among *Extroversion* and concreteness, as well as amid *Conscientiousness* and words conveying insight, lengthier words and words learned late by children (Farnadi et al., 2016).

4) SentiStrength – Strength of sentiment

SentiStrength is a heavily used sentiment extraction algorithm (For a detailed elaboration of *sentiment analysis* and the application of SentiStrength see sections 2.3.5 and 2.4.5 in the literature review). The algorithm assigns a positive, negative and/or neutral sentiment score to each text on a scale of 1 (no sentiment) to 5 (very strong sentiment). This has the ability to simultaneously identify sentiments as positive, negative and neutral (Farnadi et al., 2016).

5) SPLICE – Structured Programming for Linguistic Cue Extraction

The SPLICE tool has the capacity to extract 66 linguistic features including prompts which relate to positive and negative self-evaluations such as *I can, I don't know* (Farnadi et al., 2016).

These presented psycho-linguistic analysis tools together have the capacity to be used for detailed analysis and exploration of human personality traits and behaviour from textual social media conversations and presents immense value for organizations and researchers alike.

Research Gap in human models

As summarized above, the study unveiled the existence of a range of theories and models to capture different aspects of human behaviour and personality. This vast body of research, provide substantiated, tested models and tools to analyse different cognitive and behavioural facets of a human using data analytics. As summarized above varied models and tools have been introduced, applied and tested over the past decade utilizing varied social media data.

However, a significant research gap was discovered as there is a significant lack of an essential all-inclusive/comprehensive holistic model of a human being that can be adopted to model and monitor a stakeholder as an individual with varied facets including persona, cognitive and psychological traits using social media data.

Such a holistic, comprehensive model will open doors to an immense range of opportunities for organizations that the existing models alone cannot provide. This include the ability to understand and serve their stakeholders on a deeper holistic level (as elaborated earlier), better predictability of stakeholder future behaviours providing organizations with a competitive advantage to adopt a proactive approach to fulfill it's purpose in serving customers by strengthening customer-centricity in decision making.

3.2.3) Conclusion of the investigation into a holistic model of a digital stakeholder

Grounded on the above reported findings, we derived that it would be highly effective to model a *digital stakeholder* as; a human, a person, an individual and as part of groups (identify and represent group behaviour) in order to understand who they are, what they will do and how they will interact with organizations.

Hence, the results were summarized and categorized into major themes based on the features existing models epitomized. Existing proven models were categorized into five profiles (illustrated in figure 3.3); psycho-demographics (figure 3.1), psycholinguistics (figure 3.2), Big Five, Dark Triad and Psycho-Economics. Additionally, as presented above, existing tools and techniques used in psycholinguistic analysis were identified and classified due to their significant nature in relation to the study (Buckels et al., 2014; Celiktutan & Gunes, 2017; Garcia & Sikström, 2014; Kosinski et al., 2013; Lee & Liu, 2015; Lozano et al., 2014; Preoţiuc-Pietro et al., 2015; Warshaw et al., 2015; Youyou et al., 2015).

Utilizing the summarized existing research and models, the possibility of creating a holistic conceptual model of a human is proposed. This proposed model will have the ability to capture and monitor all major aspects of a human being in a single conceptual model. Furthermore, this model would allow the representation of a person as a *digital stakeholder* inclusive of any version/ avatar of a digital stakeholder such as a “student” in higher education (other e.g. : a customer in a bank or retail, a patient in a health care organization), thus making a significant contribution to the body of research.

This model will open doors to endless possibilities in data analytics to use social media as digital footprints to develop a comprehensive model of a *digital stakeholder* for a deeper, comprehensive understanding of organizational stakeholders. It will offer organizations the opportunity to monitor stakeholders on a deeper level than currently possible and allow all types of analytics to be performed on a deeper level and generate more meaningful and comprehensive insights. This will enable generation of predictions throughout the stakeholder life cycle (stakeholder journey).

With traditional analytics and models performed on traditional organizational data sources, when an organization attempts to explore and answer the question of “why” a stakeholder reaches the milestones they do or makes the decisions they do; organizations are actually searching for information that doesn’t exist in the data. This is because answers to these highly important business questions lie in the stakeholder’s personality traits and individuality. These highly personal and individualistic information related to stakeholders are not captured by the variables in traditional data sources such as databases and data warehouses. Hence, augmenting with social media; is a highly effective and economical solution for organizations to answer stakeholder related questions and generate deeper stakeholder insights.

Additionally, a generic model of this nature is highly adaptable, and has the capacity to transcend time and technology expansions. It has the ability to integrate any new sources of data the organization will consider down the road, such as novel social media data and public opinions in the form of forums (new big data sources), which are becoming increasingly important in data analytics.

In sum, this investigation and literature review exposed a significant gap in research in modeling human personality and behaviour for organizations using data analytics and social media and justified the requirement to conduct further research based on the research question 2 presented in the introduction;

"*How can a holistic and generic framework of a digital stakeholder be developed from an organizational perspective using organizational data and the digital environment?"*

3.2.4) Conceptual Holistic Model of a digital stakeholder

Utilizing the summarized existing research and models, the holistic conceptual model of a human as a *digital stakeholder* was designed and proposed for implementation as illustrated below in a high-level diagram (figures 3.3) and a detailed diagram summarizing the techniques used (figure 3.4).

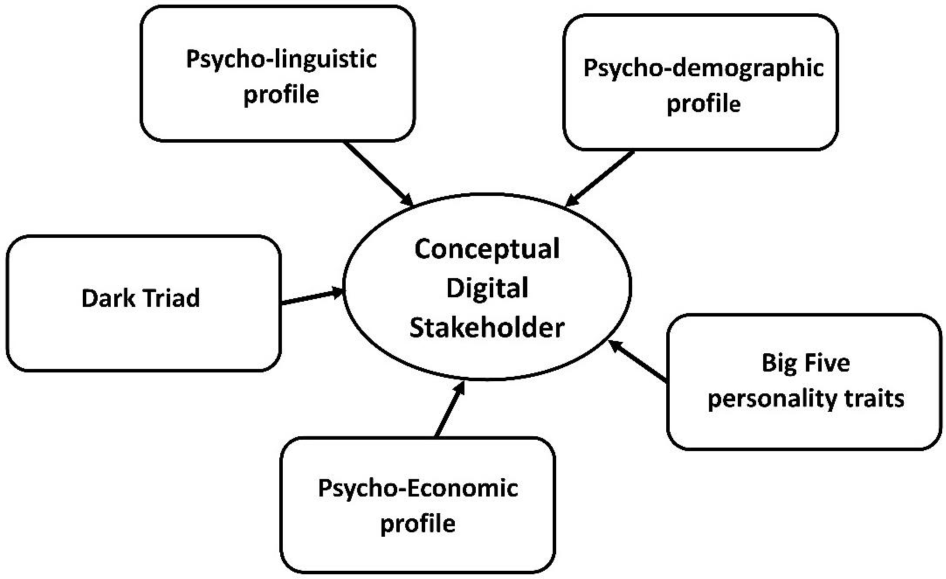


Figure 3.3: A high-level diagram of the proposed conceptual, holistic 'Digital Stakeholder' model. The model presents a comprehensive representation of an individual that can be effectively constructed using the digital environment. Each of the five profiles presented are not necessarily fully mutually exclusive and are supported by a range of existing research studies that have validated their effectiveness and applicability.

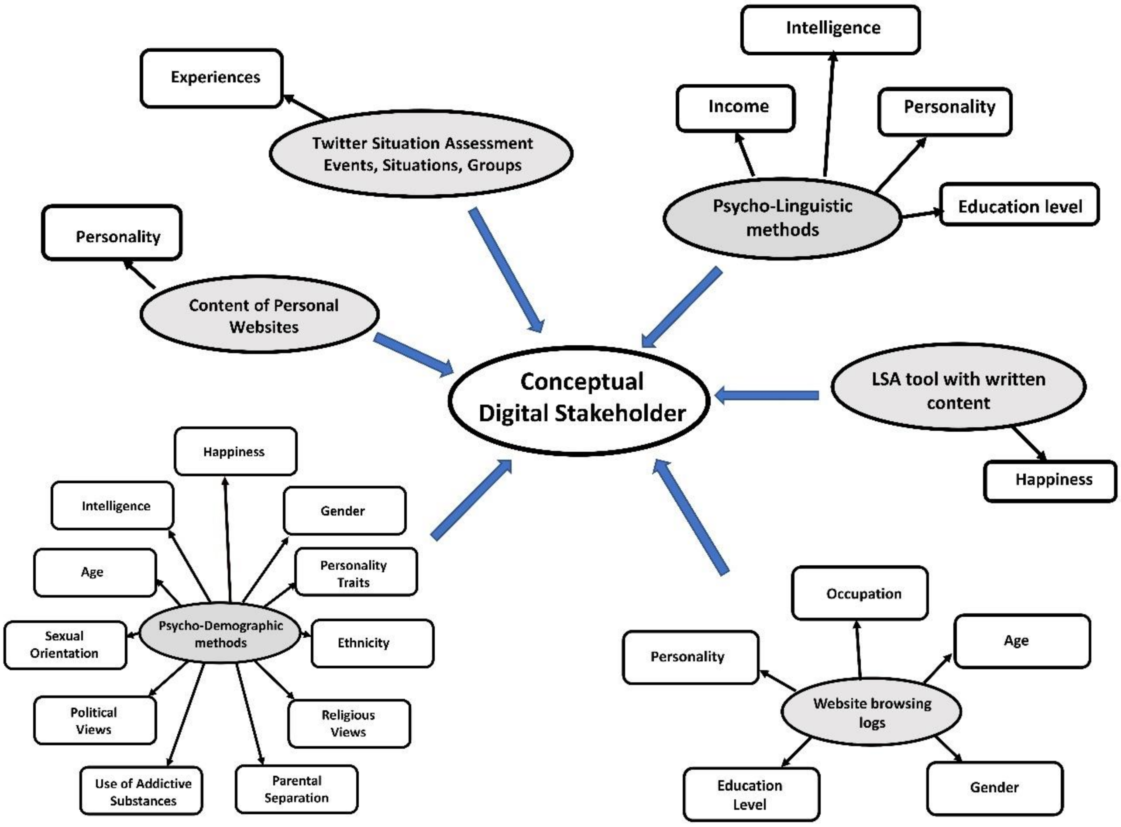


Figure 3.4: The proposed conceptual holistic digital stakeholder model (Detailed/expanded version). The diagram presents a detailed, comprehensive representation of the proposed 'Digital Stakeholder' illustrating all of the traits /features that can be extracted and observed related to an individual using social media data and a high level view of the techniques that have been used to identify these traits/features from social media. The conceptual model is constructed using a collection of profiles, traits and techniques from research literature. The profiles/techniques are not fully mutually exclusive (There are several overlaps such as several profiles/techniques being capable of generating features such as happiness and personality traits).

3.2.5) Transformation Matrix: Psych to Social media

This section presents a novel theoretical and conceptual contribution arising from the research presented in this chapter.

A novel matrix was constructed and proposed to map psychology theory/ traits to features of social media, titled; *Transformation matrix: Psych to social media* (Table 3.1). This matrix presents the first version of such a mapping that is generic and hence can be adapted and extended for future research and experiments. The matrix presents a pathway to transform features of social media to traits of an individual (stakeholder) presented in psychology theory and is based on existing research unveiled through the investigation into literature summarized in the preceding sections of this chapter. This matrix will provide a conceptual tool for researchers and business organizations to understand individuals by leveraging social media. The table 3.2 presents a method matrix, visually summarizing the methods that can be used to generate each of the proposed psychology and behaviour profiles from social media features illustrated in the Transformation matrix in table 3.1. A few example applications of the Transformation matrix and the method matrix is presented in table 3.3.

Table 3.3 presents a few examples such as how;

- Lower income was predicted through increased URLs in posts by (Preoţiuc-Pietro et al., 2015) showing that these users link to external content such as news, pictures or videos.

- User profile features-statistics computed based on user profile information such as number of Facebook friends, number of tweets, number of followers etc. can be used to predict numerus individual traits.
- Higher income users have significantly more followers albeit the number of friends not dependant on income as per (Preoţiuc-Pietro et al., 2015)
- Anger and fear emotions are more prevalent in users with higher income while sadness, surprise and disgust emotions are more connected with lower income (Preoţiuc-Pietro et al., 2015).

Psych to Social Media Transformation Matrix		***Features from social media & digital environment for identification***				
		Behaviour/ Social media usage	User generated content	language	User profile features (# of friends, tweets, followers)	Content of personal websites
Psychology Profiles	Psycho-demographic profile	✓			✓	
	Psycho-linguistic profile			✓		
	Big Five personality traits	✓		✓	✓	✓
	Psycho-Economic profile	✓	✓	✓		
	Dark Triad	✓	✓			

Table 3.1: The proposed Transformation Matrix: Psych to Social media. The matrix illustrates the mapping of theory/traits from Psychology to features of social media based on existing research literature.

Method Matrix		*Methods for identification from social media*				
		Machine learning algorithms	Regression	Psycho-linguistic tools & techniques	Word clusters	Ontology based personality detection
Psychology Profiles	**Psycho-demographic profile**	✓ (J. Li et al., 2014)		✓ (P. Liu et al., 2015)	✓ (Schwartz et al., 2013)	
	Psycho-linguistic profile		✓ (Preoţiuc-Pietro et al., 2015)	✓ (J. A. Golbeck, 2016)	✓ (Schwartz et al., 2013)	✓ (J. Li et al., 2014)
	Big Five personality traits	✓ (Farnadi et al., 2016)	✓ (Bai et al., 2014); (Bachrach et al., 2012)	✓ (J. A. Golbeck, 2016); (Farnadi et al., 2016)		✓ (J. A. Golbeck, 2016)
	Psycho-Economic profile		✓ (Preoţiuc-Pietro et al., 2015)	✓ (Preoţiuc-Pietro et al., 2015)	✓ (Preoţiuc-Pietro et al., 2015)	
	Dark Triad		✓ (Goodboy & Martin, 2015)	✓ (Garcia & Sikström, 2014)		

Table 3.2: The proposed “method matrix” to derive psychological and behaviour profiles from features of social media. The matrix illustrates the mapping of theory/traits from Psychology to methods of analysing social media based on existing research literature.

Example Applications of the Psych to Social Media Transformation matrix		Features from social media & digital environment for identification					
		Behaviour/ Social media usage	**User generated content**	**language**	**Profile features**	**emotion**	**Content of personal websites**
Psychology Profiles	**Psycho-demographic profile**	User likes on Facebook to predict intelligence, religion, gender & substance use (Kosinski et al., 2013).	Status updates used to predict subjective well-being (P. Liu et al., 2015)	Language usage to predict happiness (Schwartz et al., 2016)	Location within a friendship network in Facebook to predict sexual orientation (Kosinski et al., 2013)	Emotion scores of status updates on Facebook to predict happiness (P. Liu et al., 2015)	Web browsing logs to predict intelligence (Kosinski et al., 2013)
	Psycho-linguistic profile	User behaviour on twitter used to predict user income (Preoţiuc-Pietro et al., 2015)	Personalities predicted from user generated text using LIWC, NRC, MRC and SPLICE (Farnadi et al., 2016)	Intelligence predicted through words used (J. Golbeck et al., 2011)	Homophily-network information of Facebook profiles used to predict users' education level (J. Li et al., 2014)	Emotions associated with written language to distinguish gender (Schwartz et al., 2013)	Written content from connected personal websites to derive intelligence and education level using psycho-linguistic analysis tools
	Big Five personality traits	Facebook groups, likes and statuses to predict *openness*	User authored text samples to estimate Big Five personality traits	Linguistic analysis using LIWC for personality	Number of Facebook friends to predict extraversion (J.	Sentiment extraction from text posts used in combination other techniques to identify	Music collections to predict personality

		(Youyou et al., 2015); (Kosinski et al., 2013)	using Receptivity API based on LIWC (J. A. Golbeck, 2016)	detection (J. A. Golbeck, 2016)	A. Golbeck, 2016); (Bachrach et al., 2012)	personalities (Farnadi et al., 2016); (Kosinski et al., 2013)	traits (Kosinski et al., 2013)
	Psycho-Economic profile	high rate of twitter posting related to lower income (Preoţiuc-Pietro et al., 2015)	Lower income predicted through increased URLs in posts, posts with duplicate content (Preoţiuc-Pietro et al., 2015)	Increased swear words used to predict low income (Preoţiuc-Pietro et al., 2015)	Number of followers and number of retweets to predict high income (Preoţiuc-Pietro et al., 2015)	Anger and fear emotions to predict higher income, sadness, surprise and disgust for lower income(Preoţiuc-Pietro et al., 2015)	Written content from connected personal websites to predict user income using psycho-linguistic analysis tools
	Dark Triad	Online commenting behaviour to predict trolling and dark traits (Buckels et al., 2014)	User generated Facebook status updates to predict dark personality traits using LSA (Garcia & Sikström, 2014)	Language used in Facebook status updates to predict psychopathy and narcissism (Garcia & Sikström, 2014)	Number of Facebook friends and frequency of status updates to predict neuroticism (Garcia & Sikström, 2014)	Emotions associated with written content used to predict dark personality traits on Facebook (Garcia & Sikström, 2014); (Buckels et al., 2014)	Language and word usage in user written comments on website and blog posts to predict dark personality traits

Table 3.3: Example applications of the Psych to Social Media Transformation matrix. The table presents example applications of the transformation matrix presented in table 3.1 and the method matrix presented in table 3.2 from existing research studies. The shaded boxes present suggestions by the authors that have not yet been validated through experiments.

The next section 3.3 explores how the proposed conceptual, holistic model can be further adopted to understand stakeholder behaviours and monitor the stakeholder's journey through the organization.

3.3) Digital stakeholder (Part 2)- Behavioural Depiction in relation to the organization

The next phase of the research study presented in this chapter is dedicated to understand and model the outer-person of this *digital stakeholder* or the interactions of the stakeholder observable by the organization. The outer-person represents the behavioural aspects of the *digital stakeholder* in relation to the organization offering an understanging into how the individual (modelled in section 3.2 above) moves through the organization. This model will enable observation of the stakeholder journey through the organization or the stakeholder life cycle.

Through combining all of the research outcomes and models presented so far in this chapter, a generic, comprehensive framework was designed to model the digital stakeholder's journey through the organization and is presented in figure 3.5. This framework uses the holistic conceptual model of the digital stakeholder (proposed in section 3.2 above) and accepts social media data and organizational data sources as input. Furthermore, moving through each phase; it accepts organizational strategy and requirements as input at several stages to generate outcomes and ultimately create organizational insights as output.

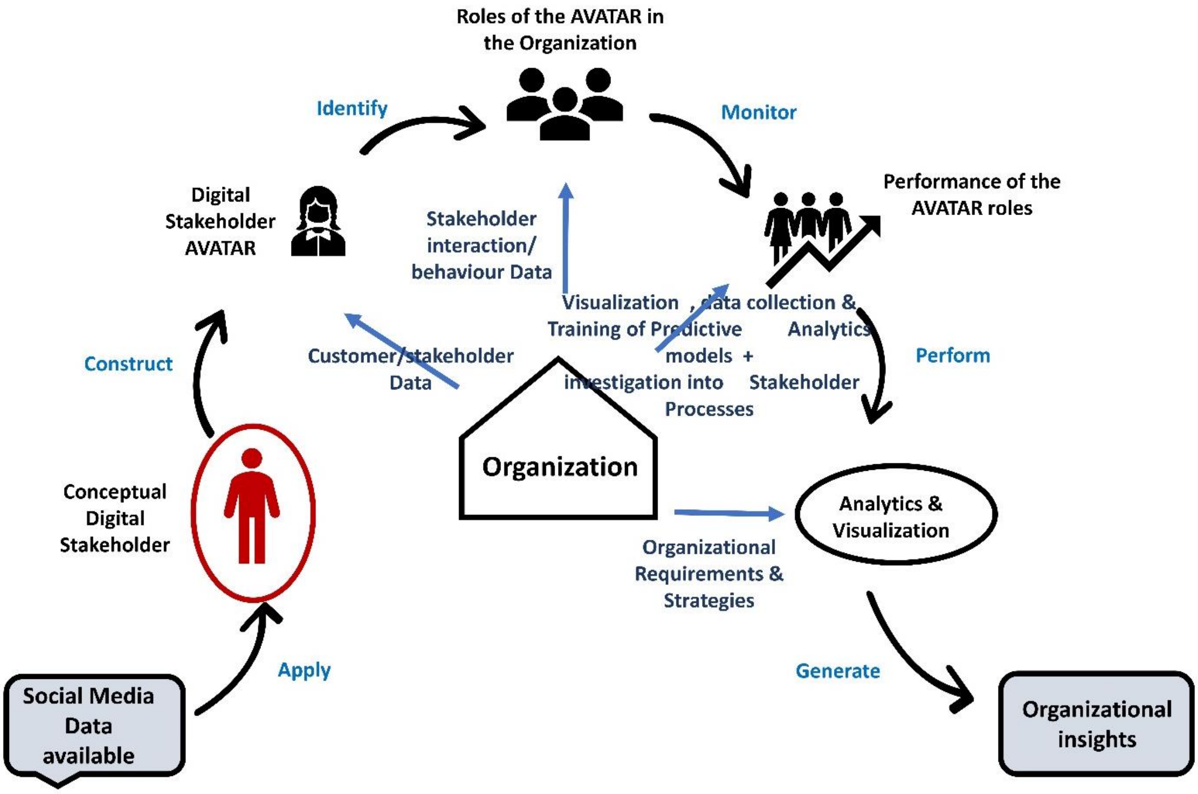


Figure 3.5: Proposed framework of the Digital Stakeholder's journey through the organization. The framework uses the proposed holistic conceptual model of the digital stakeholder (presented in figure 3.4) and accepts social media data and organizational data sources as input while receiving organizational strategy and requirements as input at several stages to generate organizational insights as output. The framework enables an organization to create comprehensive Avatars for any stakeholder the organization wants to observe (eg: customer, student, patient) and identify the different roles of the stakeholder's life cycle and monitor stakeholder performance according to the stages in the stakeholder life cycle and perform analytics to generate varied insights at each stage to serve and support their stakeholders in a proactive manner.

First, using the available social media data and the holistic conceptual *digital Stakeholder* model proposed in figure 3.4, together with stakeholder data (eg: customer data) from the organization; a *Digital stakeholder Avatar* is constructed in alignment with the requirement at hand for the organization. For example, the *Digital Stakeholder Avatar* can be any customer category that the organization wants to know more about; a student for a higher education institute, a customer in a bank or retail, a patient in a health care organization etc.

The next phase in the framework involves identifying roles of this *Avatar* (figure 3.5) using stakeholder interaction and behaviour data from the organization. This step involves identifying the stakeholder life cycle. For example, for a *student Avatar* in a higher education organization, this could be potential student, current enrolled student and graduate/alumni, or for a customer of a bank, this could be potential customer, current customer and churned customer. In section 3.4 below we demonstrate a case study application of this framework in an organizational domain.

The third phase of the framework recommends monitoring the performance of this *Avatar* using the identified Avatar roles via data analytics techniques (visualization, data collection and training of predictive analytics models) and an investigation into organizational processes (the application of these phases will be demonstrated in section 3.4 using a case study organization). This phase allows monitoring and collecting data related to roles of a stakeholder separately. For example this allows not just observing but collection and organization of data according to the roles which allow building predictive models in future to predict what traits of individuals promote a stakeholder to join the organization and move through the life cycle and also enable making predictions (based on individual traits collected by the Avatar) about *how* each stakeholder would behave, *what* decisions they will make and the reasons *why* they will behave or decide in a certain manner.

Finally, these collected data on the performance of stakeholder roles will be used in alignment with the organization's strategy and requirements to generate analytics and visualization. Every type of analytics will be explored; descriptive, diagnostic, predictive and finally prescriptive analytics; to generate organizational insights to inform organizational strategy. The application of this framework is presented using a case study demonstration in the next section, section 3.4.

This framework enables organizations to learn more about their stakeholders and further enable timely interventions to support and serve the stakeholder at each stage or milestone of the stakeholder life cycle.

The following is a demonstration of a generic application of this framework: a development of a predictive model to be used with the framework to monitor the stakeholder lifecycle and enable prediction of stakeholder decisions and behaviour.

Application: A predictive model built using the proposed framework of the Digital Stakeholder's Journey through the organization (Three-step predictive model)

By applying the proposed framework (presented in figure 3.5) and using stakeholder interaction and behaviour data obtained from phase 3: identification of the roles of the *Avatar* in organization (see figure 3.5) , phase 4: monitoring of Performance of the *Avatar* roles using analytics, together with research conducted, a three-step predictive stakeholder model was designed and proposed for organizations (Figure 3.6).

This model provides two layers,

- An observation layer for milestones (Potential Stakeholder→ Current Stakeholder→ Churned/ graduated stakeholder). This layer involves phase 3 of the introduced framework; collecting training data related on the stakeholder's interactions/behaviours with the organization (figure 3.6).
- A predictive layer (using predictive models X, Y and Z) to monitor the stakeholder in near real-time and enable prediction and observation of stakeholder performance/outcomes throughout the stakeholder life cycle as a continuous process (figure 3.6).

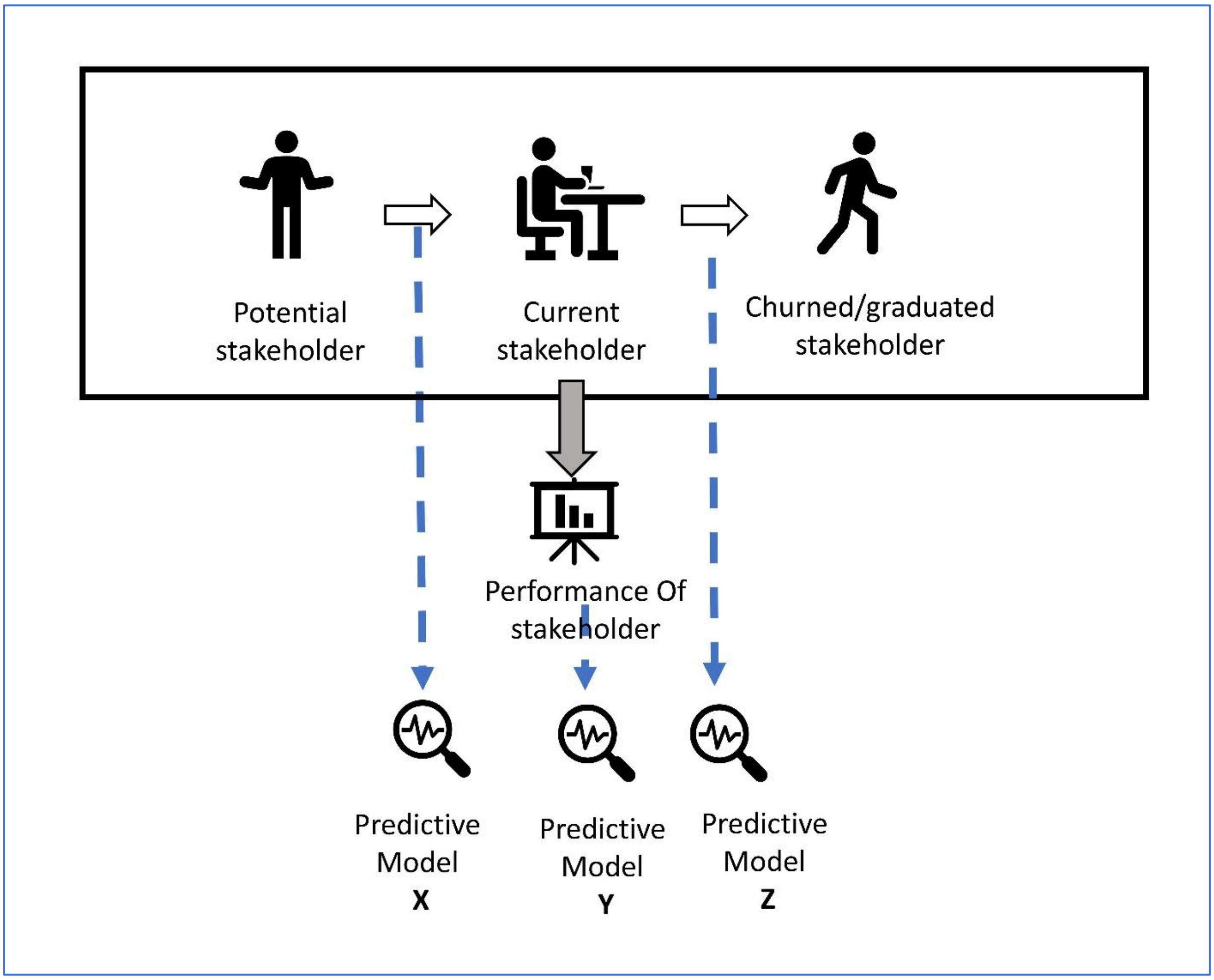


Figure 3.6: The proposed generic three-step predictive model for monitoring the life cycle of the proposed 'Digital Stakeholder' presented in three milestones; potential stakeholder, current stakeholder and churned stakeholder. This model is built to be used with the proposed framework in the phases. This model enables a digital persona/Avatar to exist as a part of the organizational processes and facilitate identification and prediction of potential next outcomes rather than a traditional 'post-mortem' analysis after the events have occurred. Furthermore, this model enables transformation of freely available information from social media to generate value in the form of the 'Digital Stakeholder'

This three-step model (presented in figure 3.6) was applied and adopted to the case study domain; higher education and presented as a life cycle of a *Student Avatar* (in three milestones of a university student) in the case study demonstration section (figure 3.8).

The next section demonstrates applications of the proposed novel conceptual model and the novel generic framework using the case study organization followed by a discussion on the varied applications and possibilities that will be enabled by this proposed framework including a case study application of the proposed three-step predictive model .

3.4) Case Study Application

The La Trobe University was used as an organizational case study to explore and demonstrate the proposed *digital stakeholder* (presented in figures 3.3 and 3.4) and to trial and demonstrate the applicability of organizational data sources in modelling the proposed generic framework of the *digital stakeholder's* journey through the organization in a business domain (higher education in this case). The university was selected due to the convenience of access to it's business intelligence processes and data sources through the university's ICT Business Intelligence unit and due to the significance of the higher education domain as a recently emerged industry into contemporary business competition .

3.4.1) Introducing Higher Education domain into the context

Currently, higher education institutes are collecting a vast amount of student data from a multitude of different sources in a variety of formats similar to other business organizations. This data constitutes not only of structured data, but semi-structured as well as a large volume of unstructured data, such as data from social media, forums (public opinions) and location awareness (student movement patterns).

3.4.2) Selection of Data Sources to build the holistic model

The main types of data sources required for this study are organizational and external, public social media data related to stakeholders (students, parents etc.). The purpose of this study is to generate the capability to use social media data to enhance the creation of stakeholder personas together with the organizational data that is directly related to the stakeholders.

1) Organizational Data Sources: The La Trobe University

The main source of stakeholder data (student data and course data when considering a higher education institute) that can be used to understand the interactions between the stakeholder and organization are selected as the data generated by the varied processes of the La Trobe University. These data sources were mapped into table 3.4. (It was apparent that many of these data sources still existed in silos with their potential for analytics untapped at the time of investigation).

Name of data source	Data source explained	Type of data contained	Other relevant information on the data source
SAP-BW:	La Trobe Business Warehouse	Student data source	centralized
SISone	Student Database	Student data source	centralized
ResearchMaster	research administration Database	Student data source	centralized
ISIS	International Student Information System	Student data source	centralized
LMS	Learning Management System	Student data source (holds course material)	centralized
CIMS	Course Information Management System	Course data source	centralized
Library system	Independent data source	Student data source	not centralized
Sports Centre system	Independent data source	Student data source	not centralized
Accommodation Services System	Independent data source	Student data source	not centralized
Security & Car park system	Independent data source	Student data source	not centralized
Child Care Centre System	Independent data source	Student data source	not centralized
Office 365	(mail box sizes, OneDrive space usage, home installations)	Student data source	with Private data owners
Building Security access	administrative data source	Student data source	with Private data owners
Room booking System	administrative data source	Student data source	with Private data owners
Other University research data	Administrative/academic data source	Student data source	with Private data owners
University calls	administrative data source	Process data source	CRM- Marketing dept
University website use	administrative data source	Process data source	CRM- Marketing dept
Websites of university clubs and societies	Independent data source	Student data source	with Private data owners
LaTrobe ALUMNI	Independent data source	Student data source	not centralized

Table 3.4: A summary table of identified current systems /sources of student and course data at the La Trobe University Melbourne, Australia.

2) External/public/social forum data

As sources of public social media data rich in information and opportunities related to the case study domain; higher education, three public online discussion forums were selected, namely;

- Whirlpool (http://forums.whirlpool.net.au/)
- ATAR Notes (https://atarnotes.com/forum/)
- Bored of Studies (http://www.boredofstudies.org/)

These three forums were selected, as they are the current most widely used public discussion forums in Australia with the largest selection of posts and discussions related to various aspects of Australian higher education organizations (see section 2.3.5 in the literature review).

Data from the public forums were collected using web scraping techniques utilizing Java and prepared for analytics and visualization using ElasticSearch and Kibana. The scraped data set used contains over 1,200,000 posts from the three selected forums.

3.4.3) Demonstration of an organizational Avatar of the Conceptual Digital Stakeholder model: A student model for higher education organizations

The generic framework proposed in figure 3.5 was applied to the La Trobe university using the university's data sources, strategy plan and requirements.

The *conceptual digital stakeholder* model (proposed in figure 3.3 and 3.4) was adopted to the currently available student data at the La Trobe University to construct a *Latrobe Student Avatar* (Figures 3.7). This Avatar is constructed based on two types of data; data available at the organization: La Trobe university, which are directly related to the stakeholder, and external data sources (big data sources: forums, other social media and research) which may be related to the stakeholder indirectly which can be combined in order to help understand the stakeholders better.

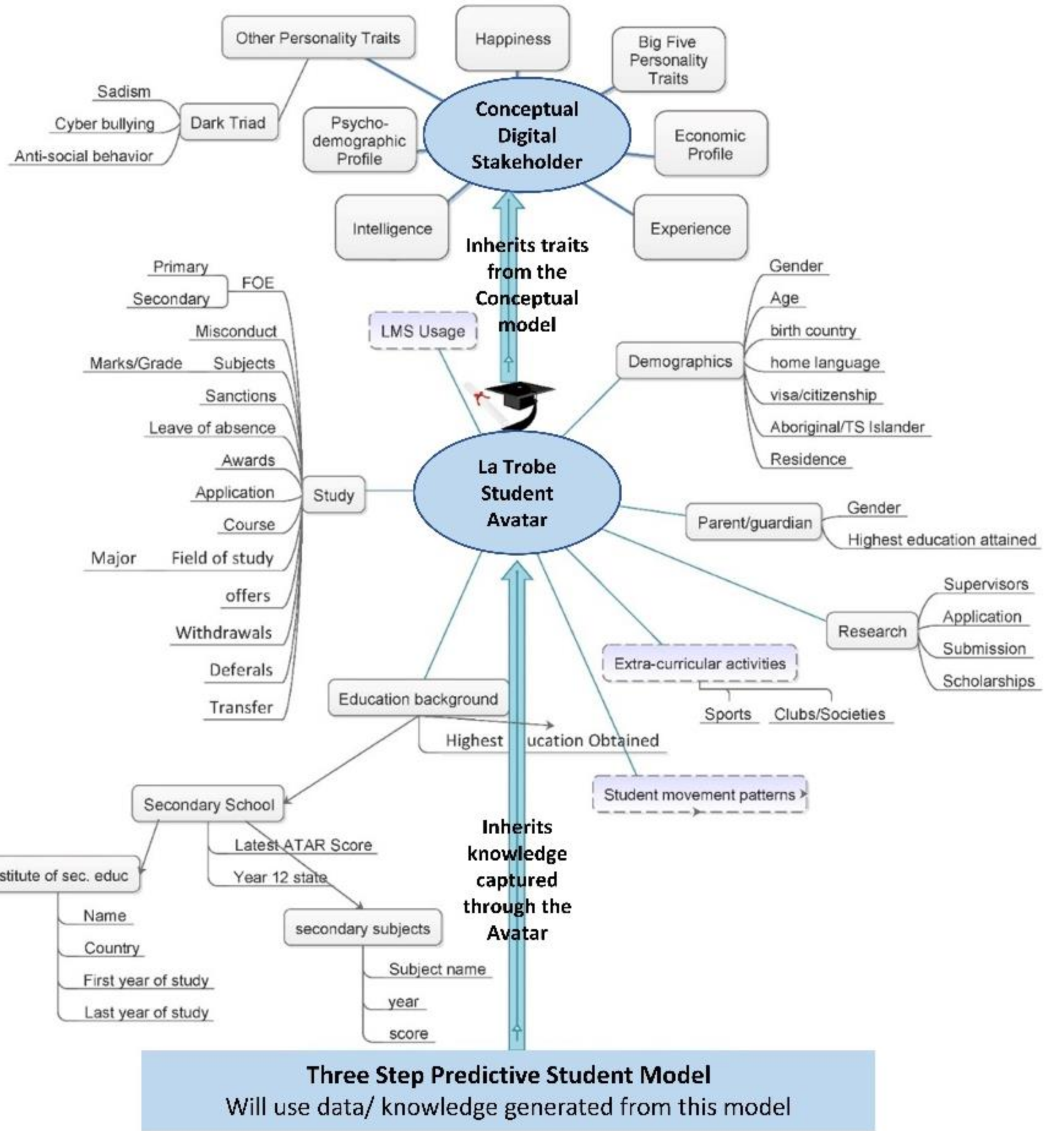


Figure 3.7: The “Digital Stakeholder Avatar” of a Student in a higher education organization (La Trobe Student Avatar) constructed by adopting the proposed *Conceptual digital stakeholder* model (proposed in figures 3.3 and 3.4) and the Framework of the digital stakeholder’s journey through the organization (proposed in figure 3). The diagram illustrate how a *Student Avatar* can be constructed by inheriting traits from the *Conceptual digital stakeholder* model and using public and organizational data sources available for the organization. The diagram further shows how this *Avatar* instance can be used to feed in

training data for a predictive model (the three-step predictive student model is presented in figure 3.8), the three-step predictive student model inherits data/knowledge from this *student Avatar model* which inherits traits from the *conceptual model*; as illustrated by the blue arrows.

3.4.4) Application of the three-step predictive stakeholder life cycle model to the case study domain

For the La Trobe university, the application of the proposed generic three-step predictive model proposed in figure 3.6 generated a unique three step predictive student model (presented in figure 3.8) that the university can train and use to observe milestones of their students. The model uses data/knowledge acquired by the La Trobe *Student Avatar* model presented in figure 3.7.

The model provides an observation layer (Potential student → Enrolled student → Graduate) and a predictive layer (using predictive models X, Y and Z) to monitor students in near real time, and thus empower prediction and observation of student success/outcomes throughout the student journey as a continuous process. From the moment a potential student (which could be a student still enrolled in high school who comes on a university tour, to students who apply for an offer) shows interest, to the student getting enrolled, moving through course work and other extra/co-curricular engagements, to graduating and becoming an alumni.

The 3 predictive models would provide the following capabilities to the La Trobe university (figure 3.8) once they have been trained with data collected over a period of time:

X- Predict how potential students (applicants etc.) may perform after enrolment

Y- Predict how enrolled students will perform during their study period at the university

Z- Predict how graduated students will perform in life/society after graduation

This model will introduce a technique to categorize students according to the university's expectation of an ideal student and help in interventions to drive students into desirable categories if and when they are slipping out of them or underperforming. Figure 3.9 illustrate this entire process of application of the framework (of the Digital Stakeholder's journey through the organization) to an organization (the La Trobe University) graphically.

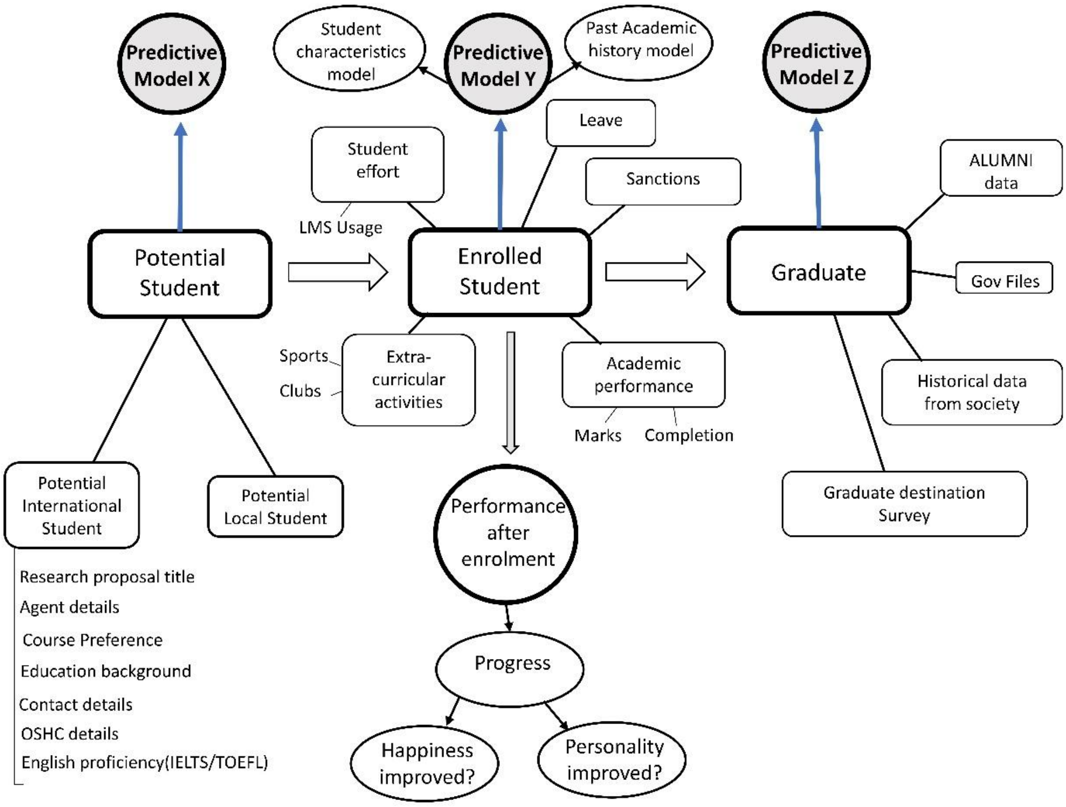


Figure 3.8: The Proposed Three Step Predictive Student Model for Higher Education Organizations. The predictive model is proposed to be trained from data collected using the *Student Avatar* presented in figure 3.7.

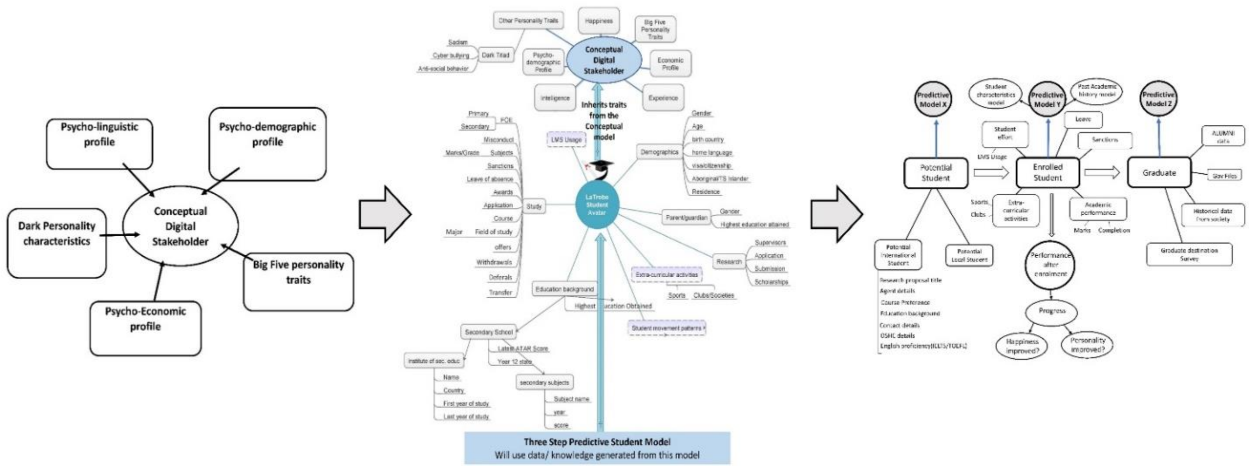


Figure 3.9: Application of the framework of the *Digital Stakeholder's journey through the organization* to a case study organization (the La Trobe University)

3.5) Chapter Summary and discussion of implications

This chapter defined and established the concept of a *digital stakeholder* from an organization perspective, utilizing organizational data and public big data in the form of social media.

The chapter presented a novel holistic conceptual model of a *digital stakeholder* that can model the diverse traits and aspects of a human being by integrating internal data (organization generated and owned data) with unstructured external data (the digital environment: public data present in social media) available for the organization. A set of novel matrices are presented as a conceptual theoretical contribution for mapping theory from psychology to features of social media. The chapter also proposed a novel generic framework for the journey of this modelled *digital stakeholder* through the organization. This framework enables observing of the stakeholders' journey through the organization while facilitating monitoring and prediction of stakeholder behaviour and performance at different stages/milestones of the stakeholder life cycle. Finally, a demonstration is made on how this generic model and framework can be adopted to a specific organization domain by using the higher education domain as a case study and developing a holistic model of a university student as a *digital stakeholder avatar* and presenting the student journey using the proposed framework. This model and framework enable organizational analytics and predictive models to be performed on a much deeper comprehensive level than was possible earlier with the use of standard organizational data sources alone (data bases and data warehouses) and traditional predictive models.

With traditional analytics and models performed on organizational data sources, when an organization attempts to explore and answer the question of "why" a stakeholder reaches the milestones they do or makes the decisions they do; organizations are actually searching for information that doesn't exist in the data. This is because answers to these highly important business questions lie in the stakeholder's personality traits and individuality. This information related to stakeholders' is not captured by the variables in data bases and data warehouses. Hence augmenting with social media, proposed in this chapter is a highly effective and economical solution (as the social media data sources are freely available) for organizations to answer stakeholder related questions and generate deeper stakeholder insights. Furthermore, this allows generation of these deep insights in near real-time as social media data is high in velocity and updated in real-time. The research outcomes from this chapter are utilized in the proceeding contribution chapters of the thesis to develop novel analytics solutions to acquire more strategic organizational knowledge and insights.

4. Perceiving the organization from a social media perspective

"Where Perception is, there also are pain and pleasure, and where these are, there, of necessity, is desire…"

~ Aristotle.

Comprehension or true understanding of anything in existance requires it to be perceived through a lense or paradigm. This chapter presents an attempt to view and comprehend the organization from a window of social media to address the research question presented in the introduction chapter; "*To what depth can knowledge and value be captured from the digital environment and social media to understand communication and interactions between organizations and individuals*".

Nowadays, people are more and more involved in social media and we are increasingly presented with evidence of how these digital footprints of individuals has the potential to be used by organizations to understand themselves, if the data can be structured into some form capable of being transformed into meaningful and useful insights.

This chapter introduces a novel framework and techniques for understanding the stakeholder perception about an organization (Perceived Organization) from a social media perspective.

Building on the Digital Stakeholder presented in chapter 3, this chapter introduces a framework to structure and transform social media into lenses or aspects of the organization (social media to organization transfiguration framework). Furthermore, novel techniques are proposed to remodel raw unstructured social media data to generate knowledge of these organizational aspects using digital traces of stakeholder emotions and opinions in social media.

The chapter presents a detailed unpacking of organizational features and aspects represented and discussed over social media in the form of public online discussion forums. Techniques are presented to leverage the social media and monitor the organizational perception from varied lenses of products, processes and focus areas as well as to understand the positioning of the organization among its competitors.

The proposed techniques are applied and the effectiveness demonstrated on a case study organization using a public social media data set of over one million textual discussions. Moreover, the chapter demonstrate developing of techniques to generate a layer of actionable insights aligned with organizational strategy to improve the perceived stakeholder experience by acquiring knowledge on the organization's performance related to the aspects of products, processes, focus areas and competitors using stakeholders' word-of-mouth.

4.1) Stakeholders' Perception of an Organization

Currently organizations are becoming increasingly fluent at Impression Management (IM). Organizations are using a variety of techniques as an attempt to influence and maintain their perception in the eyes of stakeholders. As uncovered by our investigations using the case study organization (figure 4.1) some of these techniques include social media engagement, promotions, emails, blogs and varried planned demonstrations through the company website. Regardless of the effort put into IM, organizations still lack a technique for truly understanding the perceptions held in the minds of their valuable stakeholders. The only image the organization has, is the image it sees through it's own lense or 'mirror' . This mirror image is what the organization tries to be on top of and maintain through the varied channels and strategies. This is the image held as the organization's perception by the board of directors, executive officers and the decision makers. However, the true image perceived and held by the stakeholders (the *Perceived Organization* or *stakeholder imagery*) can at times be very different to this refelection viewed through the organization's mirror. The *Perceived Organization* consists of stakeholders' deeply held emotions, believes, memories, wants and needs related to the organization. These are constantly changing and generally evoked by their experiences with the organization and its products and services as well as the influence from information diffusion such as word of mouth. The *Perceived Organization* or stakeholder imagery can be considered as an amulgamation of *user imagery*, *customer imagery* and *employee imagery* observed in organizational IM and brand literature (summarized in section 2.2.3 in the literature review).

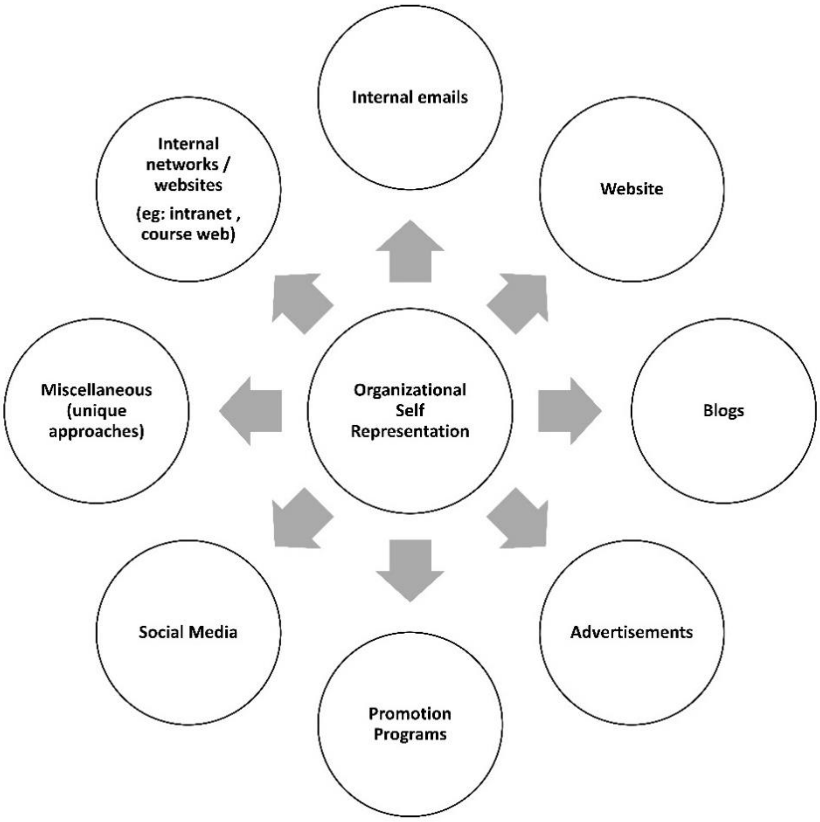


Figure 4.1: Dimensions of organizational self portrayal: Modelled using an investigation into a case study organization in higher education, the diagram illustrates the dimensions through which an organization engages in Impression Management (IM)

Organizations have struggled for decades to understand this true perception of it-self in the minds of stakeholders, in contrast to the “mirror” reflection. Organizations have traditionally made attempts to understand this *Perceived Organization* through varied techniques such as questionaires, surveys and interviews as summerized in the literature review (section 2.2). Regardless of the attempts and recent research, this continues to be a challenge due to the elusive nature of stakeholders’ true emotions.

When we look at the current digital environment, as established in chapters 1 and 2, the recent boom of social media has opened doors of immense opportunity for organizations to understand their stakeholders' emotions. Furthermore, as revealed by introducing the Digital Stakeholder in chapter 3, now it is possible to truly understand stakeholders on a deeper level more than was possible ever before by augmenting with social media. Additionally, due to the nature of social media compiled by section 1.2, the availability of data together with technology drastically reduces the amount of resources (time, money and expertise) required to understand these stakeholders. In section 3.1 the Digital Stakeholder was defined as the cognitive, affective, emotional, social and physical understanding of a stakeholder generated through their digital footprint left behind in the social media environment. Hence, understanding this Digital Stakeholder unpacked in chapter 3 is the key to develop a technique to understand the *Perceived Organization* through social media.

However, in the attempt to identify and extract the Digital Stakeholder's *Perceived Organization*, two key challenges that needs to be overcome were identified. The first challenge was presented by the four Vs (Volume, Velocity, Variety and Verasity) and the complexities inherent to social media highlighted in section 2.3. The main difficulty is that social media data constitutes not only of structured data, but semi-structured as well as a large volume of unstructured text. Transferring opinions, ideas and emotions expressed in text into a structured form that can be evaluated, quantified and reported is a highly challenging process that is still under research (Wijenayake et al., 2021).

The second challenge was unpacking the aspects and features which make up the organization in the stakeholders' mind or perception. The organization is a highly complex integration and rendezvous of varied tangible and intangible aspects. Unpacking this cosmos of the organization is an essential step in developing a technique to understand and unveil the *Perceived Organization*. The following section, section 4.2 presents the unpacking and generation of a mapping between social media and organization.

4.2) Leveraging social media to understand the Perceived Organization

The present big data environment is home to a rich and diverse ecology of social media platforms (Wijenayake et al., 2021). As illustrated in sections 1.2 and 2.3, the boom of the internet and the unprecedented increase in the number of social media users have given rise to a large number of social media platforms. These platforms vary in terms of their scope and fuctionality which can be viewed by observing a few of the most popular platforms such as Facebook, Youtube, Twitter and Linkedin. Each of these platforms demonstrate significant contrast with the others as they all have their own unique scope and functionalities. This diversity and complexity of social media necessitate a structurally rigorous model to be applied to aid in unpacking. For example, before engaging with social media data, first a decision has to be made on which functionalities of the social media should be considered to engage with for the task at hand (Wijenayake et al., 2021). Information related to an organization is distributed across various social media conversations and posts under multiple themes, topics and focus areas. Hence, the challenge is to come up with a structure or framework to bring these scattered pieces of information together in a meaningful manner.

The honeycomb social media framework proposed by Kietzmann et al. in et al. in (Kietzmann et al., 2011) and introduced in section 2.3.10 was selected to assist in unpacking social media to develop our novel framework to comprehend the *Perceived Organization*. This honeycomb framework was selected as it is the most widely used and accepted social media framework in social media literature. This model will assist us to identify the main building blocks or areas of the social media to engage with for maximum impact.

As elaborated in section 2.3.10, the honeycomb framework introduces a honeycomb of seven functional building blocks: *identity, conversations, sharing, presence, relationships, reputation*, and *groups*. The blocks allow to unpack and examine different facets of the social media user experience, and their implications for organizations. The building blocks help to understand how different levels of social media functionality can be configured (Kietzmann et al., 2011). According to this framework each social media platform is driven by primary, secondary and tertiary building blocks, which provide the foundation for important social media design decisions (Kietzmann et al., 2012). The figure 4.2 from (Kietzmann et al., 2011) illustrate the application of this framework to four popular social media platforms ; Facebook, YouTube, LinkedIn and Foursquare, demonstrating the diversity of the social media platforms through the primary, secondary and tertiary building blocks of each platform.

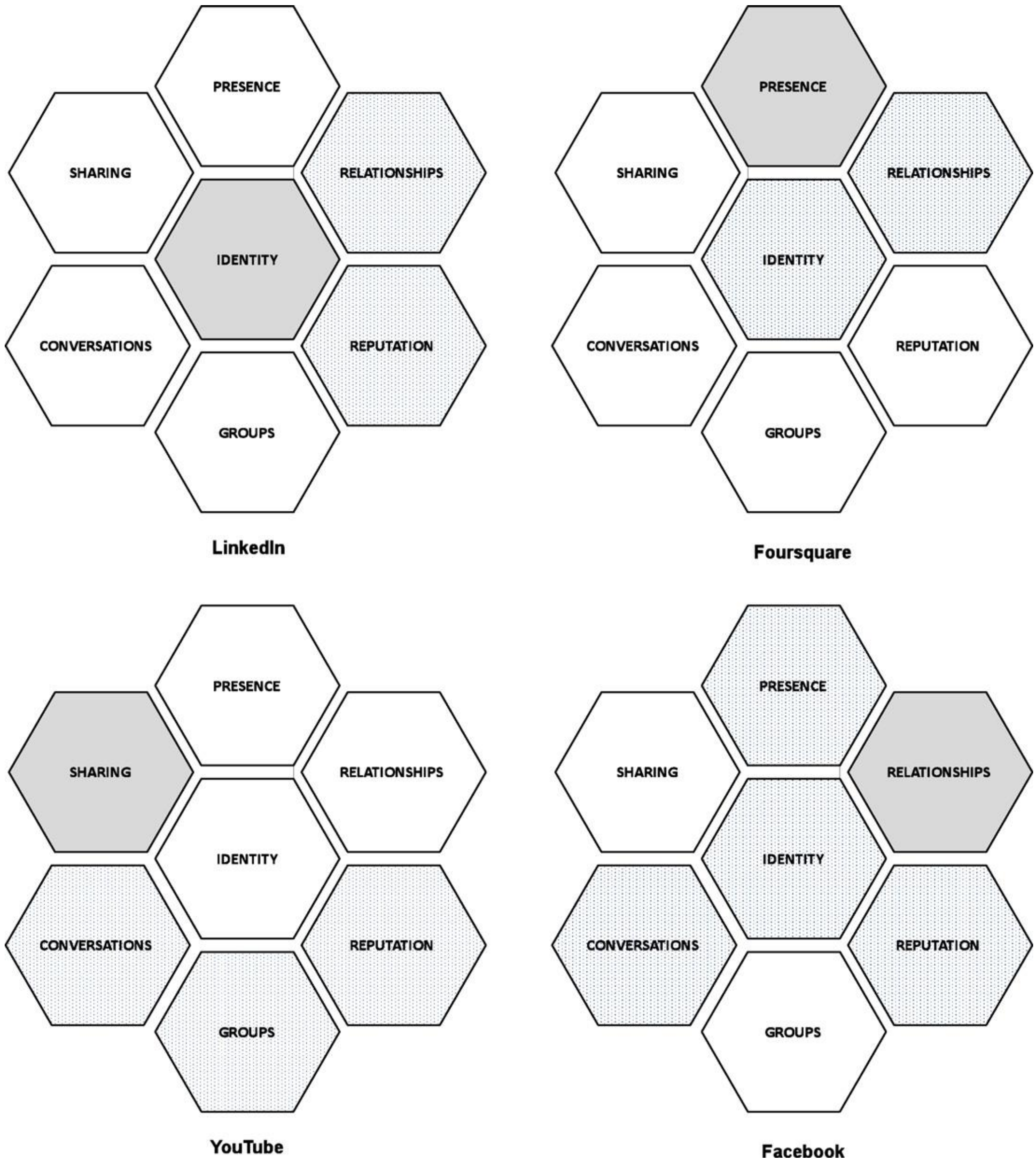


Figure 4.2: The honeycomb framework applied to four popular social media platforms, contrasting their scope and functionality. The dark shade reveals the *primary* building block while the lighter shades reveal the *secondary* and *tertiary* building blocks unique to each platform. (Image courtesy: (Kietzmann et al., 2011)).

This framework will allow components embedded in the unstructured mass of the social media content to be discovered and mapped to the organization so that a range of organizational aspects can be seen from this social media. This framework was applied to compare and contrast the functionalities and implications of different social media activities in the selected online communities to understand the most effective feature to engage with. The honeycomb framework was applied to identify the areas or building blocks of our selected social media (public online discussion forums) to utilize to prepare the dataset for the research study. The application disclosed that in terms of public online forums, “Conversation” and “groups” are the building blocks or areas where organizations should engage with in this social media for maximum impact (figure 4.3).

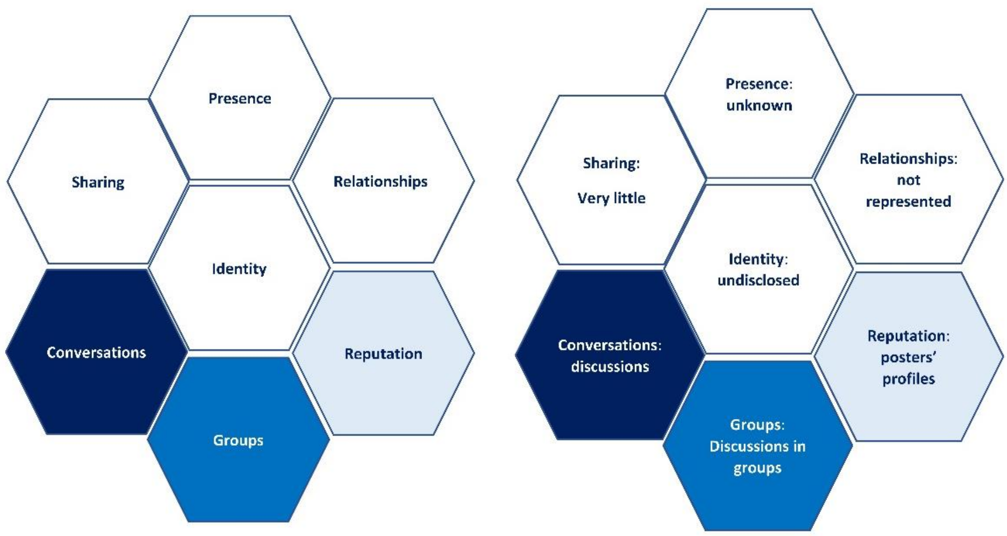


Figure 4.3: Results of Honeycomb social media framework applied to public online discussion forums; revealing the primary (dark blue), secondary (medium blue) and tertiary (light blue) building blocks of the social media platform.

Public online discussion forums are a type of social media environment designed to mainly enhance conversations. Hence, the dataset was proposed to be extracted and prepared from ‘conversations' in ‘groups’ (present in the public forums) relevant to the domain of the organization.

For example, when applied to our case study organization domain; higher education, our investigation revealed that for a higher education institution (such as a university), the dataset should be prepared from conversations in groups related to higher education. Hence, to demonstrate our framework and technique, the testing dataset was extracted from the recommended conversations in higher education groups of three of the most widely used public forums in Australia. Three different social media platforms were selected from the same category; public online discussion forums to provide validation to the presented techniques and results. As the three platforms; Whirlpool, Bored of Studies and ATAR notes were selected as all three of these platforms corresponded to the honeycomb framework presented in figure 4.3 with conversations and groups being prominent while being the most popular and widely used public online discussion forums related to higher education in Australia.

The composition of the final dataset consisted of 44% from the “Bored of Studies”, 41% from “Whirlpool” and 15% from “ATAR notes”. This was designed to correspond with the variation in the number of conversations (the amount of data present) available for extraction in each platform. The combined and completed dataset consisted of over 1.2 million text posts.

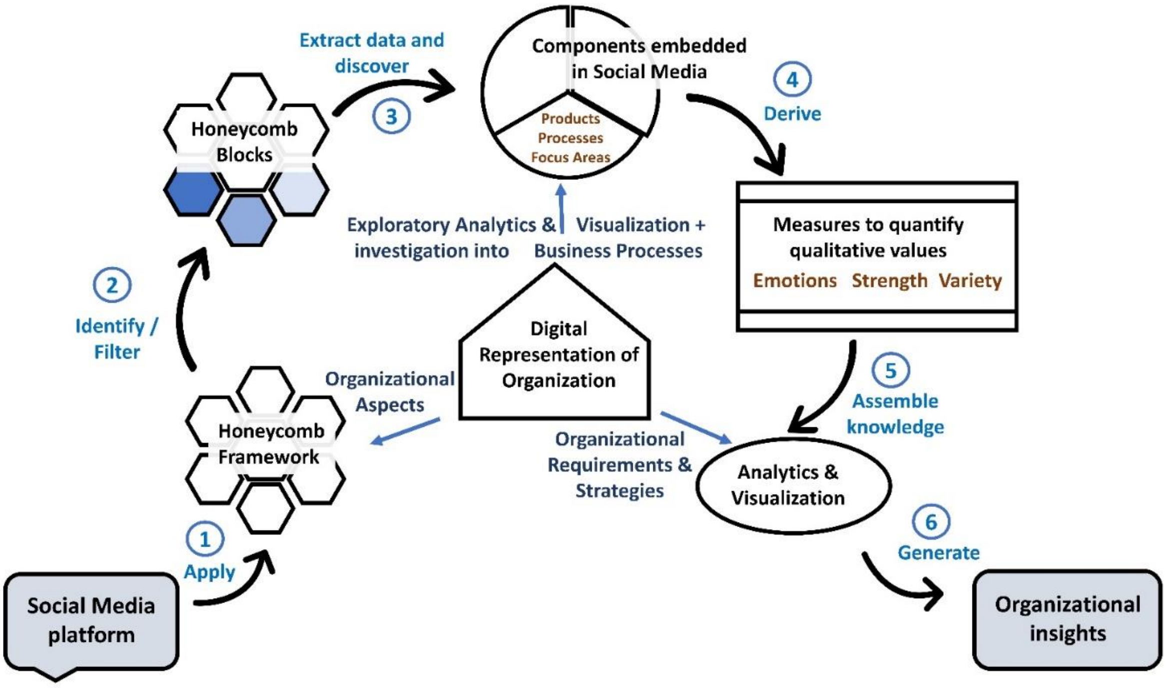
4.3) Social Media to Organizational Insights: A Transfiguration Framework

As discussed above in sections 4.1 and 4.2, the main challenge for understanding the *Perceived Organization* from social media is the lack of a generic framework to structure and transform information buried in social media data into organizational insights. To address this problem, we introduce a novel framework which the authors call the *Social Media to Organization Transfiguration Framework* (figure 4.4). The proposed framework consists of several phases which are explained in detail in the subsequent sections of this chapter. The framework has been tested using the above-mentioned social media dataset of over 1.2 million text posts and has been proven to be capable and effective at transforming social media into organizational insights related to the *Perceived Organization*.

4.3.1) The phases of the proposed framework (illustrated in figure 4.4):

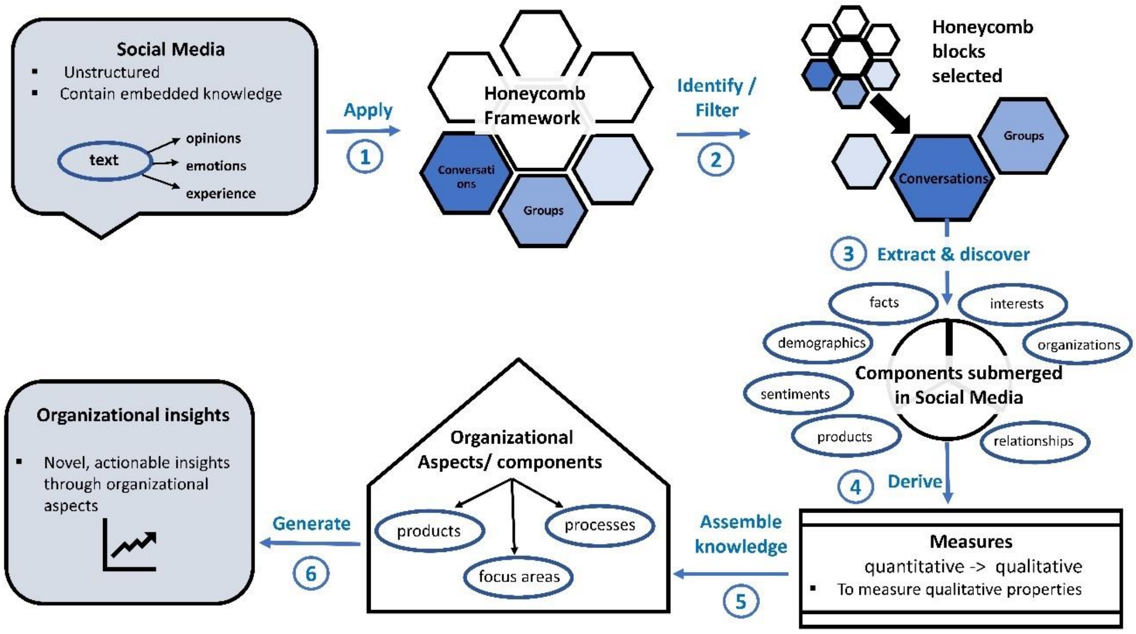
1) Application of the Honeycomb framework and organizational requirements to the chosen social media platform.
2) Identification and extraction of the social media building blocks related to the digital representation of the organization.
3) Extraction of data from social media using the identified building blocks and performing exploratory analytics and visualization together with knowledge from investigation into the business processes of the organization to discover components related to the organization embedded in social media.
4) Deriving metrics to measure the discovered components from the above phase 3, by using techniques to quantify qualitative information of value to the organization.
5) Assembling the now measurable components into knowledge aligned with the organization's requirements and strategy, using analytics and visualization.

6) Generating organizational insights to understand the *Perceived Organization* and benchmark.



Fugure 4.4: Social media to organizational insights - Transfiguration Framework : A diagramatic representation of the proposed process of structuring and transforming raw social media data into measurable features of the *Perceived Organization* to generate organizational insights. As *input,* the framework is fed-in with features and raw data of the chosen social media platform together with the features and strategic needs of the organization. Then, the framework follows through six phases of structuring and transformation and finally generate organizational insights as *output*.

The figure 4.5 provides a more detailed, expanded view of the Social media to organizational insights - Transfiguration Framework presented in figure 4.5. This diagram provides descriptive versions or lenses for each of the phases and components in the Social media to organizational insights - Transfiguration Framework.



Fugure 4.5: Social media to organizational insights Transfiguration Framework: an expanded and detailed view, displaying descriptive versions of each of the phases and components of figure 4.4.

The following sections, 4.4 to 4.10 describe each of the six phases of the *Social media to organizational insights transfiguration framework* in detail with the corresponding experimental results relating to each step.

4.4) [Phase 1] Application of the honeycomb social media framework and organizational requirements

Once the social media platform of value (related to understanding the *Perceived Organization*) is selected, phase 1 of the transfiguration can be performed. For this phase to succeed two processes need to be executed hand in hand concurrently and alternatingly with one providing value to the other process and vice versa to get the best value. The two processes are;

(1) Investigation and understanding of aspects of value to the organization (organizational requirements).

and,

(2) Application of the six dimensions of the honeycomb framework presented in figure 4.6 to the chosen social media platform to identify organizational aspects that may be explored from the social media data.

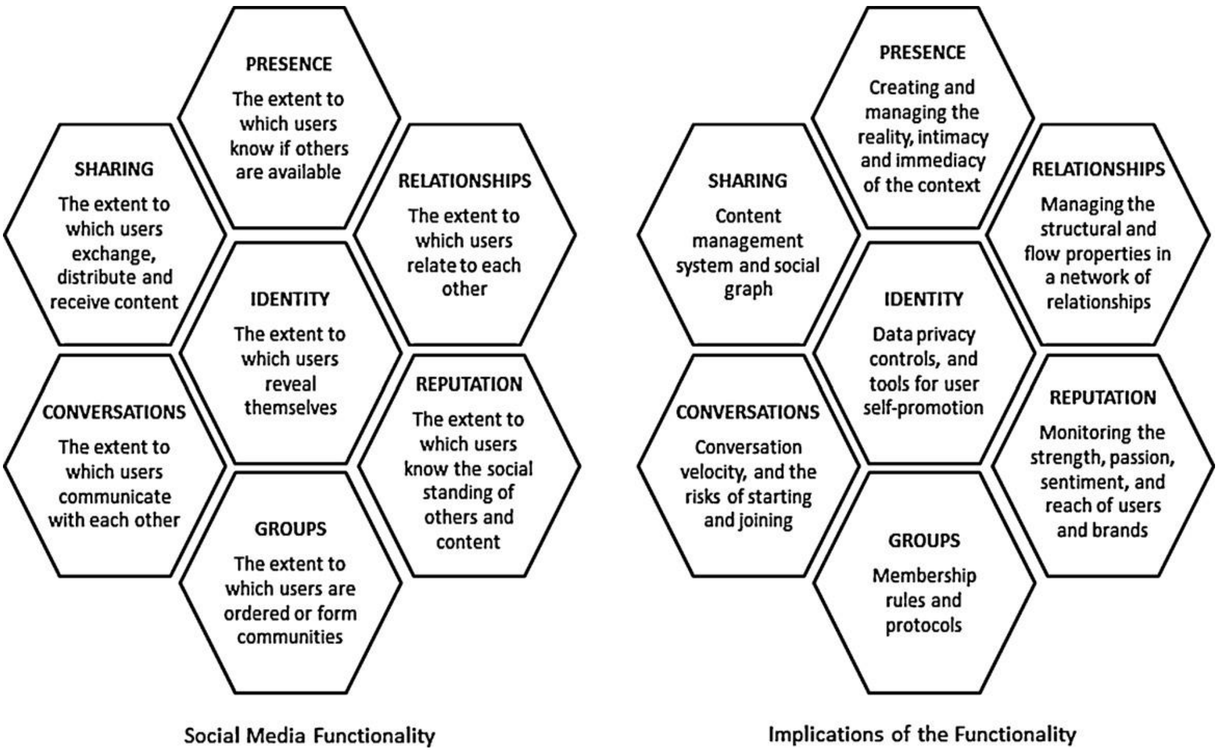


Figure 4.6: The honeycomb of social media (image credits: Kietzmann et al. in (Kietzmann et al., 2011)) introducing the 7 functional building blocks present in social media platforms (left) and their implications (right).

(1) Investigation and understanding of aspects of value to the organization/ organizational requirements

Organizational input was obtained by coming up with questions to explore from the organization’s perspective. This involve questions related to benefits for the organization.

Some examples of questions explored for experiments using the case study organization; The La Trobe University are:

a) How can the student experience be improved?

b) How can the student management process be improved?

c) How can La Trobe university be more competitive?

d) What are the courses that are doing well? And what are not doing so well?

e) What processes are functioning smooth?

f) What are the focus areas of strength?

g) What other unknown areas need attention?

h) Where can the university improve to make a difference?

i) Understand the relationship between products (e.g. courses) customers (e.g. students, parents, potential students, alumni) and processes (e.g. enrolment, transfers, graduations)

j) How can varied focus Areas of interest be evaluated? (e.g. teaching, applications, prestige, rankings)

(2) Application of the six dimensions of the honeycomb framework to identify organizational aspects that maybe explored from the social media

The social media is investigated in this step in detail to understand what data is present. For example, if there are text conversations, are there:

- Querries/questions?
- Answers to queries?
- Recommendations?
- Opinions?
- Experiences?
- Advice /suggestions given?

To explore these aspects in more detail the selected framework; the honeycomb is applied to the social media platform. The social media was viewed from each of the seven dimensions or building blocks of the honeycomb framework and a mapping was developed for each dimension. The following table (table 4.1) presents the results of this mapping. A description of the seven functional building blocks of public online discussion forums and each of their significance for an organization in the case study domain; higher education is uncovered.

The honeycomb building block	Observations through application for Public forums and Higher Education
Conversation	Discussions are the highlights of the community
Groups	Conversation takes place in groups: Groups under which community discuss issues, joining of relevant groups by users, groups based on users' interests.
Reputation	Reputation of posters as a contributor to the community conversations, reputation in terms of the quantity of conversations participated in. Posters have profiles: number of posts, number of replies, profile badge (assigned by website forum) etc.
Identity	Identity of individuals in the community is insignificant and not revealed. They have a profile specific identity-still insignificant
Presence	current online/location status of users is not relevant, significant or known.
Relationships	relationships between users not represented. just ask or reply to queries.
Sharing	very little content shared, mainly online links to direct other users towards information relevant to the discussion, no personal items shared

Table 4.1: Application of the honeycomb to public online discussion forums and the higher education domain. Table presents the results of the application; a description of the seven functional building blocks of public online discussion forums and each of their significance for higher education organizations.

4.5) [Phase 2] Identification and extraction of the Social media building blocks

One of the main benefits of application of the honeycomb social media framework, as described in section 4.2 above is the identification of the three main, most significant building blocks of a social media platform. The honeycomb framework has been proven to be universally applicable across all types of social media platforms. By applying the honeycomb results obtained from Phase 1 (section 4.4) demonstrated in table 4.1, a unique honeycomb can be obtained for any new social media framework as depicted in figure 4.2 above. In our experiments, this application resulted in a novel honeycomb for public online discussion forums (presented in figure 4.3) illustrating the three main building blocks of public forums as follows:

Primary building block: Conversations

Secondary building block: Groups

Tertiary building block: Reputation

The next step is to perform a mapping for components discoverable through each of the main building blocks of social media into the requirements of the organization.

Tables 4.2 to 4.4 presents some sample results from our case study application using public online discussion forums and the La Trobe University as a higher education organization.

CONVERSATION Components	Component in detail
1)Organization/universities	List of Australian Universities *
2)Products/courses	List of courses in Australian Universities *
3)Departments/faculties	List of departments, faculties, and colleges of universities*
4)Processes	List of university processes*
5)Infrastructure	(buildings in the university: DWB, MB etc.)
6)Focus areas	List of Focus Areas for Universities in Australia*
7) Stakeholders	Students: undergrad, postgrad, mature age &/or (potential, current, Alumni)

* See table 4.11 in chapter appendix for the lists

Table 4.2: An example from a case study application: Mapping of components of social media (public online discussion forums) into the organization (higher education organization) from the lens of the honeycomb building block "Conversations". For public online discussion forums this represents discussions which are the highlights of this social media community. The mapping is the result of application of the honeycomb framework and the investigation into the social media data (discussions) and organizational requirements.

The mapping presented in table 4.2 reveal the components of value for the organization that can be discovered through the conversations of the selected social media platform. The mapping reveal that information related to 29 other competitor organizations (the list can be found in table 4.11 in the chapter appendix) in the business domain of interest can be discovered through the social media. Furthermore, information related to units in the business such as departments and buildings as well as stakeholder groups, all of the offered products etc. can be explored. Moreover, information related to the organization's requirements in terms of business processes and strategic focus areas can be discovered through this social media content (table 4.2).

GROUPS Components	Component in detail
1) Interest/ discussion	Education forum
2) Geographical location	Victoria
3) Circumstance	Need to transfer, Low ATAR, looking to study law
4) Age	Undergrad, Postgrad, Mature age
5) Past experience	Past experiences of stakeholders in relation to the organization
6) Expertise	Expertise in a discipline: Worked in university administration, University staff, student advisers
7) Customers/students	Potential, current and past students

Table 4.3: An example from a case study application: Mapping of components of social media (public online discussion forums) into the organization (higher education organization) from the lens of the honeycomb building block "Groups". For public online discussion forums this represents the *secondary* building block: varied interest groups within the social media community structure. The mapping is the result of application of the honeycomb framework and the investigation into the social media content (groups) and organizational requirements.

The results of the application of the secondary building block (groups) using the case study organization is presented in table 4.3. Groups in public forums can be observed through the components or dimensions of interest, geographic, personal circumstance, age, past experience, expertise and seeking advice (table 4.3). The output mapping of this building block clearly demonstrates a very different lens to that of the primary building block.

REPUTATION Components	Component in detail
1) Organizations	List of Australian Universities *
Departments	List of departments, faculties and colleges of universities*
3) Products	List of courses in Australian Universities *
4) Processes	List of university processes*
5) Infrastructure	Library, Sports complex, DW Building
6) Focus areas of interest	List of Focus Areas for Universities in Australia*
7) Users	Forum user profile
8) User groups (High level sub forums to lower level topics)	List of user groups*

* See table 4.11 in chapter appendix for the lists

Table 4.4: An example from a case study application: Mapping of components of social media (public online discussion forums) into the organization (higher education organization) from the lens of the honeycomb building block “Reputation”. For public online discussion forums this represents the *tertiary* building block: representing the reputation of the social media content creators in the online community. The mapping is the result of application of the honeycomb framework and the investigation into the social media content (profiles and text posts) and organizational requirements.

Finally, the mapping of the tertiary building block: Reputation for the case study organization is presented in table 4.4. This mapping presents another additional perspective into the dimensions of knowledge hidden in the social media by presenting a completely new lens compared to that of primary and secondary honeycomb blocks (tables 4.2 and 4.3). The reputation related to the organization as a whole, its competitor organizations as well as the reputation across components at different granularities across the organization is shown to be contained in the social media.

This demonstrated potential to identify and discover diverse components and facets of the business organization through social media, enable comprehensive organizational insights to be generated related to varied organizational perspectives/aspects through analytics and visualization in the impending phases.

4.6) [Phase 3] Extraction of data from social media and discovery of components of value to the organization

Extraction of data

Using the knowledge gained from phase 2 (section 4.5) related to the functionalities of the social media for maximum impact, the data set is extracted using a combination of the three main building blocks.

For our experiments we extracted conversations as discussion posts (obtained from the primary building block) from groups (obtained from the secondary building block) related to the domain of the case study organization; higher education from three public online discussion forums as elaborated in section 4.2 above.

Components embedded in social media

Through the knowledge gained from phases 1 and 2 we identified that using public online discussion forums, stakeholder opinions related to four different perspectives of the organization can be observed. Namely;

1) Products
2) Processes
3) Focus areas
4) Competitor organizations

We present these four perspectives as four main categories for components related to the organization discoverable through this type of social media.

An example is presented below from the application of the four perspectives to the case study organization:

a) Processes: Key processes - Enrolments, Transfers, Withdrawal/defer (course withdrawals), other processes
b) Focus areas: Fees, Employability
c) Products: Courses, Subjects
d) Competitor organizations: Comparisons between universities related to a, b and c above)

These components were identified by exploring the social media data set using exploratory analytics and visualizations (such as graphs, word clouds etc) and through input from a detailed investigation into the case study organization's business processes and requirements (As illustrated in the Transfiguration framework, Figures 4.4 and 4.5).

Extraction of organizational processes

We propose a three-step process for discovering organizational processes.

First, an investigation into the organization's business processes that are aligned with the organization's strategy should be conducted and identified processes categorized and modelled using the digital representation of the organization. Thereafter, this information should be mapped and consulted with the components uncovered by the honeycomb application in phases 1 and 2. The digital representation of the organization consists of internal and external data (organizational + public) related to the organization.

The second step involves identifying stakeholder impressions/perceptions of these processes by exploring the social media dataset generated from the preceding phases using exploratory analytics and visualization. The figure 4.7 demonstrates the resulting model from the investigation into organizational processes of interest related to a student (student as a strategic focus) of the case study organization.

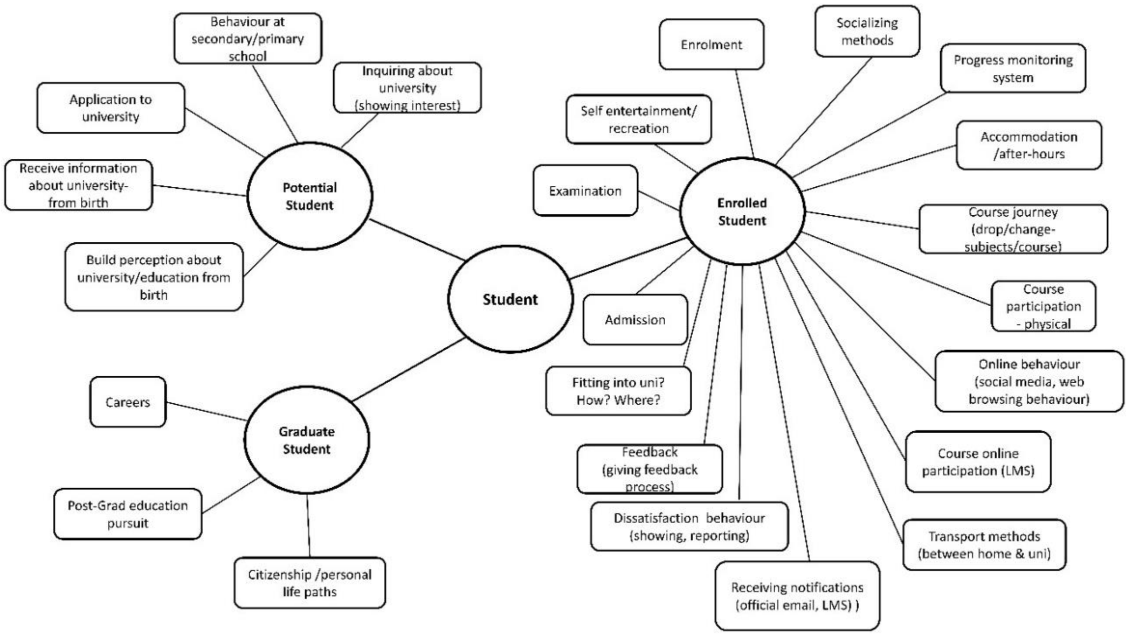


Figure 4.7: Identification and categorization of organizational processes of interest to the organization: Student processes of a higher education organization that are of value to observe. The resulting model is a diagramatic representation of the investigation into student business processes of the case study organization in the higher education domain. Student processes were focussed-on to align with the organization's strategy plan related to the goal of improving the student experience and the student management processes.

Sub process identification through word cloud:

Another discovery is that a further detailed breakdown of processes into sub processes (where needed) could be obtained by using word clouds on social media text data from the public forums. Following is a list of sub-processes discovered from the social media dataset for the processes *Enrolment* and *Employability* for the case study domain.

Discovered Sub-processes for Enrolment:

offer

acceptance

confirmation

selecting a university

selecting a campus

choosing a course

choosing subjects

Time tabling

Filling/submitting forms

Enrolment variation- enrolment variation form

making contact with university (methods of communication)

cancelation of enrolment/discontinuing

Discovered Sub-processes for Employability:

Employability for different courses, future job prospects

Employability advantage comparing courses and universities and masters/honours & majors

Courses to do to get into specific jobs

Maximizing employability

First job

Skills for employability

marks

Extraction of Focus Areas of the Organization

Focus areas were identified, again using the same three-step process as above for processes. But in this situation a higher emphasis is given to the organizational strategy plan. Then, the components in the organizational strategy plan are mapped to the components brought to light using the application of honeycomb framework (phases 1 and 2 of the transfiguration framework). Thereafter, exploratory analytics are performed on the prepared social media dataset to identify which components can be discovered from the data.

Extraction of Products and Competitors.

This step is relatively more straight forward than the extraction of processes and focus areas, as there is a well-known, named, finite list of components. The components the organization is interested in observing are explored using exploratory analytics to identify whether they are present and if so the amount/ degree of information present in the social media.

For example, for the case study organization, stakeholder impressions corresponding to the array of courses and subjects as well as other universities in Australia (potential competitor organizations) were explored.

4.7) [Phase 4] Deriving metrics to measure and quantify the discovered qualitative components of the organization

Transforming unstructure data into a structured template

The next challenge is transforming unstructured text posts in the public forums into a structured format where these observable components of processes, focus areas, products and competitors can be monitored. The challenge is to come up with a technique to measure these qualitative components in order to convert the embedded information into analytics and insights for the organization.

The first challenge was to put a structure into the unstructured text dataset that will work efficiently in an analytics environment where these organizational components maybe identified, extracted and observed in near real time using analytics dashboards.

We propose a technique consisting of 6 steps; illustrated in figure 4.8 to transform the raw text data (of user posts) into structured data that can be used to extract the *Perceived Organization* in near real time. First, a post element is created (for each post) by extracting each post from the conversations with the time stamp and author IDs (figure 4.8). Second step is concerned with extracting the sentiment related to the conversation in the post element. The next section presents our study into methods of extracting sentiments and emotions. The sentiment is extracted as a numeric value and attached onto the post element (figure 4.8). Next, the current organization and other competitor organizations that each post mention are identified, and the post element tagged accordingly (figure 4.8). This is done using string matching techniques. A list of strings that are used to mention the organization is first derived for each item (university).

For example, for La Trobe university the list of strings was: LTU, La Trobe, Latrobe etc. Word2vec presented in section 2.4.6 in the literature review was used to improve this list of words using word embedding techniques.

Technically, the transformation of data presented by figure 4.8 was performed using an Elasticsearch JSON file structure and tagging each file (*post element*) with all of the information presented in the *extract & tag* table in figure 4.8, using a java program to create the elasticsearch dataset with *final data element*s. In this scenario a single JSON file represented a single final data element (after transformation according to figure 4.8). The elasticsearch indexed dataset was connected to a Kibana visualization dashboard that works efficiently with data that flows in real-time to obtain real-time analytics. Next, the same string-matching technique is used to extract and tag each post element with products, focus areas and processes mentioned in the text post. Thereafter, the constructed *final data element* is used for all proceeding steps and phases of the transfiguration framework (figure 4.4)

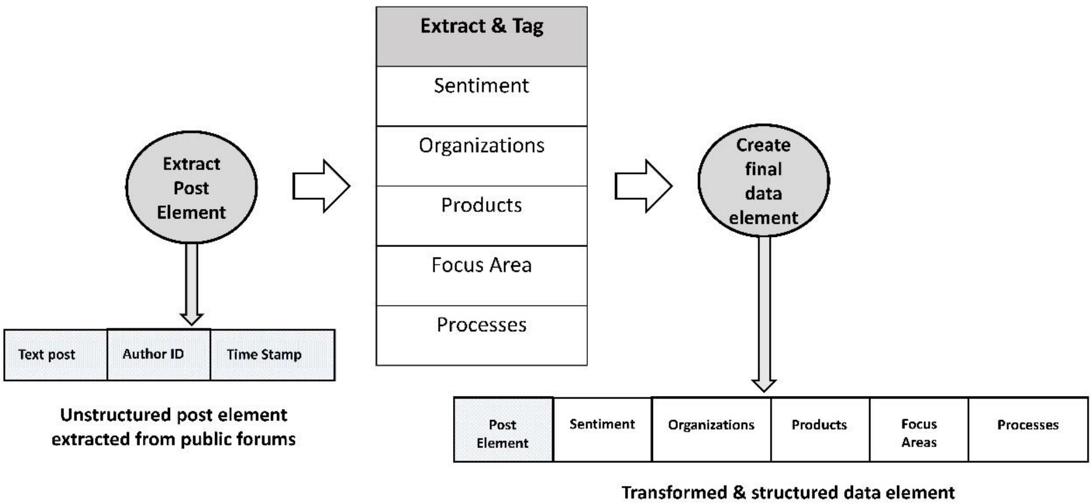


Figure 4.8: The proposed Transformation technique to convert unstructured raw text data available in public online discussion forums (*post elements* in the diagram) into structured data with the extracted knowledge components embedded to support analytics and visualization (*final data elements*).

The table 4.12 in the appendix demonstrate a sample of a set of final post elements generated for the experiments using the procedure illustrated in figure 4.8 above, performed on a public forum discussion dataset and stored in an excel file. This data is structured and ready to perform analytics and experiments on.

Extracting sentiments and Emotions

Several options for extracting emotions were explored from past research and literature as well as through new experiments using the case study dataset. Section 2.4.5 in the literature review presented sentiment analysis and an exploration of methods of sentiment extraction that have been used in past studies and experiments. For the extraction of *Perceived Organization* our experiments revealed that many of these techniques can be used effectively to convert the emotion present in text into a numerical value. Alternatively, the post elements constructed can be directly tagged using the process proposed in figure 4.8 using a sentiment categorization such as {positive, negative, neutral/unrelated}.

We recommend using the widely accepted and used sentistrength algorithm if a numerical value is preferred. Alternatively, the following proposed method can be used for sentiment categorization when a more customized approach is required. In our studies this method was discovered to be suitable for explorations on public forums related to the case study domain; higher education (for exploring public forum discussions on universities).

A proposed novel method to relate processes, products and focus areas with the emotions expressed by the stakeholder in public forums

Method (Illustrated in figure 4.9):

Filter post elements corresponding with the organization, product, focus area or process of interest. Then, filter with a set of words from the emotion index table constructed for the specific organization domain and social media under study (table 4.5), relating to the emotion under observation (Eg: “highly regarded” for a positive emotion exploration). The emotion index table (table 4.5) was created using word lists from the Plutchik’s Wheel of Emotions (Plutchik, 2001) together with words observed in the dataset through exploratory analytics methods such as word clouds.

Plutchik’s wheel of emotions (presented in figure 4.20 in chapter appendix) is one of the most popular emotion wheels that categorize the range of human emotional experiences and presents a visual map for use by researchers. According to Plutchik, humans experience eight core emotions (Plutchik, 2001).

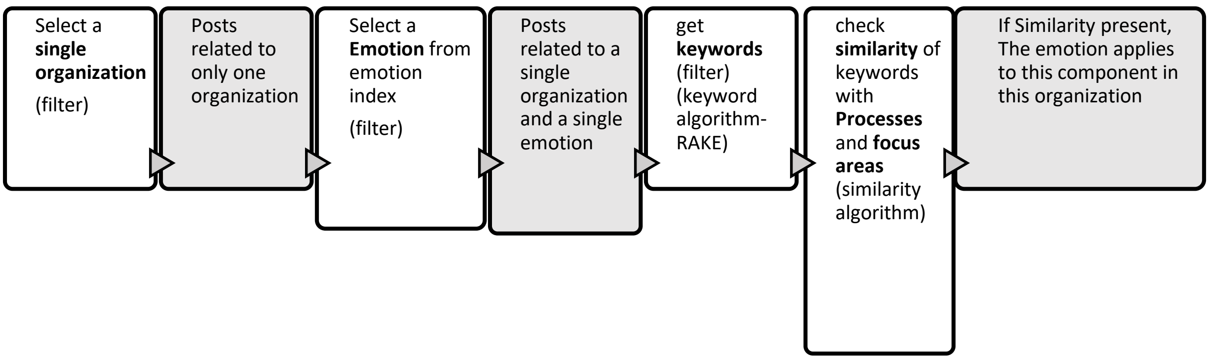


Figure 4.9: Sentiment discovery: proposed method to relate processes, products and focus areas with the emotions expressed by the stakeholders in text conversations in public online discussion forums. Grayscale boxes represent results generated by each step.

As illustrated in figure 4.9 this proposed method first allows you to filter and identify the discussion posts relevant to the organization you are observing. Then using the designed and customised emotion index, posts related to the organization of interest, indicating the emotion under observation (e.g.: positive/happy mentions, angry mentions of an organization on social media) can be screened. Next a key word or a set of key words related to the component that the user wishes to observe(e.g.: teaching, fees) is prepared and passed into a keyword algorithm to generate a list of (possibly unknown) words that denote the same as the user provided key word.

For the experiments in this study, the RAKE (Rapid Automatic Keyword Extraction) algorithm present in the Python NLTK tool kit was used. This algorithm was selected as it is a widely accepted algorithm for the purpose and our experiments revealed it to be highly effective.

Once the list of keywords is generated another filtering is performed using the keywords. This results in a list of discussions that discusses the topic the user is interested in, related to the particular organization, with the tone of the selected emotion.

Next step involves a similarity matching (using similarity algorithm on NLTK) with the list of keywords and the known processes, focus areas and products of the organization. This reveals what organizational processes, focus areas and/or products the selected emotion is related to and the filter reveals the number of conversations that were found on the social media related to this.

This allows organizations to specifically filter and observe social media conversations specifically related to a performance area of an organization. This could reveal areas where the organization need make changes and be proactive to serve and satisfy their stakeholders.

The procedure was tested on a sample dataset categorized by human observation and an accuracy of approximately 90% was observed.

Following is an example application from experiments conducted for this study:

The technique was applied to ACU (Australian Catholic University) as the organization using a dataset from public online discussion forums of higher education.

The application of the method revealed teaching was highly regarded in ACU.

Positive Emotion word list used for the application: "*highly regarded*", "*superb*", "*proud*", "*pleased*", "*best in*”, "*great*" ,"*prestigious*", ”*prestige*”

Wheel of Emotions words used:

Disgusted: Awful, disappointed, disapproving, repelled, appalled,

Happy: successful, valued, thankful, free, accepted, respected, hopeful, optimistic, inspired

Positive	Negative
great (with no negation words in the same sentence)	shit
highly regarded	not great (check reliability of word for mood)
good (exclude-good luck, good point and sentences ending with "?")	not good (check reliability of word for mood)
excited	screwed
thrilled	miserable
superb	worthless
glad	exhausted
pleased	devastated
satisfied	terrible
proud	hopeless
optimistic	horrible
relaxed	frustrated
admired	bad
appreciated	Pissed off
grateful	mad
prestigious	tense
best in	crap
successful	disgusted
valued	awful
thankful	disappointed
top5	Really bad
big4	Really poor
go8	

Table 4.5: Emotion index table generated for public online discussion forums (a simplified version. The list can be extensive and highly domain specific)

Note, certain words like "*good*" have been eliminated from the index as these words needs to have a positive prefix like very, pretty, really to denote a positive emotion.

A sentence structure to use for the extraction of emotions using this method:

If a more intense association is needed, the technique presented in figure 4.9 can be applied on a sentence level rather than on a post level (as the post generally contain multiple sentences and even multiple paragraphs).

E.g.: A positive *<emotion word/phrase from the emotion index>* with the *organization name* and a *product name* in the same sentence gives a ->*positive opinion* related to the specified product of the specified organization.

Deriving measures for components

The next requirement to generate the *Perceived Organization* was to derive measures or techniques to quantify and measure the qualitative components discovered above (processes, focus areas, products and competitors)

Through the application of honeycomb, we were able to view social media through lenses of components of the organization and come up with a mapping or structure to observe social media. These results presented in tables 4.2 to 4.4 above in section 4.5 were further explored and extended to derive measures for each of the components to convert the qualitative information into quantitative components that can be measured and evaluated. The results are presented below in tables 4.6 to 4.8 for *conversations* and *groups* building blocks with a third column; 'measures' included (an extension of table 4.2 in section 4.5 in phase 2).

The results were simplified into a three-component technique to measure any component from public forums:

We present this technique of measuring components as E S V: Emotion, Strength, and Variety of each component.

CONVERSATION Components	Component in detail	Measure
1)Organization/universities	List of Australian universities *	**E S V**
2)Products/courses	List of courses in Australian universities *	**E S V**
3)Departments/faculties	List of departments, faculties and colleges of universities*	**E S V**
4)Processes	List of university processes*	**E S V**
5) Infrastructure	(buildings in the university: DWB, MB etc)	**E S V**
6) Focus areas	List of focus areas for universities in Australia*	**E S V**
7)Customers	Students: UG, PG, Mature Age &/or (potential, current, Alumni)	**E S V**

* See table 4.11 in chapter appendix for the lists

Table 4.6: Deriving measures using the primary building block from the honeycomb framework. The table presents an extension of table 4.2 extending the components discovered using the honeycomb into measures to measure the influence and applicability of each component for the organization using the primary building block "conversations". This extends results from the case study application using public online discussion forums and the La Trobe university as a higher education organization. E S V was discovered to be an effective measure for all components discoverable through the primary building block-*conversations* related to the organization.

E S V stands for; (E) -> Emotion range for components: Very positive to very negative,

(S) -> Strength or frequency for versions of components (e.g.- Latrobe, arts),

(V) -> Variety of sub-topics under a component (e.g.- number of courses discussed for Latrobe)

GROUPS Components	Component in detail	Measures
1) Interest/ discussion	Education forum	Size of group/ number of participants. Length of time of existence of group Number of topics (variety) under the group Number of posts (depth) Velocity of posts (frequency)
2) Geographical location	The states of Australia (Victoria, NSW, SA etc.)	ESV
3) Circumstance	Need to transfer, low ATAR, looking to study law	ESV
4) Age	Undergrad, Postgrad, Mature age	ESV
5) Past experience	Varied components based on requirement	ESV
6) Expertise	Expertise in a discipline, worked in university administration, university staff, student advisers	ESV
7) Customers/students	Potential, current and past students	ESV

Table 4.7: Deriving measures using the Secondary building block from the honeycomb framework. The table presents an extension of table 4.3 extending the components discovered using honeycomb into measures to measure the influence and applicability of each component for the organization using the building block "Groups". This extends results from the case study application using public online discussion forums and the La Trobe university as a higher education organization.

REPUTATION Components	Component in detail	Measures
1)Organizations	List of Australian universities *	Frequencies (eg: number of transfers) Frequencies mapped with emotions (positive comments on a process or course)
1)Departments	List of departments, faculties and colleges of universities*	ESV
3)Products	List of courses in Australian universities *	ESV
4)Processes	List of university processes*	ESV
5)Infrastructure	Library, Sports complex, DW Building	ESV
6)Focus areas of interest	List of focus areas for universities in Australia*	ESV
7)Users	Forum user profile	ESV
8)User groups (High level sub forums to lower level topics)	List of user groups*	Website data and statistics. Profile badges, number of replies by user etc. Number of posts & velocity of conversation in sub forums, groups & topics

* See table 4.11 in chapter appendix for the lists

Table 4.8: Deriving measures using the tertiary building block from the honeycomb framework. The table presents an extension of table 4.4 extending the components discovered using honeycomb into measures to measure the influence and applicability of each component for the organization using the building block “Reputation”. This extends results from the case study application using public online discussion forums and the La Trobe university as a higher education organization.

4.8) [Phase 5] Assemble mesurable components into knowledge

The next challenge is assembling the now measurable components into knowledge aligned with the organization's requirements and strategy. This was done using analytics and visualization. Several experiments were carried out to discover insights related to the components discovered in the previous sub-sections related to the *Perceived Organization*.

Experimental Results

In this section we present experimental demonstrations of the proposed social media to organization Transfiguration framework (Figures 4.4 and 4.5). Experimental results from executing steps 1 to 5 from this proposed social media to organization Transfiguration framework are presented in figures 4.10 and 4.11 below.

Figure 4.10 show the measured strength of organizational aspects: Products, Processes and Focus areas for the case study organization (The La Trobe University) and figure 4.11 shows the same for a competitor organization of the case study organization (The University of Monash). Furthermore, figure 4.12 presents the analytics dashboards.

Courses	Positive Strength	Negative Strength
Arts	69	14
communication and critical enquiry	69	14
Accounting and Business	38	3
Allied health and rehabilitation	32	3
Law	32	5
Medical and Psychological Sciences	29	7
Information Technology	24	2
Dentistry and oral health	20	4
Chemical and Mathematical Sciences	18	4
Physical	18	4
Economics and Finance	16	1
Biological Sciences	13	7
Engineering	13	3
Social sciences	9	1
gricultural and Environmental Science	8	5
Marketing and Management	6	
Visual and creative arts	6	3
Asian studies	2	

FocusAreas	Positive Strength	NegativeStrength
Applications	72	13
Offers	63	11
Prestige/Reputation	29	10
Undergrad:	28	7
Postgrad	20	5
Honours	14	3
Rankings	13	4
Mature Age Students	9	1
Fees	6	2
International Student	5	1
Employability	2	1

Process	Positive Strength	Negative Strength
Transfer	41	9
Enrollment	16	3
Graduation	5	1
Withdrawal_Defer	4	1

Figure 4.10: Positive and negative stakeholder impressions for products, processes and focus areas demonstrating emotion strength for university Courses, Processes and Focus areas of La Trobe University). Demonstrating results of the application of phases 1 to 5 of the social media Transfiguration framework on a public forum data set.

Courses	Positive Strength	Negative Strength
Arts	696	185
communication and critical enquiry	696	185
Accounting and Business	678	156
Law	545	112
Engineering	457	111
Chemical and Mathematical Sciences	330	105
Physical	330	105
Information Technology	264	69
Economics and Finance	262	66
Medical and Psychological Sciences	245	53
Biological Sciences	182	40
Agricultural and Environmental Sciences	96	31
Marketing and Management	93	21
Social sciences	59	18
Asian studies	52	21
Allied health and rehabilitation	46	10
Visual and creative arts	29	11
Dentistry and oral health	23	4

FocusAreas	Positive Strength	Negative Strength
Offers	618	110
Applications	563	120
Undergrad:	301	58
Prestige/Repu	245	41
Honours	215	43
Postgrad	117	27
Rankings	97	26
International S	69	16
Fees	58	8
Employability	26	4
Mature Age St	19	6

Process	Positive Strength	Negative Strength
Transfer	370	77
Enrollment	147	25
Graduation	56	14
Withdrawal_Def	30	8

Figure 4.11: Positive and negative stakeholder impressions for products, processes and focus areas demonstrating emotion strength for university Courses, Processes and Focus areas of University of Monash. Demonstrating results of the application of phases 1 to 5 of the social media Transfiguration framework on a public forum data set.

Figure 4.12: Analytics and Visualization dashboards developed to test the proposed framework using public online forum data (1.2 million text posts) and the case study organization.

4.9) [Phase 6] Generating organizational insights to understand the Perceived Organization and benchmark.

Finally, the steps discussed above (phases 1 to 5 of the transfiguration framework) must lead to generating organizational insights that enable understanding the *Perceived Organization*.

Organizational requirements and strategies were fed into analytics and visualization using the case study organization and it’s strategic plan. Text mining and Natural Language Processing techniques were utilized.

The following potentials were explored:

- Monitoring organizational performance from the organizational aspects
- Identifying relationships between the organization and its components and organizations’ stakeholders. (eg: possibility to extract perceived employability of university students, impression on fees, quality of teaching, prestige of the university etc. from the data set)
- Descovering public preferences of organizations based on stakeholder perceptions of prestige: Ranking universities based on customer perceptions (using Natural Language Processing)
- Identifying specific processes and focus areas of strength in an organization (by using a combination of sentiment and emotion extraction, keyword extraction & process and focus area extraction)
- Public views on specific organizational products as compared to similar products by competitor organizations
- Methods of benchmarking against competitor organizations

4.9.1) Relationships between organizations, products, processes and focus areas

Case study demonstration: Relationships between courses, processes and focus areas

Observation of how the universities in Australia performed in terms of a focus area such as 'teaching' (for the entire period the data was extracted) yielded results demonstrated in figure 4.13. This provides insights on how Monash University is leading in Australia with a few universities from other states right behind it. This provides a tool for higher education organizations to benchmark against competitors related to different focus areas. Similarly, further exploration can be performed to identify which organizations are considered highly regarded in terms of a key focus area in the domain, as per the organization's strategic plan such as 'employability' for higher education (figure 4.14).

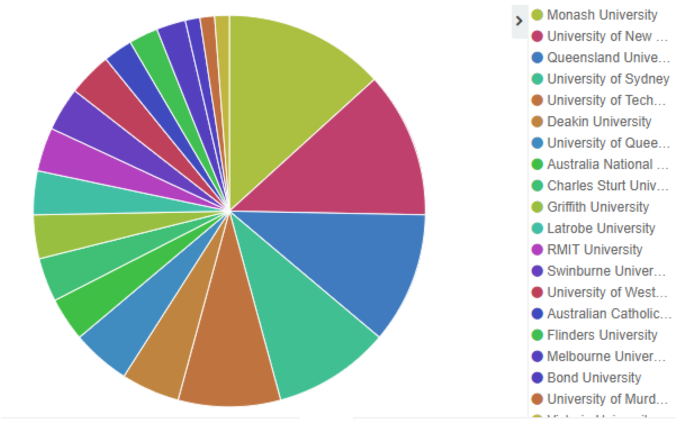


Figure 4.13: 'Teaching' focus area for universities in Australia (Monash leading with positive sentiments) from public forums.

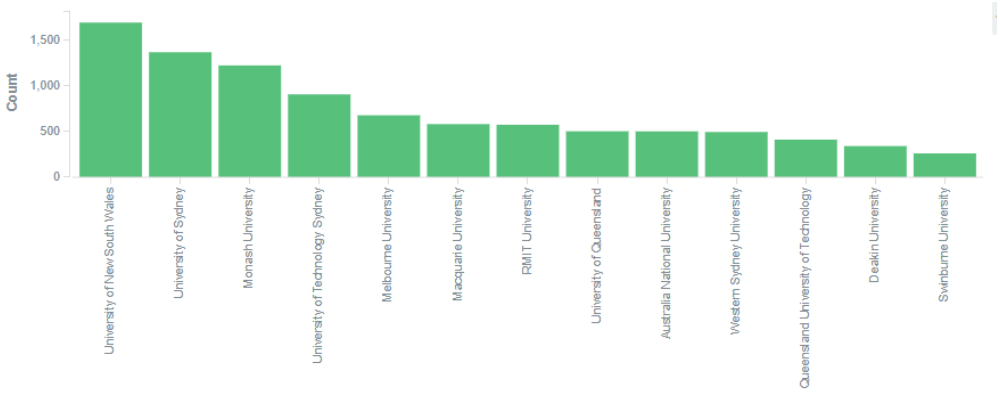
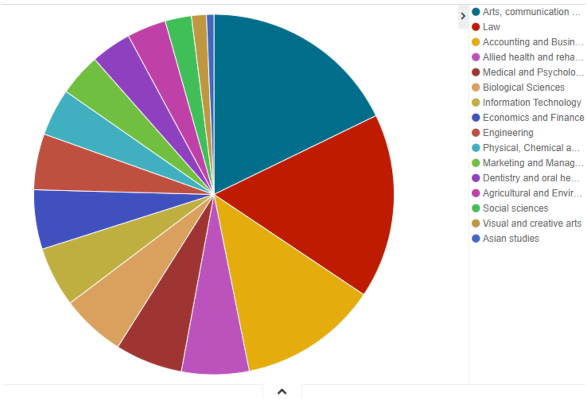
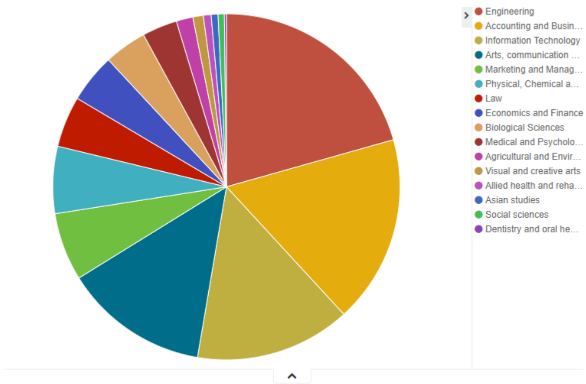


Figure 4.14: Universities where the aspect of 'employability' was discussed in a positive light in Australia (illustrates a few New South Wales universities leading with the University of Monash ahead of universities in Victoria).

Figure 4.15 below illustrates the possibility of comparing the most popular products across organizations. The diagram illustrates a comparison of courses which appear most popular in the case study organization; the La Trobe university, compared with a competing and neighbouring university, RMIT. RMIT is widely believed to have a reputation for technology related courses. This is confirmed with the results from data with La Trobe university displaying Arts on top and RMIT displaying Engineering in the lead. This allows further monitoring of fluctuations over time.



La Trobe University



RMIT

Figure 4.15: Courses that were discussed in a positive light related to La Trobe University and RMIT University.

Our experiments into observations of how well individual organizations are performing in terms of the functioning of their processes revealed information that is not predictable beforehand.

For example, Australian Catholic University which is an equity university had ‘offers‘ being extremely popular among stakeholders (Figure 4.16) and the data set further revealed that while this university does not carry a high reputation, it is highly regarded for teaching.

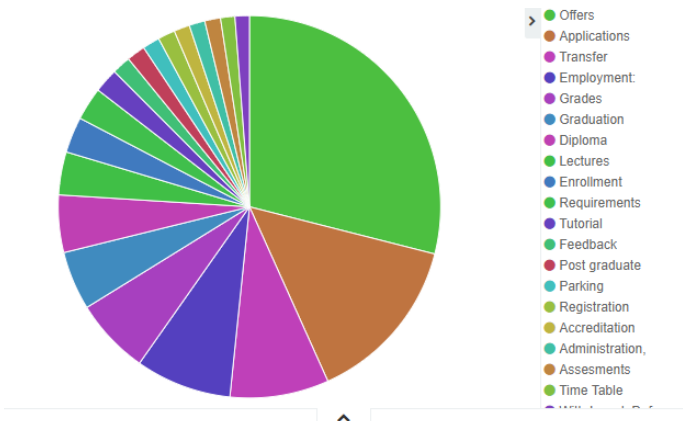


Figure 4.16: Well-functioning processes in the Australian Catholic University, illustrates 'offers' as most popular among stakeholders.

4.9.2) Ranking products and organizations based on perceived popularity

This section presents another possible application of the social media to organization Transfiguration Framework (presented in figures 4.4 and 4.5) to generate another layer of organizational insights by presenting a ranking system and an algorithm for organizations and products based on the *Perceived Organization*. The case study domain was used to explore the possibilities as follows.

Case Study demonstration: Courses in which students consider transfers in Australia

The Transfiguration framework's potential to provide organizational insights to rank products across organizations in a business domain using perceived popularity through social media is explored and the results presented.

The figure 4.17 presents the application of the Transfiguration Framework for organizations to discover most popular products for customer churn, or the products which are most at risk of losing a customer to a competitor. The insights extracted from the public forum data show that in the higher education sector in Australia, ‘Law’ is the course most considered by students for changing universities (Figure 4.17). The products were ranked by discovering matches of post elements with the course and ‘Transfer’ process occuring together and by using the E S V mesuring system introduced in section 4.7.

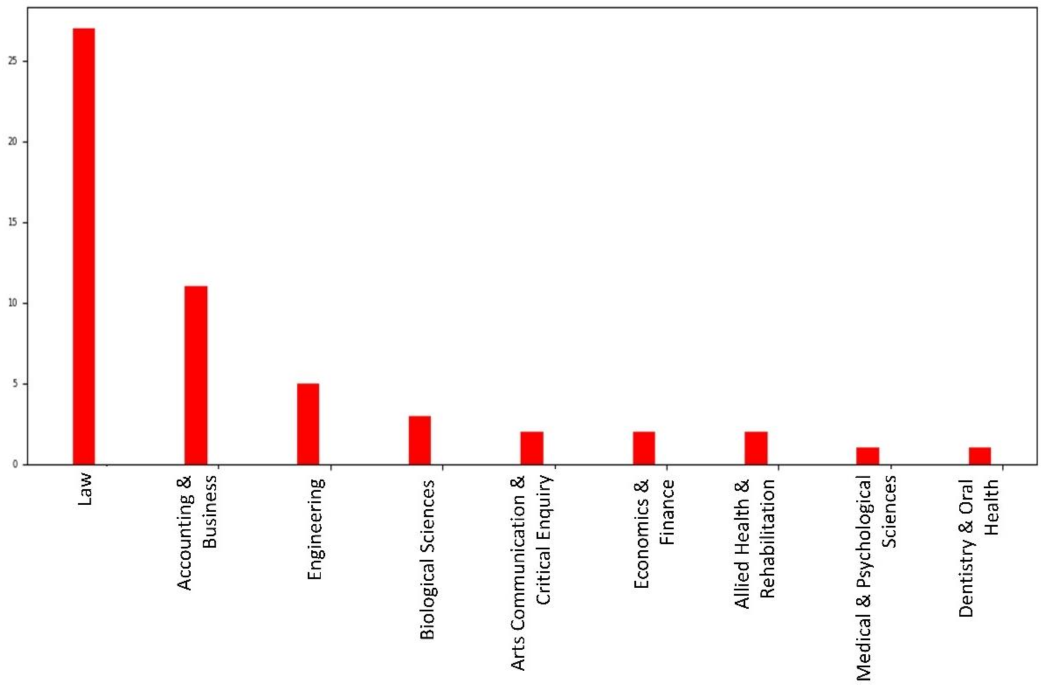


Figure 4.17: Courses in which students consider transfers between universities in Australian higher education sector extracted from public forums: The graph ranks courses (products) which are most popular for transfers.

However, the ESV system was shown to be insufficient to rank and compare organizations together with products due to the unstructured nature of text posts and challenges inherent to NLP. Hence, an algorithm was developed, tested, and presented as follows.

4.9.3) Developing an algorithm to uncover comparisons between organizations and products

The proposed algorithm was developed using a significant process observed in the organization domain from the social media dataset. This was done through investigating the context of this process extensively from the discussions present in social media. The significance and implications of this process was further explored using analytics and visualizations performed on the social media dataset. The resulting algorithm uses patterns of keywords and string matching to discover organizations and its relationships to the key process and was tested using the data set (1.2 million posts) and the case study organization. For experiments using the case study domain, the process “transfer” was selected as the key (significant) process to compare and rank organizations and products. Discussions on *transfers* in the public online discussion forums were revealed to imply students discussing their preference and potential to transfer between universities in Australia for varied courses of study.

Key categories of Transfers presented in the social media- categories for posts:

The following categories were identified after an investigation into the social media discussions and this categorization was used to develop the algorithm’s classification step.

a) Transfers not mentioning any university

b) Transfers mentioning one or more universities

c) Transfers mentioning universities and courses

Algorithm to rank organizations (demonstrated using the case study application)

Step 1 (Extraction/filtering): Get the posts that have “transfer” as a process (from data transformed and prepared as per figure 4.5)

Step 2 (Classification): Find and categorize post elements into categories *a, b, b1* and *c* using the following construct (order of text in a sentence)

Sentences with the following order of words:

<Transfer Process keyword> “to|into” <course> “at” <university X> →(a)

||

<Transfer Process keyword> “to|into” {next 5 words} <university X> →(b)

||

<Transfer Process keyword> “to| into” <university X> “or” <university Y> →(b1)

||

<Transfer Process keyword> “to|into” <university X> “from” < university Y > →(c)

Some notes:

- X and Y stand for text representations of higher education organizations.
- In case b1 the order of universities indicate the order of popularity or prestige (revealed from our investigation into social media conversations).

  E.g.: **X** > **Y** > … <any other university mentioned in the same post> means X is the most prestigious university out of the set of universities mentioned.

Step 3: (Measuring) Providing weights to measure intensity

- Give weights to universities (and consider the total weight added ultimately)
- Case *b* or *a*: university X gets weight 1.0
- Case *b1*: university X gets weight 1.0 university Y gets weight 0.75 etc.

All weights will be added in the end and a list created

For example, if the values are as follows;

University Name	Total weight
A	678
B	798
C	78

Accordingly, university B will be chosen as the top ranked university.

The algorithm has the capacity to rank products across organizations in a domain using perceived popularity through social media.

Case Study Demonstration: A university ranking system based on popularity (using how students consider transferring to and from universities)

The possible applications of using the above presented algorithm for ranking organizations based on popularity presented in *Perceived Organization* is explored.

The algorithm was tested and demonstrated to be successful at identifying the universities students are *transferring into* and the universities students are *transferring from* with an accuracy of 94%. Tables 4.9 presents the results of the application of step 3; calculating weights to measure.

The testing accuracy was obtained through performing manual testing on a sample. Figures 4.18 and 4.19 presents the final results of the application of the proposed algorithm for the case study organization. The results show the success and the potential of the algorithm to come up with valuable insights that allow near real-time benchmarking for the organization using social media.

Results generated by the algorithm for Step 3

University	Generated total weight
University of New South Wales	176
University of Sydney	136.75
Monash University	85
University of Technology Sydney	55.5
Macquarie University	36
Western Sydney University	29.25
Melbourne University	20
Australian Catholic University	15
Deakin University	10.75
RMIT University	10
University of Canberra	7
Bond University	6
Griffith University	5.75
University of Queensland	5.75
Australia National University	4.75
University of Newcastle	4
University of Western Australia	3
Latrobe University	2
Charles Sturt University	2
Flinders University	2
University of Murdoch	2
University of South Australia	2
Queensland University of Technology	2
Central Queensland University	1

Table 4.9: An example of intermediary results generated (from step 3) of the ranking algorithm applied on the higher education domain.

All added weights generated by the algorithm for 23 universities in Australia presents the potential for using the technique to obtain a successful ranking for universities.

Results: University Rankings for Australian Universities based on student preferences (using transfer process)

	A
1	University of New South Wales
2	University of Sydney
3	Monash University
4	University of Technology Sydney
5	Macquarie University
6	Western Sydney University
7	Melbourne University
8	Australian Catholic University
9	Deakin University
10	Australia National University
11	RMIT University
12	University of Canberra
13	Bond University
14	Griffith University
15	University of Newcastle
16	University of Queensland
17	University of Western Australia
18	Latrobe University
19	University of Murdoch
20	University of South Australia
21	Charles Sturt University
22	Flinders University
23	Queensland University of Technolog

Figure 4.18: Results from application of the proposed ranking algorithm on public forum data. A ranking of higher education organizations in Australia for the case study organization (Latrobe university) to monitor how its position change in near real-time using perceived social media.

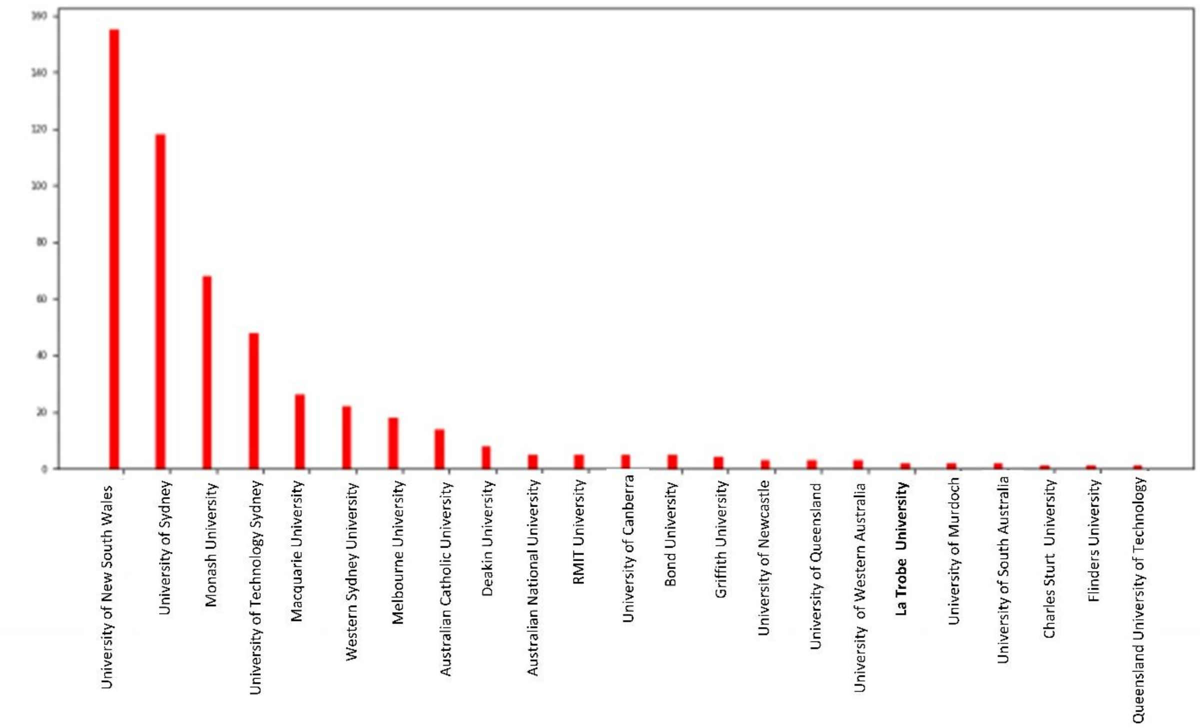


Figure 4.19: Results from application of the proposed ranking algorithm on public forum data. A ranking of higher education organizations in Australia for the case study organization (Latrobe university) to monitor how its position change in near real-time as well as the power or strength of its position among competitors using perceived social media. The results show the University of New South Wales in the lead.

Examining the results generated by the algorithm presented in figures 4.18 and 4.19 revealed its effectiveness. As illustrated with the results, this algorithm can provide organizations with a tool to benchmark against competitors in near real-time by monitoring Perceived popularity of its products and the organization as a whole.

Observations from the experiments on the dataset:

Out of the 1.2 million posts in the dataset, 39,000 posts were filtered by step 1 of the algorithm as relating to the *transfer* process. Samples for all four cases (*a,b,b1,c*) generated by step 4 of the algorithm were tested manually to check accuracy. Overall a 94% accuracy was observed for the classification.

A sample of the classification results set generated by the step 2 of the proposed algorithm

Following is a sample of a results set (social media posts) classified by the algorithm into the four cases/categories of transfers as per step 2 of the proposed algorithm. The underlined sections of the post demonstrate the applicability of the post to the case.

Case A:

1. Case A: Transfer to bachelor of arts Has anyone had any experience transferring to bachelor of arts at macquarie university from another uni? If so how did it work for you and what requirements did they need? GPA etc
2. Case A: arts at unsw or eco at macq I am wanting to transfer into commerce at unsw next yr, I received an offer for arts in the main rd and eco a macq in the late rd
3. Case A: Will I get all the subjs credit tranfered if I do economics at macq and transfer to commerce at unsw next yr

Case B:

1. Case B: when do transfer offers come out Lucky!! how did u get unsw offer soo early? lol im wanting to transfer to uts
2. Case B: And everyone knows that RMIT offers a much more hands on experience and direct pathway regarding engineering; wouldn't that be what employers are looking for? To be honest, I started to fully question myself when a friend of mine [dux of his school/easily one of the smartest people I personally know] doing biomedicine at Melbourne Uni decided to transfer to RMIT
3. Case B: He wants to transfer into Melbourne after the first year and wanted to know if these subjects would be cross credited

Case B1:

1. Case B1: I am doing a bachelor of business - accounting but want to transfer to either macq, uts or uws doing accounting

2. Case B1: transfers as in uws they are doing a common first year if i was to transfer to aeronautical at unsw or syd i wouldnt have to repeat
3. Case B1: UNSW test your best option is probably doing a year at UWS or somewhere else with low uai cutoffs, and then transferring to a better uni like unsw or usyd after a year with good uni marks
4. Case B1: Currently undertaking first year uws engineering, want to transfer to either unsw or usyd at the end of this year using uni marks

Case C:

1. Case C: Whos going to Blacktown/Parramatta Campus? ohh man why did u transfer to uws from usyd for? u should haf transfered to UTS if u wanted exemptions for your IT courses
2. Case C: Arts/Science transfer into Law I want to try and transfer into law at unsw from unsw in 2010, but I am not sure which degree to combine it with
3. Case C: I'm also transferring to uts from usyd
4. Case C: i'm planning to transfer to unsw from usyd after 1 full year of study
5. Case C: I got a 70 atar and want to transfer to commerce at usyd from uws
6. Case C: Anyone do any kind of bachelor degree in sports?? Looking to maybe transfer to deakin course from monash (doing commerce atm)

Results for comparison of organizations (using case 'C'):

An example of transfers between organizations (from one university (d2) to another more preferred university (d1)) revealed by the case 'c' classification is demonstrated in table 4.10.

	d1(prefer to transfer into)	d2(prefer to transfer out of)
Australia National University	4	0
Australian Catholic University	15	0
Bond University	6	0
Central Queensland University	1	0
Charles Sturt University	2	0
Deakin University	11	0
Edith Cowan University	0	0
Federation University	0	0
Flinders University	2	0
Griffith University	5	0
Latrobe University	2	0
Macquarie University	36	0
Melbourne University	20	0
Monash University	81	1
Queensland University of Technology	2	0
RMIT University	10	0
Swinburne University	0	0
University of Canberra	7	0
University of Murdoch	2	0
University of New South Wales	164	1
University of Newcastle	4	0
University of Notre Dame	0	0
University of Queensland	5	0
University of South Australia	2	0
University of Sydney	132	4
University of Tasmania	0	0
University of Technology Sydney	56	1
University of Western Australia	3	0
Victoria University	0	0
Western Sydney University	28	2

Table 4.10 An example comparison table obtained through the application of the algorithm (using the case 'C') on the higher education domain. Results demonstrate the strength of d1 and d2 for universities

4.10) Chapter Summary and Discussion

This chapter presented a research endeavour into techniques to view and comprehend the organization from a window of social media to address the research question; "*To what depth can knowledge and value be captured from the digital environment and social media to understand communication and interactions between organizations and individuals*".

The chapter proposed a novel framework titled, "*Social Media to Organizational Insights: Transfiguration Framework"* and a series of techniques and algorithms to be used together with the framework to understand the stakeholder perception about an organization (Perceived Organization) from social media. The chapter built on the Digital Stakeholder presented in chapter 3 and introduced the said novel framework utilizing the honeycomb social media framework, text mining techniques, Natural Language Processing, exploratory and visual analytics and feature engineering to structure and transform social media into lenses or aspects of the organization and generate organizational insights.

The chapter presented a detailed unpacking of organizational features and aspects represented and discussed over social media in the form of public online discussion forums. Techniques were presented to leverage the social media , remodel unstructured text data and monitor the organizational perception from varied aspects/lenses of products, processes and focus areas as well as to benchmark and position the organization among its competitors, using digital traces of stakeholder emotions, opinions and experiences in social media. A novel application of the widely accepted honeycomb social media framework was presented to provide structure to unstructured social media text and identify the said aspects of organizations (products, processes, focus areas and competitor organizations) embedded in social media.

The novel framework was presented in six generic steps and the effectiveness of the proposed techniques empirically tested and verified using a case study application on the higher education domain in Australia together with a public online discussion forum data set of over one million textual discussions. All in all, the chapter demonstrated a novel, generic methodological approach to develop techniques to generate actionable organizational insights aligned with organizational strategy, which enable organizations to improve the perceived stakeholder experience through the acquisition of knowledge on the organization's performance related to the aspects of products, processes, focus areas and competitors using stakeholders' word-of-mouth.

Chapter Appendix

List of Australian Universities	Latrobe University Melbourne University Monash University Deakin University Australia National University RMIT University Swinburne University Australian Catholic University University of New South Wales University of Queensland Bond University Central Queensland University Charles Sturt University Federation University Flinders University University of South Australia University of Technology Sydney University of Western Australia Victoria University University of Sydney Queensland University of Technology Griffith University University of Murdoch Edith Cowan University University of Newcastle Western Sydney University University of Notre Dame University of Canberra Macquarie University University of Tasmania
List of courses in Australian Universities	Accounting and Business Agricultural and Environmental Sciences Allied health and rehabilitation Arts, communication and critical enquiry Asian studies Biological Sciences Dentistry and oral health Economics and Finance Engineering Information Technology Marketing and Management Medical and Psychological Sciences Social sciences Visual and creative arts Physical, Chemical and Mathematical Sciences
List of university processes	Enrolment Withdrawals & Deferrals Transfers Graduation Accreditation Assessments/exams Administration
List of Focus Areas for Universities in Australia	Employability Fees Offers Applications Prestige/Reputation Rankings Undergrad Postgrad Honours Mature Age Students International Students

	Living (on campus) Transportation Accommodation Library
List of departments, faculties and colleges of universities (represented with codes)	SHE (Science college) ASSC (Arts college) Business school, School of Education, ..etc School of Computing *(different for every university)*
List of user groups (High level sub forums to lower level topics)	*Education* group in whirlpool *Uni Stuff* section in ATARnotes *Transfers* forum in bored of studies *GAP years* forum in bored of studies *Exchange & Overseas study* forum in bored of studies

Table 4.11: Lists of components discovered from the social media dataset related to higher education domain of Australia

A sample data storage table:

PostDate	Title	Content	processes	courses	universities	focusAreas	SentiStrength	SentiWordNet	AuthorID (hidden)	MessageIndex
18/10/2009	what can i do with a bachelor of biomed?	There is also Deakin uni, which has just opened a brand new medical school, although I'm pretty sure their campus is in Geelong. If you really want to do Med, there are 12 graduate medical schools in Australia to choose from.	[]	[]	['Deakin University']	[]	2	0.7459477		20
27/11/2012	Law Degree	If you've done mainstream English and scored over 35, your English is better than around 80% of the state and that's a very competent score, so I think you should be fine!	[]	['Arts, communication and critical enquiry']	[]	[]	3	1.7962839		7
15/12/2016	Commerce/Computer Science vs Commerce/Biomedicine	Commerce/Computer Science would be a really good combo for algorithmic trading and in my opinion go together better then commerce/biomedicine. Biomedicine really lends itself to be a premed degree or research.	[]	['Accounting and Business', 'Information Technology']	[]	[]	2	1.9387254		2
27/11/2012	Law Degree	Why are you wasting 3 years doing a degree you don't plan to use? Especially a degree like law which can be extremely dry, boring and expensive. And in a market which is already oversaturated with law students... Yes, to get your foot in the legal industry your marks do matter, as well as work experience, co-curriculars (e.g. mooting) and extra-curriculars (e.g. hobbies - law firms like to see that you have other things going on in your life). After your first job, your marks will start to matter less. But you have to get in first!	[]	['Law']	[]	[]	0	0.867345		11
04/06/2012	Commerce Double Major & Electives?	For what it's worth if you are doing an accounting major, just make sure you meet all the subject requirements of the respective industry body/association which you intend to pursue subsequent to finishing uni (e.g. CA vs. CPA requirements).	[]	['Accounting and Business']	[]	[]	2	0.7941089		8
23/01/2015	Help with timetable! Combined commerce!	Yes - once again, you may do courses in whatever order you want, except of course if there are prereqs etc. etc. I suggest 2 COMM + 2 SCI courses in most semesters.	[]	[]	[]	[]	0	0.2166272		2
26/09/2013	How to manage the workload?	Hey guys, I'm interested in studying law and science next year however I am worried that the workload will be too great! Has anyone studied these two courses before and if so could you please let me know how you went? Thanks	[]	['Law']	[]	[]	-2	0.5474793		0
01/07/2010	Server/Database career	Hi, so i was looking through tafe courses and i didn't see a category that fit this type of role. Possibly networking it may have some server units in it, but no official course for server/database administration stuff? i'm wondering where is a good way to learn this stuff?	[]	[]	[]	[]	1	1.2149082		0

04/02/2014	Telco Engineering- UNSW vs USYD vs UQ	Hi guys, i really need you advice regarding the best place to study master in telecommunication engineering. currently i received offer form UNSW and USYD both of them for master of telecommunication engineering and University of Queensland for Master of Science Engineering (Management). i'm indonesian and currently working as network engineer for a mobile operator. i am interested in the unsw program because there will be 60 days of industrial training where i should gain valuable experience. but regarding rating, USYD and UQ are better than UNSW for engineering subject.	[]	['Engineering', 'Asian studies']	['University of New South Wales', 'University of Queensland']	['Offers']	1	3.2763996		0
23/08/2011	Music / Drum Teacher	Just wondering if anyone has done further studies or if they exist for teaching drums ie. AMEB. I am looking at getting recognized accreditation so i can teach drums, if anyone can help point me in the right direction.	[]	['Arts, communication and critical enquiry']	[]	[]	0	0.5498382		0
10/12/2011	print materials and online courses?	Hey guys, i've done online courses before and frankly I enjoy them, always do really well in them lol. I'm trying to pick an elective for next semester and i've seen many courses that say print materials....what's the difference between that and an online course? This is at griffith.	[]	[]	[]	[]	2	1.2058876		0
30/12/2015	After Uni Degree HELP!!	Can you become an architect with a Graduate Diploma? In Australia, at least, you need to do a 3 year Bachelor's and often an additional 2 years Masters. Have you thought of doing a Diploma in Building Design? I believe it's similar to architecture (not as creative, but still somewhat design based), and has good employment opportunities. It also has links with construction, if you're interested in that sort of thing.	[]	[]	[]	[]	1	0.5357617		22
03/08/2010	New QUT students, Questions and Answers	What is the best place to park if I want to go to uni during nights say around from 8 to after midnight?	[]	[]	[]	[]	1	0.8158613		716
25/05/2006	Anyone used Excom for Linux Professional	I'm interested in the Suse Linux Professional Practium on offer from Novell (via Excom) to become a certified Suse Linux Professional. Just looking for opinions from those who may have done the practium. Thanks	[]	[]	[]	['Offers']	1	0.7129353		0
08/08/2011	Law/Commerce or Law/Economics?	I'm going to be doing a dual degree in either law/commerce or law/economics at uni next year, but I can't decide which. The field of investment/merchant banking really appeals to me, and I suspect that is the path which I will eventually follow (but who knows!). Considering that, which degree would be the best option? In terms of investment banking, what placements/internships/part time jobs are available for a student whilst studying? In a very best case scenario, what is the earliest year in which one could find such a job? (I hear it is possible to find a law clerkship in second year if you are really good enough, but I haven't heard anything regarding banking). Thanks!	[]	['Law', 'Economics and Finance', 'Accounting and Business']	[]	[]	0	3.3942447		0

Table 4.12: A sample data storage table (from *Excel*) demonstrating a set of the resulting, transformed and structured *final data elements* presented in the data transformation technique illustrated in figure 4.9. In this table each row presents a final data element that has been structured and prepared to perform analytics to generate insights.

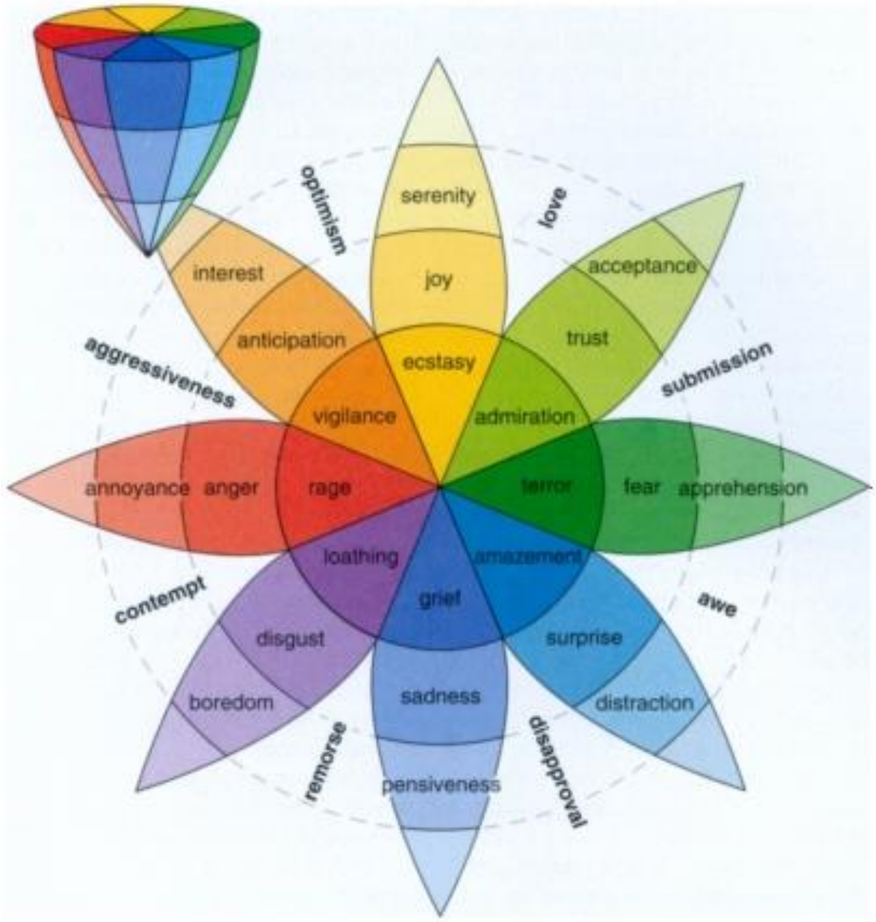


Figure 4.20: The Plutchik's wheel of emotions. Image courtesy of (Plutchik, 2001)

5. Organizational brand from the opinions of the Digital Stakeholder

"Your Brand is what other people say about you when you're not in the room."

~ Jeff Bezos

This chapter extends the knowledge gained from the framework introduced in chapter 4 related to understanding the *Perceived Organization* into organizational strategy by exploring the possibility of understanding organizational brand from social media. Chapter 4 presented novel techniques to understand and map components embedded in social media into organizational aspects that can be measured and evaluated. This chapter explores the possibility of extending this knowledge by mapping these aspects of the organization uncovered in chapter 4 into accepted features of organizational brand. This chapter proposes a novel method of extracting an organization's brand from social media by introducing a technique to map components of social media into accepted features of brand, presented in marketing literature. Finally, this chapter presents a generalizable technique of brand management for contemporary organizations to transform decision-making utilizing machine learning and social media in combination with a set of widely accepted scales from Marketing literature; Brand Personality Scales and an accepted social media framework; the honeycomb social media framework.

The proposed technique presents a method of automated monitoring of stakeholders' perception of organizational brand image, brand consistency and harmony of brand architecture in near real-time, at different granularities across a spectrum of organizational constituents facilitating benchmarking against competitors. The proposed approach utilizes high volume, high velocity, freely available, freely expressed and heavily underutilized social media data sources and incorporate deep learning, word embedding and emotion extraction techniques and present three alternate methods that can be adopted by organizations according to the level of human and technology intervention possible based on the resources, technological capacity, maturity and culture of the organization. The proposed technique has been successfully tested and demonstrated using a dataset with 1.2 million text posts on the higher education sector in Australia.

Figure 5.1 offers a diagrammatic, summarized, presentation of the contributions of this chapter. The diagram illustrates how this chapter utilizes output generated from the novel framework proposed in chapter 4 (Social media to organizational insights transfiguration framework); specifically the mapping of components embedded in social media to measurable aspects of the organization; together with theory on branding from marketing literature to draw brand personality scales; to ultimately map organizational components embedded in social media to components of organizational *Brand Personality* and generate more structured organizational insights.

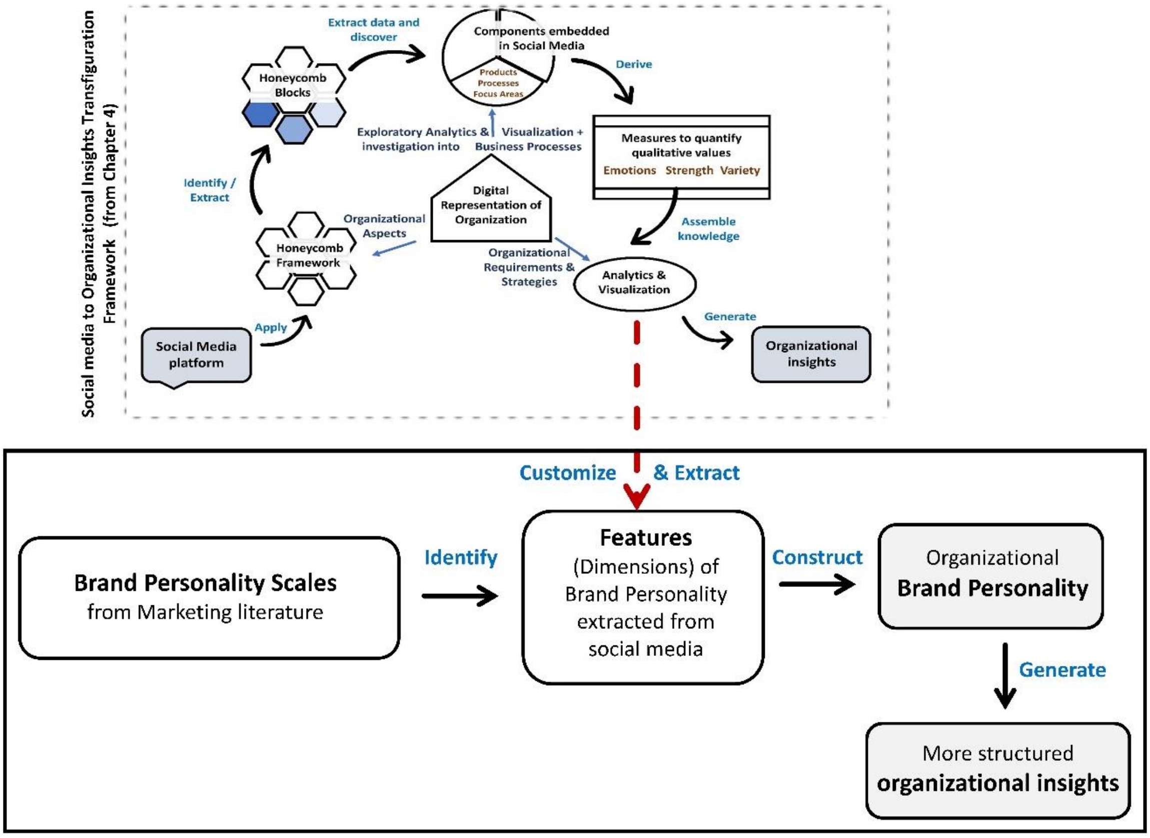


Figure 5.1: A diagrammatic representation of the summarized contribution of chapter 5 (highlighted in grey) and the input utilized from the Social Media to Organizational Insights Transfiguration Framework from chapter 4. This chapter utilized theory on *Brand Personality* from Marketing literature to identify features or dimensions that make up an organization's brand (using Brand Personality Scales) and uses the mapping of organizational components embedded in social media to measurable aspects of the organization (from the framework proposed in chapter 4) and ultimately generate the mapping of social media into organizational brand. This enables the construction of organizational Brand Personalities (observable in real time) and generate highly structured organizational insights (that can be generated on top of the constructed brand personality) using social media text data available in public online discussion forums.

5.1) Introduction to Brand

The "*Art of enchanting the soul*" has been explained by Plato as *rhetoric*. The *rhetorical* analysis introduced by Aristotle for observing the means of persuasion, presents three modes or appeals of persuasion, namely; *ethos* - ethical appeal, *pathos* - emotional appeal and *logos* - logical appeal; which has become part of many theories, frameworks and concepts in contemporary research literature (McCloskey, 1998, 2000; Iglesias and Bonet, 2012; Sigrell, 2008 as cited in Urde, 2016) . According to (Urde, 2016) these *rhetoric* perspectives are core aspects of a brand management strategy.

As elaborated in the literature review (section 2.2.1), today corporate brands have ascended to the status of providing distinctive institutional-identities in the business world, and holds a prominent position in corporate marketing and strategic management due to its ability to create business and shareholder value. Therefore, organizations that avail themselves of current research and best practices concerning brand management will be in a better position to succeed (Duesterhaus & Duesterhaus, 2014).

In such an environment, having access to tools and techniques, which enables an organization to understand and monitor their brand image offers immense business value.

5.2) Measuring Brand: Brand Personality and Brand Personality Scales

When investigating into methods of measuring and evaluating brand image ; the concept of *Brand Personality* (introduced in section 2.2.2 in the literature review) appears in business and marketing literature as a proven method to differentiate products and companies (Clifton & Simmons, 2003). The term *Brand Personality*, was first introduced by Martineau in 1958, and refers to a set of human characteristics or stakeholder perceptions associated with a brand (Xu et al., 2016 ; Martineau, 1958). Brand personality scales have been demonstrated to be a reliable, effective and generalizable tool for evaluating brand personality (A. Xu et al., 2016).

An in-depth inspection of the essence of these brand personality scales revealed that they are in fact designed to extract the brand's performance in the three key appeals of persuasion introduced in rhetoric (and discussed above in section 5.1); ethos, pathos, and logos; i.e. the ethical, emotional and logical appeal of a brand in the eyes of the public or consumer.

However, extraction of this brand personality has been challenging for decades. Traditional techniques, most widely used to date as elaborated in section 2.2.3 in the literature review are questionnaires and interviews conducted online and offline (Chung & Park, 2017; Gordon et al., 2016; Hultman et al., 2015; Matzler et al., 2016; So et al., 2017; Sung et al., 2015; Zenker et al., 2017). Some of the more contemporary additions into this repertoire of techniques are online product reviews and social media (Callarisa et al., 2012; Gensler et al., 2015; Hudson et al., 2016; Saboo et al., 2016; Unnisa et al., 2016).

While the traditional methods have been effective, they require time, resources, money and experts in domain knowledge to design, execute and analyse. Furthermore, the quantity of data (sample size) that can be collected and processed is limited due to the time and resource consuming nature of data collection and analysis, which confines the quality of results. Furthermore, there is minimal to nil opportunity to extract this information for competitor brands.

Thus, here onwards this chapter will present our proposed direction to address these limitations in current brand management strategies by utilizing the untapped potential and opportunities in social media elaborated in the literature review.

In the endeavour for improved techniques for capture and understanding of an organization's brand, a glance into the current digital environment in which all business organizations operate (introduced in section 2.3.1 of the literature review), reveal that information is generated at high volume, velocity and variety in relation to every aspect of human life. As elaborated in the literature review, this is known as the big data environment (Ammu & Irfanuddin, 2013 ; Pournarakis et al., 2017). Through the rise of social media platforms, this environment has witnessed a fundamental shift in how consumers convey their experience during interactions with a brand, with consumers all over the world presently influencing and shaping perceptions of brands, leaving no choice for organizations than to adopt, invest in and monitor these channels (Pournarakis et al., 2017).

Measure of brand personality through social media Analytics (SMA) can provide a highly dynamic and near real-time assessment of what stakeholders are experiencing and expressing at the moment of brand interaction (Pournarakis et al., 2017).

However, despite the presence of this data and technology, organizations' ability to harness the opportunities presented by social media is still in it's infancy as emphasized in the introduction chapter and the section 2.3 in the literature review (Alahakoon & Wijenayake, 2018; Dhar & Mazumdar, 2014; Wijenayake et al., 2020). In this chapter, we introduce a novel methodology and technique that will enable organizations to leverage this potential of social media data using a combination of machine learning techniques presented in the literature review (section 2.4).

5.2.1) Introducing the case study organization to Brand

With the modern pace of business, higher education (HE) has become an industry that is newly being exposed to business competition (Khanna et al., 2014). Today, many higher education Institutes (HEIs) are struggling to maintain their brand image, keep up with official evaluations and rankings, constantly implementing improvements to attract high achieving students and staff and to retain their already enrolled students. Hence, we selected higher education as the ideal case study to demonstrate our approach.

5.2.2) University Brand Personality Scale

Although there is agreement that understanding institutional branding and clearly developing and communicating that brand is of great value to HEIs or universities, research on *ethos pathos* and *logos* of higher education Institutes, (e.g. a university's ; brand image, identity, reputation, and meaning) remains underdeveloped (Rauschnabel et al., 2016).

Literature reveal that *ethos pathos logos* have been used to identify brand personality of higher education Institutes as well as to understand different aspects of HEIs for marketing, such as how higher education Institutes can engage students and the community using social media (Brech et al., 2017) and how international partnerships and perceived service quality relate to student satisfaction (Wilkins et al., 2017).

However, these studies lack the utilization of a robust brand scale to represent and benchmark higher education organizations due to the deficiency of a suitable scale in *Brand Personality* literature. In order to fill this gap, (Rauschnabel et al., 2016) introduced the University Brand Personality Scale (UBPS) for brand management in higher education. UBPS can effectively capture the *Brand Personality* for universities across countries.

According to (Rauschnabel et al., 2016), university brand personality represents the human-like mental associations stakeholders have with and about a particular university. Hence, using a series of seven qualitative and quantitative studies the UBPS was developed and consist of the six dimensions;

“1) prestige (accepted, leading, reputable, successful, considerable)

2) sincerity (humane, helpful, friendly, trustworthy, fair)

3) appeal (attractive, productive, special)

4) lively (athletic, dynamic, lively, creative)

5) conscientiousness (organized, competent, structured, effective)

6) cosmopolitan (networked, international, cosmopolitan)” (Rauschnabel et al., 2016)

In this case, *ethos* is represented by the dimension sincerity and *pathos* by prestige and appeal, whereas *logos* is implied by dimensions conscientiousness and cosmopolitan. Thus, we selected this brand personality scale as the most suitable scale to demonstrate our technique on the case study domain; higher education sector.

5.3) Related Work (Social Media, Brand Management, Higher Education and Machine Learning)

Several recent attempts have been made to make use of social media content to extract opinions. For example, recent publications such as (Unnisa et al., 2016) have used Twitter data for sentiment extraction by clustering tweets into positive and negative clusters. The research by (T. R. Bandaragoda et al., 2017) used unsupervised incremental machine learning to monitor topic pathways in twitter discussions and (T. Bandaragoda et al., 2018) utilized an ensemble of deep learning and machine learning algorithms to extract the quality of life of cancer patients from self-reported information in online cancer support groups.

Many studies have used online product reviews, a less dynamic form of e- Word–of-Mouth (eWOM) which are direct feedback on products rather than unstructured discussions in social media to understand brand image of organizations. A few recent studies have attempted to use social media for university brand related activities. In the higher education sector, (Rutter et al., 2016) have attempted to discover if institutions with lower reputational capital can compete for students by increasing their brand presence. The study analysed the impact of social media interaction on student recruitment (the response to organization intervention/presence in social media) using content generated by 60 social media pages (Facebook and Twitter) managed by higher education institutes.

Another study by (Bolat & O'Sullivan, 2017) made an attempt to understand how higher education institutes can utilise student-generated social media data for higher education marketing and branding using behavioural expressions (such as likes, shares and opinion comments) and individual experiences (stories and comments of personal experiences and views). A mixed method analysis was used with qualitative data analysis focused on the content of posts, impressions in the form of text-based comments and quantitative analysis based on frequencies and quantitative interaction metrics revealing that students engage well with student-generated content. This demonstrated how in practice, higher education institutes can learn from student-generated social media content by performing a three-step analysis of social media data: descriptive analysis, sentiment analysis and network analysis.

5.4) Public Online Discussion forums

As discussed in sections 2.3.5 (in the literature review) and 4.2 (in the previous chapter), public online discussion forums are a form of social media that is used heavily in modern society for domain specific information exchange. Individuals seek this type of social media to gain more information and clarity related to different organizations, products and services to support their decision making. Hence, these forums are rich in stakeholder opinions, expectations and experiences related to many industries. Furthermore, as discussed earlier, this data is freely available, freely expressed and updated often. Moreover, as presented in the literature review, our investigation into existing research revealed that this type of social media has not been fully utilized by organizations to enhance their processes and organizational image. Hence, public online discussion forums were selected as the data source to illustrate our technique.

The higher education sections of these forums possess all of the above stated information directly related to universities in the relevant country. In relation to the higher education sector in Australia, there are several public online forums that display a high level of engagement of prospective students, enrolled students and parents seeking advice on courses and universities as well as university staff members and experts in higher education providing guidance and information.

The dataset used for experiments was the same social media dataset used for experiments in chapter 4 (elaborated in section 4.6). To prepare this dataset, higher education sections of the three most widely used public online forums were selected as sources rich in information and opportunities related to higher education in Australia. Namely;

- Whirlpool (http://forums.whirlpool.net.au/)
- ATAR Notes (https://atarnotes.com/forum/)
- Bored of Studies (http://www.boredofstudies.org/)

5.5) Deep Learning and Long Short-Term Memory (LSTM)

This chapter focusses on developing techniques to construct organizational brand personality through stakeholders' *Perceived Organization* (explored and constructed in chapter 4) modelled and extracted through text conversations over social media (figure 5.1). One of the key challenges of constructing this organizational brand is the extraction and classification of the dimensions of *Brand Personality* from textual social media. Machine learning techniques summarized in section 2.4 in the literature review were explored to design and trial an effective solution to overcome this challenge through sequential data modelling.

Deep learning (introduced in section 2.4.4 of the literature review) is a Neural Network (NN) architecture that allows computational models consisting of multiple layers of processing to learn data representations (such as sequences) with multiple levels of abstractions. A composition of multiple such transformations allows even highly complex patterns to be learned. Hence, for classification tasks, higher layers of representation amplify aspects of the input that are important for categorization (LeCun et al., 2015). This technique has dramatically improved the state-of-the-art in many domains that require sequence and pattern recognition such as speech recognition and hand writing recognition (LeCun et al., 2015; Wijenayake et al., 2021).

As elaborated in the literature review, there are several deep learning architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) (Wijenayake et al., 2021). Long Short-Term Memory (LSTM) is a RNN architecture introduced by Hochreiter (Hochreiter & Schmidhuber, 1997) that has yielded successful results in sequential data modelling with recent studies (Breuel, 2017; Graves et al., 2013; W. Liu et al., 2017; Sak & Senior, 2017; Shi et al., 2017). Hence, this architecture was selected as the most effective method to learn and classify brand personality dimensions from text data for the research presented in this chapter. For experiments, a Bidirectional Long Short-Term Memory (BiLSTM) which uses a forward pass NN and a backward pass NN to learn sequences in both directions was used (Wijenayake et al., 2021).

5.6) Methodology/ Proposed approach to generate Brand Personality from social media (Experiments with the University Brand Personality Scale)

As presented in figure 5.2 below, we propose three novel methods (which we call 'text processing pathways') to model and extract *Brand Personality* (the University Brand personality in this case study demonstration) using social media text data with varied levels of human-machine intervention. The intention is to give organizations (higher education organizations in this case) a choice to select the data processing method most appropriate for their institute based on the human and technology resources deployable.

The three novel methods (text processing pathways) in summary are as follows;

1) Skilled Mapping Method: which utilizes an expert's (an individual familiar with the domain) intervention to map the discovered processes and focus areas from *social media to organizational insights transfiguration framework* (presented in chapter 4) into brand dimensions by selecting the keywords associated with each *Brand Personality* dimension using human skills and domain knowledge.
2) Word2vec Method: which makes use of word embedding techniques introduced in section 2.4.6 of the literature review to partially automate the process of generating keywords using a trained word embedding (word2vec) model (steps presented in figure 5.2).
3) Deep Learning Method: which presents a completely automated approach using the word2vec model from the text processing pathway No. 2 together with a trained LSTM neural network (figure 5.2).

The proposed technique will be presented using a step-by-step approach from here onwards.

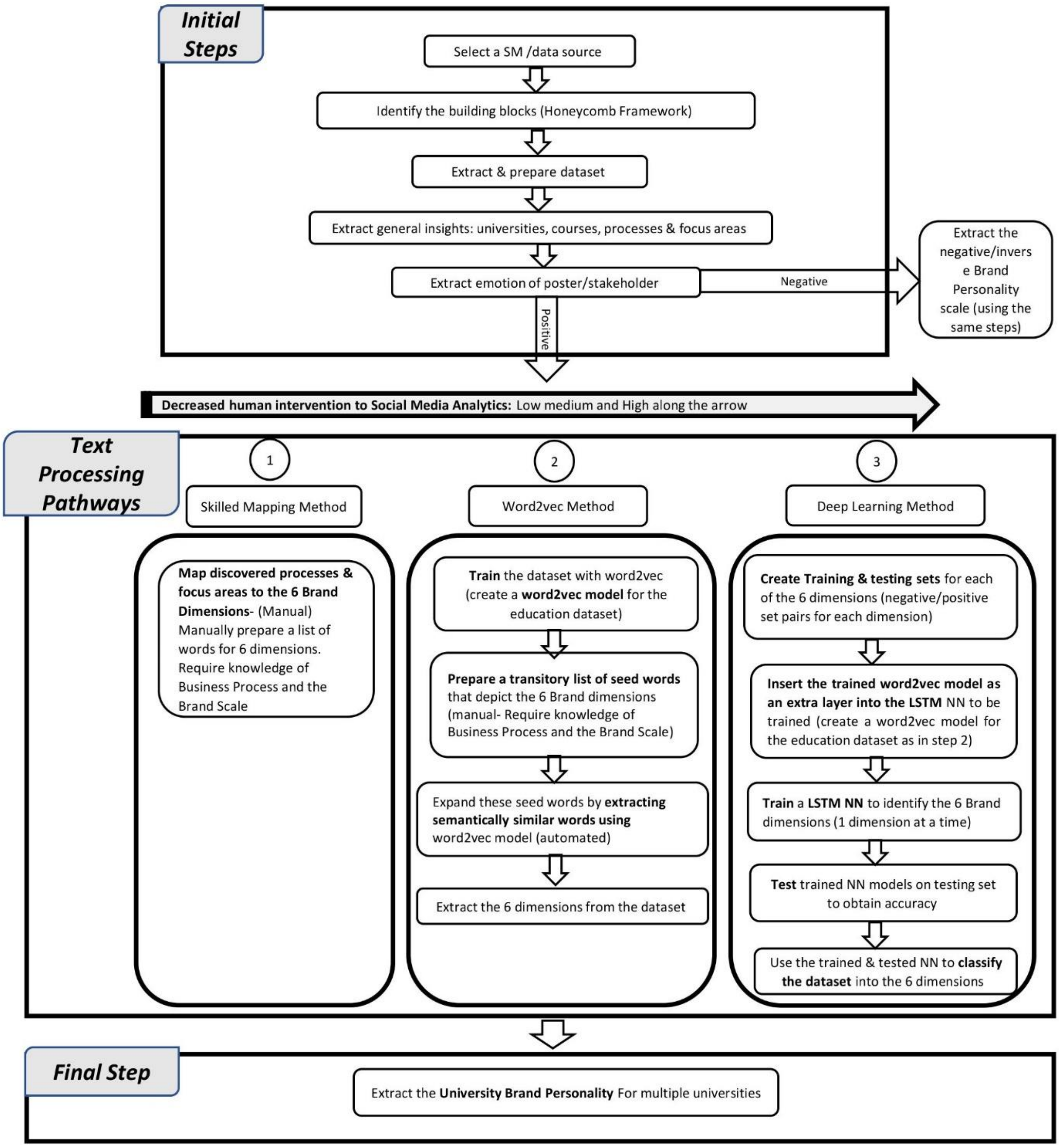


Figure 5.2: The flow diagram of the technique for extracting University Brand Personality Scale (UBPS) using textual social media word-of-mouth, illustrating the three proposed alternative methods (or data/ text processing pathways)

Steps of the Proposed approach:

5.6.1) Initial steps: Data extraction and pre-processing

The initial steps of data extraction and pre-processing are covered in chapter 4; extraction of the *Perceived Organization* from social media (Social media to organizational insights Transfiguration Framework: figure 4.4). However, as the chapters in this thesis are independent, a summary of the initial steps of selection of data source, identification of the building blocks of the data source, extraction of general insights, organizational insights and emotions are presented herein.

1) Selection of data source

For demonstration of the proposed methodology, as the first step, public online discussion forums were selected as the data source as mentioned in the section 2.3 in the literature review.

2) Identification of the building blocks of the datasource – Honeycomb framework

The next step was to identify the main building blocks or areas of the social media to engage with for maximum impact using the honeycomb social media framework introduced by Kietzmann (Kietzmann et al., 2011). As elaborated in section 2.3.10 of the literature review, this framework was selected as it is the most widely used and cited social media framework in literature. As articulated in the literature review, the framework introduces a honeycomb of seven functional building blocks: *identity, conversations, sharing, presence, relationships, reputation*, and *groups*. The honeycomb blocks allow to unpack and examine different facets of the social media user experience and their implications for organizations. The building blocks help to understand how different levels of social media functionality can be configured (Kietzmann et al., 2011). According to this framework each social media platform is driven by primary, secondary and tertiary building blocks, which provide the foundation for important social media design decisions (Kietzmann et al., 2012).

As presented in sections 4.4 - 4.6 in the previous chapter, the honeycomb social media framework was applied to identify the areas or building blocks of the social media platform to utilize for extraction of the dataset. The application disclosed that in terms of public online forums, “Conversation” and “groups” are the building blocks or areas where higher education institutes should engage with in this type of social media (refer to the honeycomb framework application presented in figure 4.3 in the preceding chapter). Public online discussion forums are a type of social media environment designed to mainly enhance conversations. Hence, our dataset was extracted from conversations in higher education groups from the public forums. The composition of the final dataset consisted of 44% from the Bored of Studies, 41% from Whirlpool and 15% from ATAR notes.

3) Extraction of general insights

Subsequently, an initial descriptive analysis was carried out to extract basic insights such as the organizations mentioned in the conversations (higher education institutes), products (courses), organizational processes and focus areas to build the foundation needed to proceed into the extraction of the brand personality and the extraction of emotions or sentiments for each post (contribution from chapter 4).

The following steps were carried out to extract general insights from posts using the techniques presented in chapter 4 (figure 4.4); *social media to organizational insights transfiguration framework*.

4) Extraction of organizational insights

Universities (organizations) mentioned in discussions, university courses (products), university processes and higher education focus areas were extracted from the dataset. This was done by identifying lists of domain specific keywords from posts (Examples provided in section 5.7 below). University processes were identified by a separate study into the university business processes in the student management system (presented in figure 4.6). The resulting model from this pre-study (figure 4.6) was used to identify processes through text mining from the above-mentioned social media text data set prepared for experiments.

5) Extraction of emotions

Sentiments of stakeholders' related to the extracted posts were mined using the Sentistrength algorithm (presented in section 2.4.5 of the literature review) introduced by Thelwall (Thelwall et al., 2010). Sentistrength's generated value related to every single post was calculated and stored in the dataset (using the method presented in figure 4.5 in chapter 4). Threshold values were used within the range -5 to +5, to categorize posts as *positive*, *very positive*, *negative*, *very negative* or *neutral*. The developed tool allows the range to be selected by the user depending on the intensity of positivity or negativity required based on the scenario and insight needed. Generally, a Sentistrength value above 1 can be considered as a positive sentiment and a value below -1 can be considered as a negative sentiment. However, if for example the management need to view extremely positive feedback, a Sentistrength value above 3 can be selected. To extract the UBPS, posts that were categorized as positive were used.

5.6.2) Method 1: Skilled human mapping Method

The UBPS was applied to the extracted university insights. As illustrated in the flow diagram in figure 5.2, the *processes* and *focus areas* sections of the extracted university insights were mapped into the six dimensions of the UBPS (in other words, the six dimensions of UBPS were identified from the dataset using the discovered *processes* and *focus areas*). This is a manual step requiring skilled human intervention. This step utilizes human intellect and domain knowledge: understanding of university business processes and the UBPS. Next, the posts discussing the six brand dimensions were identified and categorized accordingly (some overlying is possible between dimensions).

5.6.3) Method 2: Word2vec based technique

In this method, as exemplified in the flow diagram in figure 5.2, the dataset was used to first train a word2vec model and create a word embedding model unique to this dataset. Word2vec is a Neural Network (NN) based algorithm introduced by (Mikolov et al., 2013). As elaborated in section 2.4.6 of the literature review this model has been successfully used for learning Word Embeddings or vector representations of words (Le & Mikolov, 2014). This was automated and carried out using the Gensim Python library (Rehurek & Sojka, 2010). This created model possesses word vectors for every single word in the dataset.

Next, as presented in figure 5.2 (text processing pathway No. 2), a transitory list of basic words that depict each of the six dimensions in the UBPS was created. Again, this is a manual step that needs to be carried out by a person knowledgeable in the UBPS, the social media dataset and higher education business processes. However, the human intervention needed is not as intense as method 1 (section 5.6.2), as the list needed to be prepared manually is a basic list.

As the next step, each of the basic words were expanded by extracting their corresponding word2vec vectors in the trained model. The table 5.2 in the results section demonstrates a sample of the word vectors obtained from the trained word2vec model for the dataset, for four words (‘prestige’, ‘reputation’, ‘transfer’ and ‘accreditation’) in a basic list.

Then, all words given by the word vectors above a threshold of 70% (>0.7) were selected into the final word list for each dimension. Alternatively, if more human intervention is possible/ available, it is recommended to manually go through the vector and select the most meaningful and best fit words to identify the dimensions (a semi-automated approach). The threshold too can be changed depending on the requirement. Thereafter, using the generated word lists for the dimensions, posts in the dataset were categorized into the six dimensions of the UBPS (again some over lapping is possible).

5.6.4) Method 3: Deep Learning based technique (Deep LSTM Method)

This method uses a novel deep Recurrent Neural Network (RNN) of Bidirectional Long Short-Term Memory (BiLSTM) the authors call the “Deep LSTM Technique” (presented in detail in figure 5.3). As elaborated in section 2.4.4 (of the literature review) Long Short-term Memory (LSTM) is used for sequence learning. Words are a sequence of characters and a sentence is a sequence of words. A BiLSTM uses a forward pass NN and a backward pass NN to learn sequences in both directions. Deep learning, as presented in section 5.5 above, allow computational models consisting of multiple layers of processing to learn data representations such as sequences with multiple levels of abstractions, and have yielded successful results in many domains that require sequence and pattern recognition such as speech recognition, hand writing recognition and genomics (LeCun et al., 2015). The combination of these methods with the LSTM-RNN architecture have yielded successful results in past studies related to sequential data modelling (Breuel, 2017; Graves et al., 2013; W. Liu et al., 2017; Sak & Senior, 2017; Schmidhuber, 2015; Shi et al., 2017).

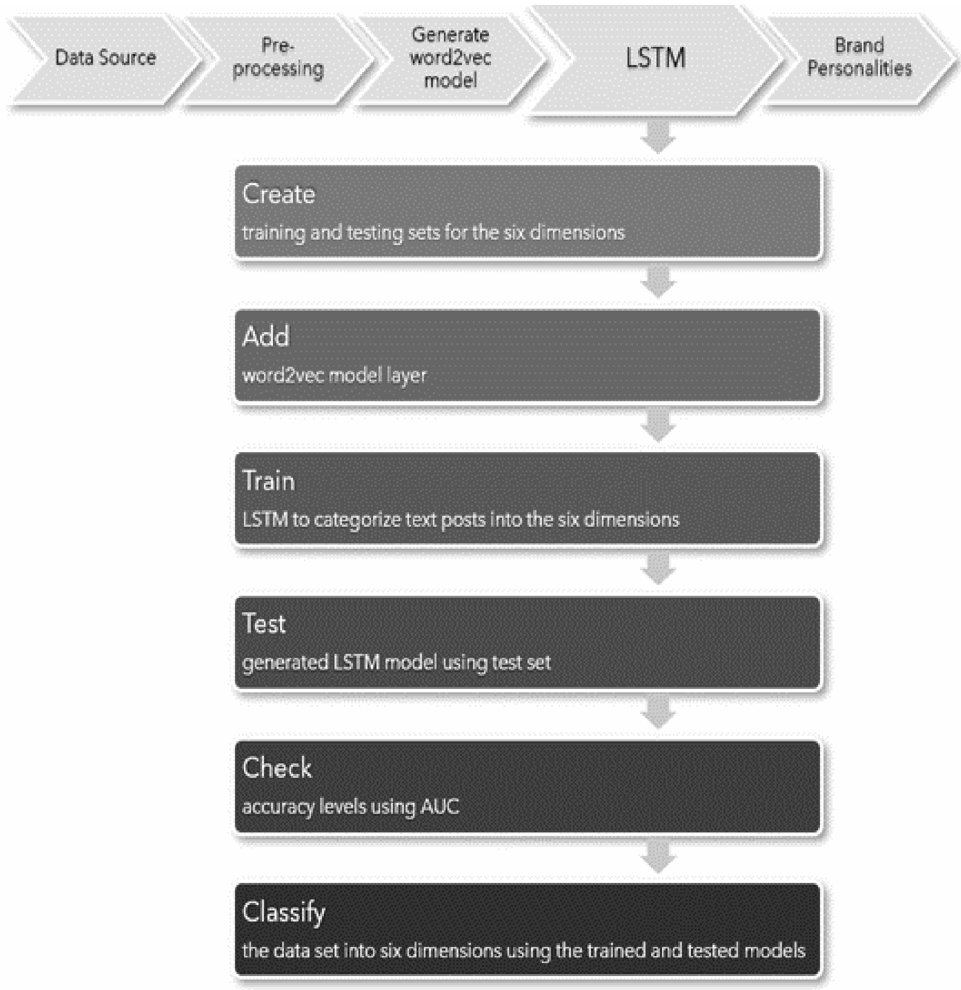


Figure 5.3: Deep LSTM Technique (A detailed illustration of the text processing pathway No. 3 of the flow diagram presented in figure 5.2) for the entirely automated extraction of brand personality from social media using social media text data (from public online discussion forums), a trained word embedding model (word2vec) and a LSTM neural network (Wijenayake et al., 2020, 2021).

In our proposed method No. 3, as demonstrated in figure 5.2 and figure 5.3, first; a training set and a testing set were created for each of the six brand dimensions. A negative and positive set was created for each dimension. These data sets can be created manually or by using any of the two methods introduced above (sections 5.6.2 and 5.6.3) based on the availability of human resources. The demonstrated results are based on the use of the method 1 presented in section 5.6.2. A random set of 20,000 posts were used for each of the training sets and a random sample of 50 posts were used for the testing sets.

Next, the word2vec word embedding model trained using the dataset (in method 2 in section 5.6.3) was fed into the NN as an intermediary layer to provide extra knowledge (depth) for training (illustrated in figure 5.2).

Then, the BiLSTM was used to train and identify the six brand dimensions. Each dimension was identified independently using the created training datasets resulting in separate trained models for each dimension.

As the next step, the trained NN models were tested using the testing sets to obtain accuracy levels. The validation sets were randomly created (through automation) for 5 epochs or iterations and the testing phase was repeated for 10 runs to maximize accuracy.

Finally, after checking the acceptability of accuracy levels, the trained and tested NN models were used to classify the entire dataset into six dimensions.

5.6.5) Final step: Extracting the UBPS

Once the posts were categorized into the six dimensions, the number of posts that mentioned each of the dimensions in a positive light was obtained. This count was used to provide the strength for each brand dimension and repeated for the different granularities under consideration such as university and course.

There by, the UBPS for the case study organization; La Trobe university as well as its' competitors (other universities in Victoria and Australia) were extracted and visualized.

5.6.6) A comparison of the three methods

Table 5.1 presents a comparison of these three methods proposed and illustrated in figure 5.2 for extracting Brand Personality.

	Method 1	Method 2	Method 3
Name/Description	Skilled human mapping	Word2vec based technique	Deep Learning based technique
Level of automation	Manual	Partially automated	automated
Skilled human intervention	Required	Required for some steps	Not required after the initial training
Domain knowledge	Required	Required for some steps	Not required
Deep Learning/Neural Network algorithms used	none	word2vec	Word2vec and BiLSTM
Technical skills required to setup	Low	Medium	High

Table 5.1: A comparison of the three proposed methods (text processing pathways) for extraction of Brand Personality (presented in figure 5.2).

5.7) Experiment Results

5.7.1) Results for Method 1

Word lists generated using skilled human mapping:

The following example illustrates the steps of output generation from method 1 for two brand dimensions 'Prestige' and 'Sincerity'. First, basic words were identified using the UBPS description and domain knowledge. Then, the suitability of each of these words were decided by mapping each word to the practical aspects of the dataset.

1) ***prestige*** *(~~accepted~~, ~~leading~~, reputable, ~~successful~~, ~~considerable~~)-*

Basic words *accepted*, *leading*, *successful* and *considerable* were removed due to their high usage in general figure of speech. They are used for positive as well as negative expressions in the dataset.

<u>Processes mapped to Prestige</u>: *Transfers, Withdrawal/defer, accreditation, university evaluations*

<u>Focus areas mapped to Prestige</u>: *Prestige/reputation, rankings, employability*

2) ***sincerity*** *(~~humane, helpful~~, friendly, ~~trustworthy~~, ~~fair~~) ~~safe~~*

<u>Processes mapped to Sincerity</u>: *teaching quality, enrolment, honours, application, ~~teaching (because teaching is a course)~~*

<u>Focus areas mapped to Sincerity</u>: *fees, offers, staff, administration, academic staff, ~~student centric~~, student centred, uni/support staff, requirements.*

Words *humane*, *helpful*, *trustworthy*, and *fair* were removed due to their generic use in the dataset or due to their lack of use in the dataset. (e.g.: Fair is used as a figure of speech to give different ideas in different situations, words *humane* and *student-centric* had insignificant usage in the dataset)

Similarly, the remaining four dimensions; lively, appeal, cosmopolitan and conscientiousness were extracted, and the outcome demonstrated in figure 4.5.

5.7.2) Results for Method 2:

Word lists generated using word2vec word embedding:

The following is a demonstration of the results for the brand dimension 'prestige' at each step.

First, a basic list of words was created.

Basic list of words: prestige, reputation, transfer, ranking, evaluation, accreditation, employability

Then, each of these basic words were expanded by extracting their corresponding vectors from the trained word2vec model. The table 5.2 illustrates the expanded results of the basic words '*prestige', 'reputation', 'transfer' and 'accreditation'.* Similarly, all of the words in the basic list were expanded to obtain a larger list of words.

The words highlighted in bold in table 5.2 such as *elitism*, *stature*, *prejudice*, *vibe*, *snobbery*, *go8* and *affiliations* are some of the completely new words discovered by word2vec and were not discovered by method 1 (in section 5.7.1). These words have been used in the dataset in similar situations to *prestige*. The similarity value shows the word embedding similarity of the word to the term, with 0.1 indicating a 100% similarity.

Since the word2vec model used was specifically trained on the dataset (not a general standard model) these words are highly relevant to the dimensions and increases the knowledge and accuracy of the output.

Term: prestige	Term: reputation	Term: transfer	Term: accreditation
('**elitism**', 0.7677)	('reputations', 0.8033)	('trasnfer', 0.8957)	('cpeng', 0.7807)
('reputation', 0.7244)	('rep', 0.7803)	('tranfer', 0.8559)	('recognition', 0.7791)
('go8', 0.6997)	('prestige', 0.7244)	('transferring', 0.8057)	('qualification', 0.7512)
('**g08**', 0.6896)	('repuation', 0.6961)	('transfering', 0.7916)	('accredited', 0.7335)
('popularity', 0.6840)	('atmosphere', 0.6922)	('trasfer', 0.7768)	('cacpa', 0.7182)
('perceived', 0.6832)	('**vibe**', 0.6778)	('xfer', 0.7551)	('cpaca', 0.6994)
('**snobbery**', 0.6704)	('perception', 0.6665)	('switch', 0.6983)	('accredition', 0.6949)
('glamour', 0.6588)	('cachet', 0.6624)	('reapply', 0.6907)	('approval', 0.6912)
('g8', 0.6552)	('roi', 0.6512)	('apply', 0.6736)	('cpa', 0.6847)
('sandstone', 0.6538)	('presence', 0.6493)	('transferred', 0.6669)	('ea', 0.6811)
('exclusivity', 0.6481)	('tradition', 0.6367)	('tranferring', 0.6601)	('designation', 0.6750)
('prestigious', 0.6443)	('facilities', 0.6364)	('enroll', 0.6553)	('qualifications', 0.6697)
('practicality', 0.6442)	('**ranking**', 0.6279)	('bcommerce', 0.6520)	('accrediation', 0.6680)
('roi', 0.6388)	('**affiliations**', 0.6247)	('transfered', 0.6513)	('rg146', 0.6673)
('**stature**', 0.6367)	('aura', 0.6229)	('ballb', 0.6475)	('registration', 0.6647)
('employability', 0.6317)	('repuatation', 0.6153)	('enter', 0.6447)	('rpeq', 0.6633)
('stigma', 0.6315)	('arguably', 0.6077)	('bcomba', 0.6445)	('acreditation', 0.6627)
('nongo8', 0.6301)	('perspective', 0.6036)	('enrol', 0.6384)	('cpaa', 0.6508)
('prestiege', 0.6259)	('stigma', 0.6023)	('bcomm', 0.6366)	('chartered', 0.6453)
('competitiveness', 0.6168)	('ties', 0.6007)	('transer', 0.6259)	('exemptions', 0.6444)
('**prejudice**', 0.6162)	('quality', 0.6005)	('unenroll', 0.6259)	('fiaa', 0.6417)
('attractiveness', 0.6111)	('opinion', 0.5953)	('reenrol', 0.6248)	('pr', 0.6404)
('greed', 0.6045)	('elitism', 0.5886)	('trasnferring', 0.6228)	('citizenship', 0.6224)
('prestigous', 0.6043)	('resourced', 0.5885)	('proceed', 0.6157)	('supervision', 0.6187)
('nepotism', 0.6041)	('rivalry', 0.5873)	('transfers', 0.6130)	('endorsement', 0.6180)

Table 5.2: Expanded lists using word2vec for the basic words; 'prestige', 'reputation', 'transfer' and 'accreditation' using the trained word2vec word embedding model (illustrates the words with the generated word embedding similarity for four decimal places from the method No.2 presented in figure 5.2. With 1.0 being equal to 100% similarity).

Finally, these word lists were trimmed by manually browsing the vectors and selecting a short list of most meaningful and best fit words to identify the dimension considering the words and their similarity values.

Alternatively, this step can be completely automated by using a threshold for the similarity (e.g: >=70% or above 0.7) and creating the short list accordingly.

Final word list for the extraction of the dimension 'prestige':

go8, group of eight, group of 8, prestige, reputation, prestigious, prestigious, elite, reputable, reputed, transfer, transferred, transfered, transferring, transfers, withdraw, withdrawal, withdrawals, withdrawing, withdrawn, withdrew, deferment, defer, deffer, defered, deffered, accreditation, accredited, accredit, qilt, ranking, rankings, ranked, rank, employment, employment opportunities, job opportunities, jobs, job prospects, employability

Similarly, final word lists were obtained for the other five dimensions, sincerity, appeal, lively, conscientiousness and cosmopolitan.

5.7.3) Results for Method 3:

Accuracy levels for the six brand dimensions using the deep learning method

Values for Area Under the Curve (AUC) were calculated for all dimensions, for 10 runs with 5 epochs (iterations) each. Table 5.3 illustrates the average AUC obtained for each of the six dimensions. AUC refers to the area under the Receiver Operating Characteristic (ROC) curve. AUC is used as a performance measure for machine learning algorithms and is highly accepted as an indicator of accuracy. A higher AUC number indicates a higher accuracy with the maximum number being 1.0 indicating an accuracy of 100 percent (Bradley, 1997).

In this proposed method, high accuracy levels were obtained for each of the six dimensions with the AUC approaching 1 for the positive sample and AUC approaching 0 for the negative sample. The high accuracy levels would be due to the use of a word embedding model providing an extra layer into the learning of the BiLSTM.

Brand dimension	average AUC
Prestige	0.984882829
appeal	0.936647972
sincerity	0.983898515
lively	0.979387417
conscientiousness	0.987680669
cosmopolitan	0.987671367

Table 5.3: Training accuracy - Area Under the Curve (AUC): AUC for training of the six brand dimensions (Average results for 10 runs with 5 iterations each)

Testing set results for the trained NN models for the six dimensions

brand dimension	average prediction for test sample	
	Positive sample	**Negative sample**
Prestige	0.805096746	0.121755627
appeal	0.913727908	0.002702125
sincerity	0.890718808	0.074712933
lively	0.782405805	0.030281347
conscientiousness	0.843108274	0.009791983
cosmopolitan	0.606904834	0.042598566

Table 5.4: Average testing set predictions for the NN models trained for the six brand dimensions

As illustrated in table 5.4, the BiLSTM model for ‘appeal’ dimension demonstrated the highest testing accuracy for positive as well as the negative sample. The final UBPS outcome for La Trobe University is demonstrated in figure 5.4.

5.7.4) Extracted University Brand Personality scale: using a combination of the three methods:

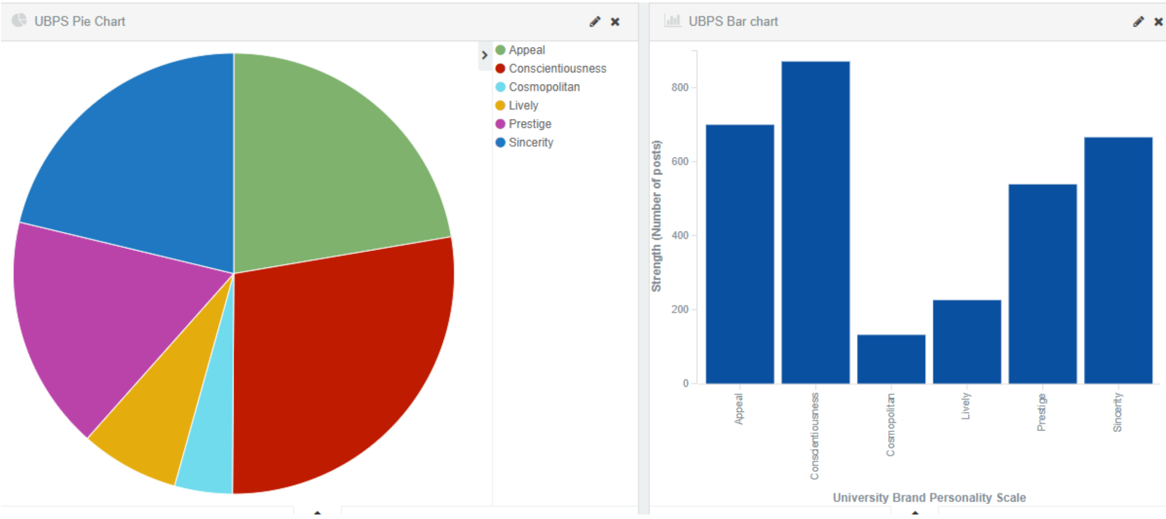


Figure 5.4: Extracted University Brand Personality Scale for La Trobe University using a combination of all three method: demonstrating the strength for the six dimensions; *appeal*, *conscientiousness*, *cosmopolitan*, *lively*, *prestige* and *sincerity*

5.7.5) Brand Harmony:

Figure 5.5 demonstrate the brand harmony of La Trobe University across several courses revealing that the *Brand Personality* is consistent within the university. Similarly, brand harmony can be extracted for different granularities such as *processes* as well as a specific *focus area* across several courses.

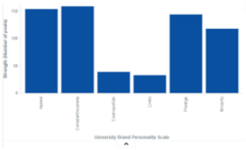
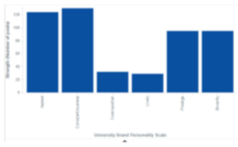
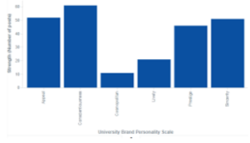
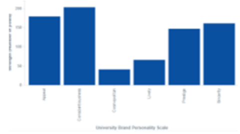
			
Law	Accounting & Business	Biological Sciences	Arts communication and critical inquiry

Figure 5.5: Brand Harmony across courses/ programs/ departments - The UBPS for different courses at La Trobe University; illustrating the strength of the *Brand Personality* dimensions; *appeal*, *conscientiousness*, *cosmopolitan*, *lively*, *prestige* and *sincerity* in order. The diagram reveals that in general, there is consistency of brand across the courses with minor fluctuations that can reveal interesting information when observed closely. The results show that the ‘Law” program has a slightly higher *prestige* than the other courses at the university (the general reputation/standing of the university’s law school validate this to be true). And that the *liveliness* is higher in the Biological Sciences and Arts compared to Law and Accounting /Business.

The diagram (figure 5.5) demonstrates that the Brand Personality for the La Trobe University is largely consistent and harmonious across the major departments. This dashboard will allow the organization/university to monitor and identify in near real-time, if any department is demonstrating a different trend which would indicate an event that is recent and significant and worth looking into. The dashboard presented in figure 5.5. reveal a slight variation in the perceived *prestige* and *liveliness* of courses which can be validated through the widely accepted understandings and believes of the university.

5.7.6) Brand consistency over time:

Figure 5.6 below illustrates the results for the six dimensions of UBPS extracted using a combination of the three proposed methods for five different universities over the period from 2003 to 2016.

Figure 5.6: Brand consistency of five uniquely selected universities over time (illustrating the performance of the six dimensions (*conscientiousness*, *appeal, lively, prestige, sincerity* and *cosmopolitan* in order) of the UBPS for the time period from 2003 - 2016)

5.7.7) Benchmarking against competitors:

The figure 5.6 above demonstrates a comparison of the extracted UBPS (with the six dimensions) for La Trobe University against four competitor universities in Victoria (selected for their unique significance) for the period 2003 to 2016.

It can be observed that Melbourne University and Monash University (which are the two Go8 universities in Victoria), exhibit a very similar pattern to each other, and the *Brand Personality* appear to be relatively stable and consistent throughout the period with minimal fluctuations across all six dimensions. For both universities 'conscientiousness' (how organized, competent, structured and effective the university is) appear to be comparatively lower. This is likely a result of the high expectations students have in relation to these top ranked universities and possibly the universities are failing to live up to the expectations.

Whereas the Charles Sturt University, which is a relatively low ranked university, demonstrates a significantly lower rating for the dimension 'prestige' and a significantly higher rating for the dimension 'sincerity' compared to other universities. This is likely due to students receiving more flexibility and consideration for university admissions compared to other universities. Correspondingly the two Go8 universities (Melbourne and Monash) are both displaying consistent lower levels of "sincerity". This is a clear demonstration of how established, high ranking universities could be falling short in relation to human touch (humanity, helpfulness, trustworthiness etc.) as they do not have a need to step up to attract and retain students compared to other less privileged universities.

Furthermore, the Australian Catholic University, which is an equity university, displays a comparatively lower level of 'appeal'. This is likely a demonstration of the lower investments on modern technology, attractiveness, productivity and superiority which is common among equity universities. However, 'appeal' seems to be increasing steadily with time across all five universities. This is likely due to commercialization and increasing adaptation of technology which increases productivity in general.

Another interesting observation is the dip in *Brand Personality* in the year 2008 across all brand dimensions and all universities. This signifies an event that took place at that time affecting universities. Some likely explanations are the curriculum reforms and university staff redundancies that took place in 2008 in par with the global economic crisis.

5.8) Discussion

This chapter proposed three novel methods to extract an organization's *Brand Personality* from social media and demonstrated their applicability and efficacy by utilizing the higher education sector as a case study and a public social media text dataset of over 1.2 million text posts.

Traditionally, *ethos pathos logos* of a brand (*Brand Personality* or stakeholders' impression of an organization's *brand*) were extracted using data collected through surveys designed specifically for this purpose utilizing instruments such as questionnaires and interviews. While these techniques have been successful, they are highly resource intensive and time consuming. These instruments need to be carefully designed using knowledge engineering which require domain knowledge and expertise. Furthermore, they need redesigning for samples and iterative phases to monitor the brand continuously.

Likewise, the surveys too need to be promptly planned and executed. Decisions need to be made on how to conduct surveys (online or offline) and how to select the samples for data collection. Again, the surveys too can require experts' participation in situations such as conducting unstructured interviews for the data collection to be successful. All of these are yet again time and cost intensive for organizations. In addition, a certain sample will always need to be rejected during pre-processing and sometimes due to varied circumstances, surveys end up as failures and require reconstruction.

Due to these strenuous procedures, the sample size that can be collected and analysed using these instruments is always limited. Also, due to the pre-structured nature of the procedure, the versatility of insights that can be extracted is; in most cases limited to what the instrument designers were able to predict and come up with at the onset of the study.

With the novel methodology proposed in this chapter, the performance results obtained for the six brand dimensions of the university brand personality scale (UBPS) illustrated how the automated technique proposed herein is capable of disclosing certain unexpected insights which would not be possible with the traditional techniques; such as how certain Go8 universities are actually underperforming related to 'human touch' and how a university generally assumed to be a low performer (due to university rankings) is actually performing well in certain areas.

Furthermore, the results presented in Table 5.2 illustrate how the automated technique captured completely new words to represent *brand* dimensions from the dataset; automatically selecting novel words highly unique to the dataset such as "*elitism*","*go8*","*snobbery*" and "*stature*" to represent the brand dimension *Prestige.* These results clearly would not be possible with the above mentioned traditional manual techniques.

Hence, this is a clear illustration of the advantages in adopting these social media based automated techniques. In traditional methods, the subjects of the study (stakeholders or customers in this scenario) are enforced to think and provide feedback in a structure provided by the questionnaires creating an unnatural environment or a thinking paradigm for the subjects. This can tamper with the accuracy of the feelings and emotions the subjects share related to a *brand* and it's products and processes.

The results demonstrate that this type of social media data can be highly effective in extracting *ethos pathos logos* of a *brand*. Three alternative techniques (decreasing human- to increasing automation) were introduced which can be used based on the choice of the amount of human intervention an organization is willing to endow. The deep learning technique used an LSTM and yielded high accuracy, due to extra knowledge gained from the word embedding layer.

Having three different methods to choose from will increase an organization's data analytics competency and also blend in existing domain knowledge and expertise with new machine learning, Natural Language Processing (NLP) and sentiment analysis techniques. This can be used as a method of gradual and incremental introduction of these new technologies into an organization which can reduce traditional reluctance and trepidation related to such technological change. Organizations will have the opportunity to select the method that will be a best fit for their institute based on the company's availability of human IT resources, technological and analytics competency and culture.

According to several existing studies, social media technologies incorporated into organizations' traditional practices, culture and institutional norms, supports organizations to embrace and benefit from social media (Huang et al., 2013), whereas direct reliance on social media feedback and technology can lead to initial tension between the technology and existing organizational/management practices and thus incorporating both in harmony leads to better organizational strategy (Baptista, Wilson, Galliers, & Bynghall, 2017 ; Huang, Baptista, & Newell, 2015 ; Baptista, Newell, & Currie, 2010). Hence, this chapter recommends a union of human and machine competencies for the contemporary organization for enhanced, continuously better performing organizational processes through a new strategy of brand monitoring for the purpose of serving and satisfying their stakeholders.

5.9) Chapter Summary

This chapter demonstrated the adaptability of social media data in the form of public online forums for organizational brand management to address the research question; "*How can an organization utilize digital stakeholder opinions present in social media to develop a model of the external viewpoint of themselves and their organizational brand, to generate insights to provide a better stakeholder experience?*".

The chapter proposed a novel technique to generate brand personalities of organizations using publicly available, freely expressed and frequently updated social media text data available in public online discussion forums. The technique was trialled, and the effectiveness proven using a case study domain, the higher education sector. The proposed technique enables monitoring of *brand image* of the organization as well as it's competitors in near real-time. Furthermore, it enables monitoring of *brand consistency* and *harmony of brand architecture* at different granularities, across a spectrum of elements within an organization. The technique has been successfully demonstrated for an Australian university, using an e-word-of-mouth (eWOM) dataset of more than 1.2 million text posts.

The proposed approach introduces a novel technique to map components of social media into accepted features of *brand*, presented in marketing literature, by extending the previously presented Social Media to Organizational Insights Transfiguration Framework (proposed in chapter 4) and by incorporating techniques of deep learning, word embedding, emotion extraction, text mining and NLP. The technique provides three alternate methods that can be adopted by organizations according to the level of human and technology intervention possible based on the resources, technological capacity, maturity and culture of the organization. Hence, this chapter presented a novel approach of maximizing human and computer capabilities for organizations by a methodology portraying union of technology, domain expertise and traditional methods for brand management.

6. Online Social Roles of Digital Stakeholders

"All the world's a stage, and all the men and women merely players. They have their exits and their entrances; And one man in his time plays many parts."

~ Shakespeare

This chapter presents an exploration into the *online social roles* played by the *Digital Stakeholder* presented in the preceding chapters in online communities to answer the research question (presented in the introduction chapter): "*What key roles do digital stakeholders play in online communities from an opinion providing and information creation and diffusion perspective and how can a generic framework of automation be provided to observe these online stakeholder roles?*".

The previous chapters focused on establishing the *Digital Stakeholder* and their relationship with the organization (chapter 3), understanding organizational perceptions of the *Digital Stakeholder* (chapter 4) and establishing an organizational brand through stakeholder perceptions (chapter 5) captured from social media conversations.

This chapter focuses the lens on this influential digital stakeholder's behaviour in terms of the varied dynamic online roles performed in online social media communities from an opinion providing and information creation and diffusion perspective that impacts the image and brand of organizations. Having established the value and importance of understanding stakeholder conversations in social media in order to understand stakeholder opinions and emotions related to an organization; *social roles* were identified in this research as a highly useful and versatile tool for organizations to take a step further in knowledge discovery to understand 'who' is responsible and 'why' for creating these online conversations (that were utilized in chapters 4 and 5 to generate organizational insights) and drawing-in and diffusing information related to the organization from the real world into the digital world and influencing the organization's image and brand.

Social roles were identified as a key concept proven to provide a structuring, coordinative, and supportive function for communities in research literature that can facilitate an organization to discover and monitor their online influencers from the vast unstructured social media conversations of high volume and velocity.

By utilizing theory and concepts related to online social roles found in communication and media literature, together with a novel application of the honeycomb social media framework and a combination of machine learning techniques, this chapter presents how social media conversations can be transformed into online social roles by presenting a generic framework of automation to observe online stakeholder roles for organizations. The proposed framework presents a methodological approach for automated detection of social roles in online communities by utilizing a deep recurrent neural network, a word embedding model and the stated novel application of the honeycomb framework. This technique can be applied to any online community to automatically identify social roles, their influence and interactions. Given the large volume of text (in social media) and the value of content, it is no longer viable to manually encode and detect social roles and contributions.

Furthermore, the chapter introduces the traditional concept of online social roles from a novel lens/perspective; as a technique of adding a layer of rules to social media data to reduce its complexity, by enabling social media conversations to be viewed through the lenses of social roles. Also, the presented research contributes to the growing body of literature on online social roles by identifying six roles built upon/based on existing literature specific to the case study domain; higher education(HE) sector to monitor stakeholder opinions and behaviours from public forums. A dataset consisting of over 1.2 million textual posts extracted from an online community relevant to the case study domain was used to demonstrate the technique. (Wijenayake et al., 2020).

6.1) Introducing online social roles

Chapters 2, 3, 4 and 5 established that online communities are an increasingly important aspect in the digital age, for business organizations, diverse industry sectors and overall, for modern society. Online virtual communities have been in existence on the Internet for many decades and have drawn the attention of researchers and businesses alike (Ridings & Gefen, 2004). In response, a large body of research has emerged with online communities and social networks, in areas of members' online behavior, members' characteristics, members' motivation analyses and the dynamic flow of information and resources (Benamar et al., 2017).

One of the main concepts used by researchers in varied domains to observe online communities are social roles (Benamar et al., 2017; Wijenayake et al., 2020). A social role perspective highlight patterns of recurring behavior in a community (Jahnke, 2010). Hence, a social role has often been defined as a set of activities performed by individuals, and the range of expected behavior within a group (Jahnke, 2010). Research shows that social roles provide a structuring, coordinative, and supportive function for communities (Benamar et al., 2017). The social role of each end-user, influencers to followers, and content providers to receivers is a primary consideration when evaluating the purpose and contribution of any online community. Online social roles have been identified throughout literature using ethnographic studies of the content of interactions (Himelboim et al., 2009; Wijenayake et al., 2020).

The critical role of social classification in knowledge creation and its contribution towards the evolution of scientific fields is highlighted in current research (Bail, 2014; Gieryn, 1999). In the context of online communities research, social roles have been identified via patterns generated by the exchange of messages in reply to one another (Himelboim et al., 2009). Extending the works of (Füller et al., 2014) and (Johnson et al., 2015), a study by (Benamar et al., 2017) provides a rich core roles typology by defining roles present throughout literature more specifically through deep analysis of activity, content and structure of social network. For example, two significant social roles based on distinct patterns of communication are *Question Person* and *Answer Person* (descriptions presented in Table 1). These two roles were observed in many previous studies (Benamar et al., 2017; Fisher et al., 2006; Turner et al., n.d.; Welser et al., 2007) and were identified as two distinct populations through the construction of egocentric social networks, through patterns of reply and via visualizations of patterns of initiation, reply, and thread contribution rates over time (Wijenayake et al., 2020).

When researching into the methods used in past studies for the extraction of online social roles, we discovered that mainly manual coding methods, qualitative analysis, discourse analysis, occasional cluster analysis, hierarchical and partitioning methods have been used (de Valck et al., 2009; Mooi & Sarstedt, n.d.; Ridings & Gefen, 2004). Tools such as Microsoft Research Netscan have been popularly used (Himelboim et al., 2009).

A further study by Gunawardena et al. used quantitative content analysis and studied only 80 discussions (Gunawardena et al., 1997). Five independent coders were used to manually code the levels of knowledge construction in the network. The quantitative content was applied based on an interaction analysis model introduced by a previous study (Gunawardena et al., 1997; Wijenayake et al., 2020).

Furthermore, another study used the popular in-degree, out-degree measures, a social network analysis using Gephi2 software graphs, 1,150 manually coded posts and ran a cluster analysis based on six variables (Benamar et al., 2017). The six variables were number of posts (member's engagement); number of likes and comments (member's reactions); number of received likes and comments (member's influence); average number of received comments per post (feedback density); average number of given comments per post (reaction density); self-comments (weight of comments on own posts in total of member's comments, indicating whether the member is oriented toward others or self-oriented). A recent literature review revealed social network analysis and content analysis research methods are the primary techniques that have been widely applied to study online social roles and introduces a multi-method approach based on social network analysis, content analysis, statistical analysis, and social-cognitive network visualization (Ouyang & Chang, 2019; Wijenayake et al., 2020).

6.2) Existing work and the research gap

As summed by Bail (Bail, 2014) and summarized in the literature review (section 2.3), while the rise of the Internet and social media has produced a colossal volume of text-based data in recent years, and computer scientists have produced powerful new tools for automated analyses of big data, there is a lack in theoretical mapping necessary to extract meaning from this data. On the other hand, cultural sociologists have produced sophisticated theories of the social origins of meaning, but lack the technical capacity to explore them beyond micro-levels of analysis (Bail, 2014; Wijenayake et al., 2020).

A wide range of role conceptualizations and methodologies have been used in the context of online communities (Benamar et al., 2017). However, most early studies have relied upon in-depth interviews or case studies that highlight the social construction of ranking within groups (Bail, 2014). To quote Bail, "cultural sociologists have scarcely explored the promise of automated text analysis to classify texts. Where these techniques have been used they have been relatively primitive approaches to automation that simply identify keywords or phrases" (Bail, 2014; Wijenayake et al., 2020).

Bail further suggests qualitative coding techniques pioneered by cultural sociologists and anthropologists can be leveraged to improve already powerful automated text analysis techniques produced by computer scientists and recommends application of text classification techniques that rely exclusively on computer algorithms to create meaningful groupings of texts (Bail, 2014). Furthermore, on text mining for social theory (Evans & Aceves, 2016), it is recommended that computational approaches with machine learning methods can now make substantially improved inferences acting as extensions of human cognitive capacity; in other words as cognitive prosthetics. However, these techniques have not yet been demonstrated on social roles classification (Wijenayake et al., 2020).

Furthermore, in a review on current state of social network research (Riemer et al., 2015), it is suggested to use mix methods (combination of trace data and qualitative approaches) to best address social network research questions with big data analytics, illustrating how existing work have mainly used either quantitative or qualitative techniques alone. For example, Riemer et. al (Riemer et al., 2015) used standard statistical approaches to analyze 61,945 messages created by 3158 users of an enterprise social media system and Bhattacharya et. al (Bhattacharya et al., 2015) analyzed social media data for 2507 undergraduate students producing a mean 145 posts each, using Markov Chain Monte Carlo techniques to analyze big data (Whelan et al., 2016). A study by Nicholas and Lee (Nicholas & Lee, 2017) used text clustering techniques to analyze factors related to customer satisfaction in hotel industry from online reviews in the Trip Advisor website (Wijenayake et al., 2020).

To address this research gap, we introduce a novel generic framework for automation of online social roles using a machine learning approach. This framework enables automation of any social role based on a classification system that includes all existing social roles as well as any that could be defined in future. The technique is built using Natural Language Processing (NLP) techniques, a deep recurrent neural network and a word-embedding model. The framework has been empirically tested and evaluated and effectiveness proven using an online community of over 1.2 million text conversations in the case study domain; higher education.

6.3) Proposed Technique for automating the extraction of online social roles

The development of the proposed technique was executed in several steps , requiring several phases of study as presented in this section: selection of social roles for experiments, selection of a data source to conduct the experiments and testing, utilization of the honeycomb social media framework, exploration for a universal categorization of social roles and exploration into the functionality and capabilities of deep learning.

6.3.1) Selection of social roles for experiments

First, we commenced with experiments to understand the value of observing online social roles for an organization and to experiment approaches for automation of social roles.

Our first observation was that social roles are varied and diverse, with a large variety of roles which are at times ambiguous, highly context dependent, social media platform dependent, are not mutually exclusive and have continuously been introduced based on the domain and requirements at hand by researchers over decades.

After much exploration and analysis, we excavated six social roles from existing literature as relevant and of value for the case study organization's domain; higher education and applicable to our selected (as presented in the introduction and literature review chapters) social media category of focus; public online discussion forums. Out of all of the social roles present in literature, first a subset of social roles that are representative of the *Digital Stakeholder* introduced and presented in chapter 3 was created. Then, using this *Digital Stakeholder* subset, two more subsets were created; one for social roles that can be discovered using the features of public online discussion forums (features presented in chapters 4 and 5) and the second subset for social roles that are applicable for higher education organizations using features of higher education organizations that were discovered and presented in chapter 4. Next, the overlap area of these two subsets were used to filter and create the set of social roles that are representative of the *Digital Stakeholder*, can be discovered through online public forums and are applicable for higher education organizations. These six roles are presented in table 6.1 and were evaluated to be of value for higher education organizations to monitor, for the purpose of understanding the behavior of key *Digital Stakeholders* who influence information creation and diffusion in an online community of interest.

The table 6.1 presents each of these six identified roles with the justification for selection. These roles will enable a higher education institute to identify who are the *gate keepers* in an online community; the key stakeholders who decide which topics related to their organization will get discussed and therefore which information will be diffused into the online community from the physical world. *Questions Person* enables discovery of the *Digital Stakeholders* who are seeking information related to the organization. *Answer Person* will enable discovery of the content creators who are pushing information related to the organization into the online community from the real world. *Active contributor, Passive contributor and Self-Centereds* enable observation of the *Digital Stakeholders* that play varied roles of engagement in the particular social media platform /online community of interest.

Social Roles	Justification for selection: Applicable for HE forums
Gate-keeper (Benamar et al., 2017; Himelboim et al., 2009; J. Xu et al., 2004) (used the definition from early literature)	Conversation starters. Decide what topics are discussed. Responsible to draw information into the social media network from external sources by activating other users.
Question person (Benamar et al., 2017; Fisher et al., 2006; Turner et al., n.d.; Welser et al., 2007)	Discussion catalysts: Ask questions, information seeking behaviour. high reply attracting participants.
Answer person (Benamar et al., 2017; Fisher et al., 2006; Turner et al., n.d.; Welser et al., 2007)	Content Catalysts: Provider of helpful information, Expertise.
Active contributors (Benamar et al., 2017)	Active participants of the social network.
Passive contributor/lurker (Benamar et al., 2017)	Participants of the social network.
Self centereds (Benamar et al., 2017)	Special type of active participant. Nature of seeking behaviour- more centred than distributed, more posts on their own threads.

Table 6. 1. Description of the six online social roles identified as representative of the *Digital Stakeholder* (proposed in chapter 3), discoverable through public online discussion forums (identified as the focus social media category in the research gap presented in chapter 1) and relevant and of value for the case study organization domain higher education(presented in chapters 4 and 5) from literature with the justification for value creation (Wijenayake et al., 2020).

6.3.2) Selection of data source for experiments

As the data source to conduct experiments on the selected social roles, public online communities were selected as these are a type of social media rich in stakeholders' opinions, expectations and experiences related to many industries and thus a highly valuable source for insight generation that remains underutilized (As justified in section 2.3.5 in the literature review and reiterated in chapters 4 and 5). In order to demonstrate this approach, as applicable to the case study organization domain, we selected the higher education (HE) sector of Australia (Duesterhaus & Duesterhaus, 2014; Khanna et al., 2014; Wijenayake et al., 2020). The following three online communities used for experiments in chapters 4 and 5 were again selected for experiments on online social roles of the *Digital Stakeholder* as the value and credibility of these sources were proven and justified by the outcomes generated and presented in the preceding chapters that utilized them for generation of the *Digital Stakeholders'* perception of organizations(*Perceived Organization* in chapter 4) and for the creation of organizational brand based on the perceptions of the *Digital Stakeholder*(chapter 5) using conversations on public online forums.

- Whirlpool (http://forums.whirlpool.net.au/)
- ATAR Notes (https://atarnotes.com/forum/)
- Bored of Studies (http://www.boredofstudies.org/)

Furthermore, existing research report an increased presence of social roles in such communities (Füller et al., 2014). In order to identify the areas in the selected online communities to engage with for maximum impact, the honeycomb framework proposed by (Kietzmann et al., 2011) was used as proposed in section 2.2.10 in the literature review and followed through in chapters 3, 4 and 5. As elaborated in previous chapters, this framework was selected as it is the currently most widely accepted framework for social media and the application disclosed that in terms of public online discussion forums, “Conversation” and “Groups” are the building blocks or areas to engage with this social media for impact. Thus the dataset extracted from conversations in higher education groups from the public online communities (utilized for experiments in chapters 4 and 5) were used for the experiments (Wijenayake et al., 2020).

6.3.3) Application of Honeycomb Framework

As previously presented and justified in chapter 4; section 4.2 of this thesis, the honeycomb social media framework was applied to identify the building blocks to engage with for the social media selected for the experiments. Since again public online discussion forums were utilized, the primary building block uncovered and utilized in previous chapters; *conversations* were utilized again.

Furthermore, we believed application of this framework may generate interesting insights as achieved in chapter 4 section 4.5 in mapping components of social media to organization. Table 6.2 presents the application of honeycomb social media framework to map components of social media into components of social roles.

The honeycomb framework enabled to add structure into unstructured social media posts to discover possibilities to extract knowledge of value to the organization through the representations and definitions of varied online social roles. The six social roles presented in table 6.1 were selected based on the justification for value creation generated by the application of the honeycomb social media framework presented in the table 6.2. This research presents the first application of the widely accepted honeycomb social media framework to discover online social roles in social media.

Honeycomb building blocks → Roles ↓	H1: Conversations	H2: Groups	H3: Reputation	H4: Presence	H5: Relationships	H6: Identity	H7: Sharing
Role 1- **Self-Centereds**	✓ Which topics, processes, institutional granularities influence self-centred behaviour and why?	✓ Which groups have more self centereds and why?	✓ The nature of profiles of self centered users	N/A	N/A	N/A	Relative amount of sharing of external content compared to other roles?
Role 2- **Question person**	✓ Which topic areas, processes, institutional granularities get more questions?	✓ Which groups have more questions persons?	✓ The nature of profiles of Question person users?	N/A	N/A	N/A	Relative amount of content shared- in general and per post?
Role 3- **Answer person**	✓ Questions in which topic areas/focus areas, processes, institutional granularities get more answers (more Answer persons) and why?	✓ Which topics/questions draw most answers?	✓ The nature of profiles of Answer person users?	N/A	N/A	N/A	How much of external content is present in answers?
Role 4- **Gate keeper**	✓ Which topics, processes, institutional granularities are focussed by topic starters and how it changes over time?	✓ Gate keepers' distribution across groups?	✓ The nature of profiles of gate keeper users?	N/A	N/A	N/A	Any usage of external content for starting topics?
Role 5- **Discussion catalyst/ Active contributor**	✓ Which topics, processes, institutional granularities draw discussion catalysts and how it changes over time?	✓ Discussion catalysts' distribution across groups?	✓ The nature of profiles of distribution catalyst users?	N/A	N/A	N/A	Relative amount of content shared- in general and per post?
Role 6- **Passive contributor/ Lurker**	✓ Which topics, processes, institutional granularities draw most lurkers and how it changes over time?	✓ Lurkers' distribution across groups?	✓ The nature of profiles of Lurker users- lower reputation indices?	N/A	N/A	N/A	Relative amount of content shared- in general and per post?

Table 6.2: Application of the honeycomb social media framework to the identified and selected social roles relevant to the case study organization domain; higher education. H1, H2 and H3 present the three most significant building blocks for public online forums. H4, H5 and H6 represent three building blocks that are irrelevant or insignificant for this category of social media (public online forums) as users' presence, relationships and identity are not shared or represented.

6.3.4) A universal categorization of social roles for an Automation Framework

In order to introduce a generic framework to automate the extraction of online social roles, it is essential to come up with a categorization technique that would include all social roles.

While it is not practical to list all social roles presented in literature, our research uncovered that all roles can be categorized using a framework that have been introduced and tested in recent studies.

Through the literature review, it was observed that all of the online social roles can be categorized into a typology of three categories based on the method used to identify them. A study conducted by (Benamar et al., 2017) introduces three categories extending the work of (Füller et al., 2014) and (Baym, 2012) (Wijenayake et al., 2020).

The three categories of social roles are:

1) Activity analysis (based on members' activity): participation level in group discussions, community core's identification; comments, likes, sum of posts.

2) Content analysis (based on shared content of members): post content, analysis of content and context.

3) Structural analysis (based on member's position in network): systemic approach by considering both member's position & interactions.

Mapping social roles into these three categories enables positioning of roles, comparing and contrasting different social roles, conceptualizing social media conversations, standardizing scattered roles as well as adding more meaning and value to social roles (Wijenayake et al., 2020).

This categorization presented a significant opportunity for our automation endeavor. The six social roles that were selected for experiments (presented in table 6.1) were grouped using these three categories. All six roles relevant to the case study domain isolated into the first two categories; categories of content based and activity based social roles as follows.

Categorization of the six social roles selected for experiments into the three categories based on method used to identify the role:

1) Activity analysis (members' activity): *Gate-keeper, Active contributor, Passive contributor/lurker, Self centereds*
2) Content analysis (shared content): *Question Person, Answer Person*
3) Structural analysis (position in network): *Nil*

This categorization was selected as the foundation to generate procedures to automate extraction of all social roles in literature based on the method used to identify and describe the role. The proposed solution is introduced and presented in the next section; *A novel framework for automation of social roles,* and summarized in table 6.3 (Wijenayake et al., 2020).

6.3.5) A novel framework for automation of social roles

First, a mapping of requirements for automation was performed for the six selected social roles (see Table 6.3). Then, each of the identified requirements for automation were mapped with machine learning (ML) and Natural Language Processing (NLP) features and techniques that are capable of solving the requirement (Table 6.3). Thereafter, a set of recommendations were derived from ML and NLP techniques to extract the social role and summarized and represented in table 6.3 (Wijenayake et al., 2020).

Social Role category	Requirements for Automation	Features of Machine Learning	Recommended ML/NLP techniques to extract the social role
Activity analysis-based roles (eg: Gate-keeper, Active contributor, Passive contributor/lurker, Self centereds)	classification of users, Heuristics extraction from SM activity, counts, frequencies, statistics, splitting into sentences & words, filtering out punctuation & stop words.	predictive modelling, classification, probabilities, optimization	heuristics based automation, Decision tree, Naïve Bayes, SVM Tokenizing, normalizing (text cleaning)
Content analysis-based roles (eg: Question person, Answer person)	classification of users, ability to identify questions from text sentences, ability to identify answers from text, splitting into sentences & words, filtering out punctuation & stop words.	text classification, text sequence learning, modelling of words specific to questions, text cleaning, extracting sentences, feature extraction, structure: nouns, verbs.	LSTM+ word embedding, bag of words, POS tagging (using nouns, verbs), word counts, word frequencies, tokenizing, normalizing (text cleaning), stemming, parsing, hashing (for optimization of algorithms)

Table 6.3. Suggested machine learning methods to automate the capture of online social roles from social media (SM) (Wijenayake et al., 2020)

Our experiments succeeded in demonstrating the applicability of the recommendations presented in table 6.3 by extracting activity based social roles and content based social roles from the dataset of 1.2 million textual posts from public online discussion forums in higher education and is presented in the proceeding sections 6.3.6 and 6.3.7 (Wijenayake et al., 2020).

6.3.6) Heuristics based technique for extraction of Activity based social roles

We propose that a simple heuristics-based procedure can be developed for activity based social roles (table 6.3). Furthermore, we demonstrate a successful application by presenting a simple procedure (illustrated in figure 6.1) we developed for extraction of gate keepers from the public forum posts. This procedure was implemented using a Python program to successfully identify gatekeepers from the public forum dataset of over 1.2 million conversations.

The procedure (figure 6.1) presents a threshold value(x) that can be changed according to the requirements of the organization. For example, whether the organization wants to monitor the top 1% of gate keepers from the online community population or more/less.

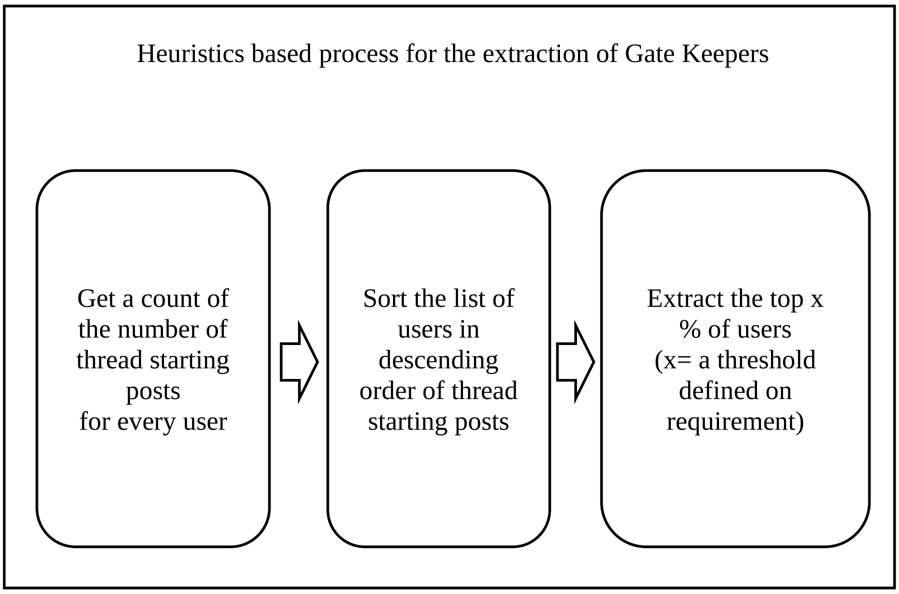


Figure 6.1: Automated Extraction Method for Activity based social role: *Gate keepers* from public forums.

This procedure was successful in enabling the identification and monitoring of fluctuations of gate keepers over time including any seasonal or interesting behaviours.

Some example applications:

- Get a list of all gate keepers for 5 years.
- Get lists of gate keepers for every year separately.
- Divide year into sections as applicable to organization domain (e.g. semesters for higher education institutes) and observe seasonal behaviour of gate keepers.

A heuristics-based procedure for extraction of Self Centereds:

Self centereds comprises of members with most of their comments made on their own posts and rarely *like* or *comment* on other members' posts.

<u>Procedure to identify Self centereds:</u>

1-Define a self-centered according to current requirement and set a value for the variable/threshold parameter (e.g: members with more than 50% of comments made on their own posts are defined as self-centereds, then 50% becomes the threshold value)

2-Filter all users who have started at least one thread. (msg index=0)

3-Get count of all posts/comments for each user

4- Get counts of posts on *own thread* and *other*

5- If own thread count > 50% => self centered

Consequently, automating the extraction of activity based social roles using the proposed recommendations from table 6.3 was demonstrated to be a straight forward and simple task.

One of the key challenges faced in this research was the automation of extraction of content based social roles from public forums. This required a suitable machine learning strategy to be selected. Our method utilizes deep learning and accordingly section 6.3.7 presents a brief introduction to deep learning and justification for selection as a key technological building block to solve the challenges faced in identifying and categorizing content based social roles from public online discussion forums.

6.3.7) Deep Learning based techniques for extraction of Content based social roles

While automating the extraction of Activity based social roles was proven to be a relatively straightforward task, Content based social roles were more challenging to automate due to the need to 'read' and 'understand' the content of text posts.

Hence, we propose a deep learning-based solution that is capable of successfully automating this 'read' and 'understand' steps for text content.

Deep Learning is a Representation Learning method. Representation Learning stands for a collection of methods that allows a machine to automatically discover representations needed for detection or classification from raw data. Deep-learning methods use multiple levels of representation, obtained by composing simple, non-linear modules that transform the representation at one level, into a representation at a higher more abstract level, starting with the raw input. A composition of multiple such transformations allows even highly complex patterns to be learned. For classification tasks, higher layers of representation amplify aspects of the input that are important for categorization (LeCun et al., 2015; Wijenayake et al., 2020).

There are several deep learning architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Network (RNN). To automate the extraction of content-based social roles, we propose a method that uses a novel deep RNN of Bidirectional Long Short-Term Memory (BiLSTM). Long Short-term Memory (LSTM) is used for sequence learning. A BiLSTM uses a forward pass Neural Network (NN) and a backward pass NN to learn sequences in both directions. Since deep learning allows computational models consisting of multiple layers of processing to learn data representations such as sequences with multiple levels of abstractions, it has dramatically improved the state-of-the-art in many domains that require sequence and pattern recognition such as speech recognition, hand writing recognition and genomics (LeCun et al., 2015). The combination of these methods with the LSTM-RNN architecture have yielded successful results in past studies related to sequential data modelling (Graves et al., 2013; Schmidhuber, 2015). To implement and demonstrate our technique, we used Python 3. For the implementation of the deep learning model, Python's NN library Keras was utilized (Wijenayake et al., 2020).

In our method, as demonstrated in figure 6.2, first; a training set and a testing set were created for each of the social roles. A negative and positive set was created for each set. A set of 200 posts were used for each of the training sets and a random sample of 100 posts were used for the testing sets. Next, a word2vec word embedding model trained using the dataset was fed into the NN as an intermediary layer to provide extra knowledge (depth) for training (illustrated in figure 6.2) (Wijenayake et al., 2020).

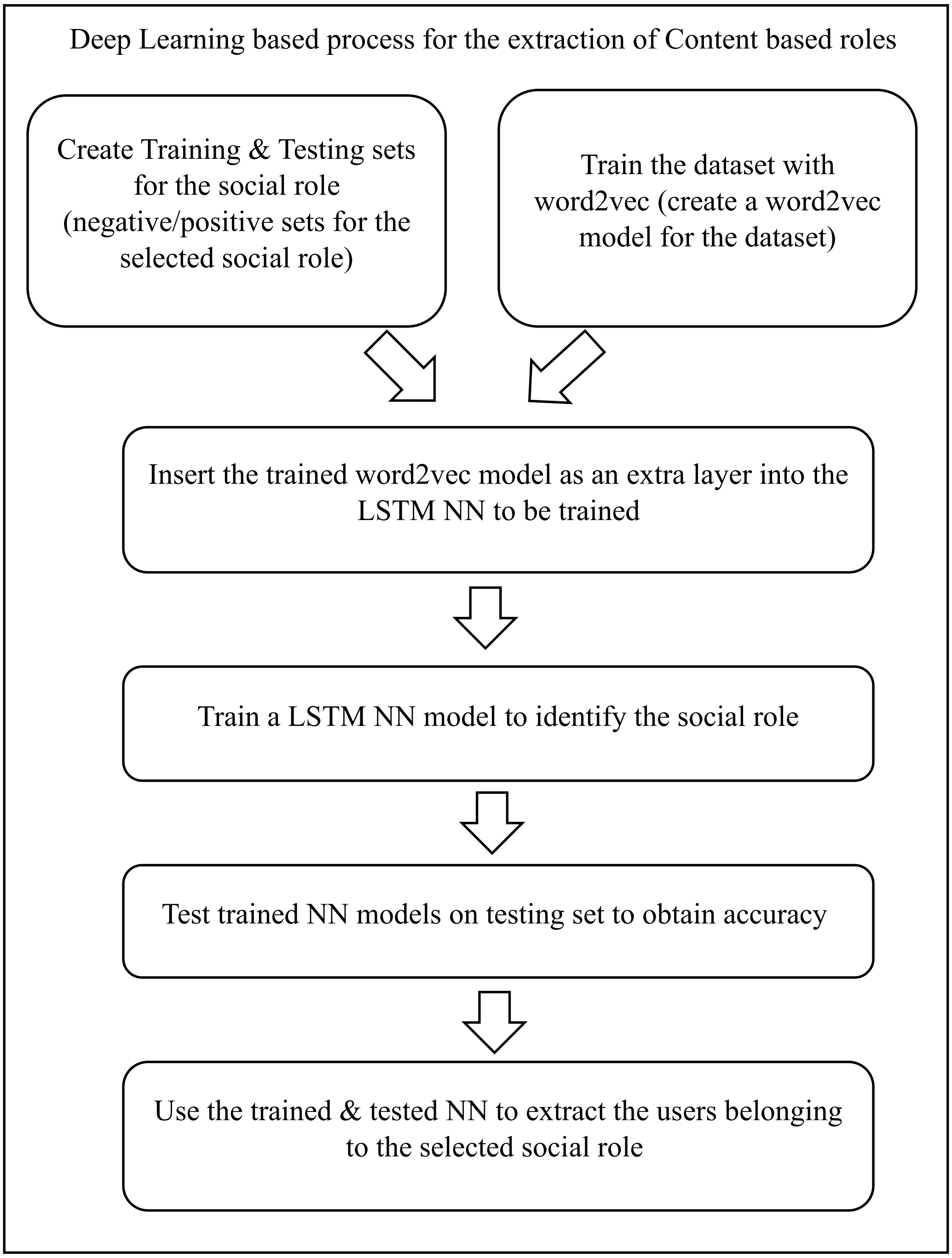


Figure 6.2: Automated extraction method for content based social roles using LSTM and Word Embedding (Wijenayake et al., 2020)

Then, the BiLSTM NN was used to train and identify the social role (Table 6.4). Each social role (*question person* and *answer person*) was identified independently using the created training datasets resulting in separate trained models for each role. The validation sets were randomly created (through automation) for 5 epochs or iterations and the testing phase was repeated for 20 runs to maximize accuracy (Wijenayake et al., 2020).

To observe accuracy of the method, values for Area Under the Curve (AUC) were calculated for 20 runs with 5 epochs (iterations) each. Table 6.4 illustrates the AUC values obtained for each of the two content based social roles presented. In our results, high accuracy levels were obtained for each of the two social roles with the AUC approaching 1 for the positive sample and AUC approaching 0 for the negative sample (Wijenayake et al., 2020).

As the next step, the trained NN models were tested using the testing sets to obtain accuracy levels (Table 6.5). Finally, after checking the acceptability of accuracy levels, the trained and tested NN models were used to extract the content based social roles from the online community (Wijenayake et al., 2020).

Content based Social Role	**AUC**
Question Person	0.815
Answer Person	0.921

Table 6.4: Training accuracy - Area Under the Curve (AUC) for training of the content based social roles.

Social Role	average prediction for test sample	
	Positive sample	***Negative sample***
Question Person	0.720013	0.200483
Answer Person	0.784021	0.135042

Table 6.5: Testing set results for the trained deep neural network models for the two content based social roles; *Question Person* and *Answer Person.*

Tables 6.6 and 6.7 below, demonstrate some sample results; text posts from the social media data set that were categorized by this algorithm as “Questions/Information seeking” and “Answers/information providing” behaviours. The categorized samples illustrate the effectiveness of the proposed novel algorithm (for identifying content based social roles/ content based behaviour) as well as the diversity of conversations (related to an organizational domain) present in public online discussion forums as an underutilized category of social media for business value creation.

Once the most challenging part of classifying posts based on post content has been done the post elements can be tagged with the behaviour (Questions or/and Answers) as presented in figure 6.3 (An extension of the figure 4.5 presented in chapter 4). Next, identification of *Question Persons* and *Answer Person*s can be achieved by using a simple heuristics procedure (similar to the procedures presented above in section 6.3.6 for the extraction of activity based social roles) to count the number of posts in each category by each user and identify users with high volumes of questions and answers by defining threshold values according to preference.

For example, for the case study organization, users with over 50% of their posts classified as Questions/information seeking behaviour were labelled and observed as *Question Persons*.

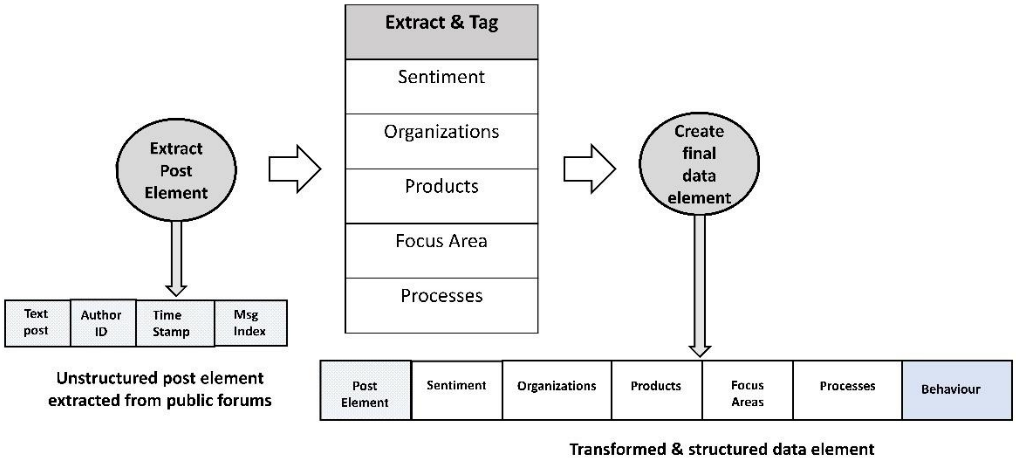


Figure 6.3: The updated data transformation diagram (The extension to the transformation technique proposed in figure 4.5 in chapter 4 highlighted in blue) to convert unstructured raw text data available in public online discussion forums (*post elements* in the diagram) into structured data with the extracted knowledge components embedded to support analytics and visualization (*final data elements*). The figure illustrate the extension of adding the *behaviour* data (e.g: Questions/Answers) related to the content of the post for extraction of content based social roles into the final data element and also the extraction and inclusion of the *msg index* (related to the positioning of the post in it's discussion thread) into the *extract post element* step.

A sample of questions posts (information seeking behaviour) captured by the proposed technique

Title of thread	Content of the post categorized " Questions/seeking information"
Griffith or UQ	O.P. Ive been offered a excellence scholarship at both unis but cant decide which 1 to choose. I was originally going to attend griffith for the bachelor of engineering (advanced with honours) and do the honours college but I am not certain now. Travelling to UQ will mean 1hr 15 one way or i will have to live up there which is over $15k/yr for on campus. Is UQ that good to choose over griffith for engineering and worth it or not really? Thanks for all your help
Cross-institutional transfer	O.P. I have applied for radiography/medical imaging through QTAC as my primary preferences at both CQU and QUT. My first choice is at QUT rather than CQU. The problem is I am more likely to be offered a place at CQU which isn't a problem because I can do the first year externally and doesn't require me to relocate until the second year. If i get into CQU for first year would i be able to transfer to QUT for second year onward? Does anyone have any information/experience on cross-institutional transfers, specifically in radiography? Would I have to repeat any units?
Radiography in Sydney	O.P. I've applied to USYD bachelors and masters, newcastle and CSU. What are my chances of getting into any of these courses with a TER (atar) of 94 but a WAM from a completed degree just under 60? Curtin uni would of been my preferred course, they let you use your TER instead of uni results, but I didn't do the prerequisite subjects of physics and maths so I don't think I can get in.
Best OUA units for science,electronics?	O.P. I'm really wanting to get into either one of these courses at Macquarie uni: Bachelor of science with a major in electronics: <url>Bachelor of engineering with the degree of Bachelor of science with a major in electronics: <url> As a pathway to these courses I am planning to take OUA units over 2014 so that I am eligible to apply for these courses in 2015. What would be the best units to complete at OUA as a pathway to these courses? Thank you!
not declaring quali' to get gov' spot	O.P. just curious if anyone has not declared a past qualification when enrolling into a new course and still got the gov subsidised spot. thanks in advance
Software or Civil Engineering	O.P. I'm a QUT student and I just switched from Interactive Entertainment to Engineering because there's a lot more opportunities and money, and it's in one of my main interests to acquire the dosh. I am most interested in Software at the moment and I enjoy programming, however I'm sure I'd learn to like Civil if I chose it, and I'm enjoying Physics/Mechanics/Statics at Uni. What are the opportunities and salaries like for Civil and Software Engineering? Anything else I should know? I think that Civil Engineers get more at the moment but I'm not altogether sure... I think it may or may not be because there are people who unfortunately don't consider Software Engineers 'real' engineers.

Table 6.6: A sample of the text conversations (post content and thread topic the post belongs to) categorized as "Asking questions/ seeking information" behaviour by the proposed algorithm from the higher education data set. Text segments representing *Questions Person* behaviour has been highlighted in grey. The categorized sample illustrate the effectiveness of the proposed algorithm as well as the diversity of content related to an organizational domain present in public online discussion forums (as an underutilized social media).

A sample of posts "providing answers/ information"- classified by the proposed algorithm

Category of information provided	Content of the post categorized "Answers/ Providing information"
Examinations	what are you majoring in? if it's accounting or finance, you'll need a calculator especially for exams. a basic scientific one is ok if you're not doing finance... otherwise might be best to get a financial calculator.
Courses	I started a JD at USYD and left to do an LLB(Hons) graduate entry through QUT. I would say, without a shadow of a doubt, that the study load and expectations were more onerous in QUT's program. This might be due to QUT taking a more practical approach to their program (assessments include client interviews, letter drafting, negotiation, etc), rather than USYD's very theoretical approach. But as someone who's experienced both types of programs, I strongly attribute the JD to be nothing more than a regurgitated and re-branded LLB. The University of Queensland caught on pretty quick, and have scrapped their JD program. For those contemplating doing a JD, I would implore you to consider the "scarlet letter" curse that's plagued a lot of American JD graduates. If you are not intending to use a JD to practice law, you will likely be in a worse employment and financial position that you would have otherwise found yourself in if you didn't pursue it. TL;DR \u2013 JD's are money making machines, and an economic bubble. Don't be in the bubble when it bursts.
Course selection	Im doing first year electrical engineering and im loving, this decipline is so huge! it encompasses power, telecommunications, electronics, computer systems, control and automation, biomedical, mechatronics engineering and others... BUT TRUST ME, if you dont like maths or if you dont do well in it, then engineering is not for you full stop. Not only is there a lot of maths in engineering, but even the other engineering units need a mathematical mind to tackle them.
Degree types (Honours , PhD)	Honours is just an additional two semesters long research component. For 3 year degrees, if you do one and then get a high result, you can start your Ph.D right after. For most 4 year Bachelor degrees(such as Engineering), the you have to write a thesis, and that is sort of your "honours year", however your honours award will be determined by your entire 4 year coursework. For instance at QUT, if you get GPA of 6 or above for any 4 year bachelor degree you would get first class honours award.
Substitutes/ alternatives	I would recommend you try completing a cousera or edx course. You can complete them for free. You can use these courses to bridge your skills and establish if you would prefer to study online or physically at a UNI. https://www.coursera.org/learn/introduction-to-research-for-essay-writing
Employability	Don't fall for that, most companies (medium to small) are not too fond of hiring academics because they are too theoretical. Universities usually talk a lot when it comes to Industrial training. Engineering is not a science field, i.e. you have to able to provide practical solutions to your clients. Good but they really need to be directors or associates, as this is not year 10 work experience.
Comparison of quality/ course content	O.P. Well my grandfather who lives in Texas owns and is the director of an oil drilling company over there. I know Engineering is not a science field, thats why Griffith provides more practical work as compared to theoretical work, whereas UQ offers more theoretical work. I can understand that some UQers think they are better than most Griffith students, but that doesnt mean all. I've been offered a scholarship for a reason, one also from UQ which must proves i must be up there with the cream of the crop :) Maybe the electrical engineering course may seem abit easy but the civil is alot harder.
Examination tips	Have a late night but DON'T do an all nighter, you won't take anything extra in and will be too tired to do your best in the exam. TBH you are in trouble but for damage control I would simply read the introduction and conclusion of each textbook chapter (if applicable) and make some notes. I would also try to do any practice exam papers if you have them and memorise the answers/methods. Ditto any tutorial questions. Otherwise, go through the pages and pages of notes, write out the main headings and try to write a few lines summary on each.

Selection of study mode	I'm a distance education student, at present. After studying on campus, I can strongly reccomend distance education to those serious students who don't have the time or resources to study on campus. However, I do agree with the above. Languages via distance often require you to record your reciting activities and send them in. Ideally, you should be on campus so you can really benefit from the interaction between the prof/tutors and other students. Philosophy would also be better on campus, overall, due to the injected debate between peers. It can be done, though, via distance \u2013 but would take a lot of dedication in order to get the most out of it. OUA is fine with most of their degrees. The one exception, though, is Griffith's Arts degree. Stay away from it like the plague. All of the others are good, though, from what I hear.
Employability related to courses	An engineering degree won't necessarily land you a good job in the IT&Telco arena. This is different to mining and civil engineering, where the degree is a requirement. In IT & Telco, it's only a requirement for a handful of companies. Most places want experience first. But an engineering degree with experience and certs will trump over someone with experience and certs without the degree. The important thing is - you will never study so hard as you did for your degree, the rest is easy ;)
Software and tools: recommendations and tips	Linux itself is the kernel, which is released under the GPL. It is freely available. Many people think of Linux as the complete package, which it's not. Strictly speaking, Linux is only the kernel; a distribution consists of all of the other components you get. However, these days, Linux has become more commercialised. You can look at the Fedora Core, Ubuntu or Debian distributions. There are quite a few others around the place, but these will certainly allow you to experience Linux. I'd recommend the VMware approach first, as per below. Yes, VMware is a good way to experience Linux without dual boot setups or dedicated machines. Of course, depending on how much RAM you have, and how much you can or will dedicate to a VM, the use of a GUI environment might be somewhat sluggish. J.
Assignment tips	For under $5 at Officeworks you can buy a large desktop calendar (in a frame thing) with tear-off pages for each month. You can write in any assignments etc that are coming up, and have easy access to it everytime you sit at your desk. Plus it doubles as a doodling pad . . . just keep your doodles to the days of the month that have already been and gone.
Criteria for university offers	It depends on the ranking mix each university has set for that particular course for CSLX (mature age) students. For example, Newcastle might set a weighting of 75% WAM and 25% ATAR in order to determine your competitiveness for that particular course. You need to do two things: 1) Call Curtin and ask them whether maths and phys are "assumed knowledge" or compulsory prereqs. If they are compulsory, many times such prereqs don't apply to non-current school leavers, so ask if that is the case. Also enquire about bridging courses to make up the deficit if none of the above options work. 2) Call the other unis and ask the relevant faculty what mix of WAM/ATAR they are using to determine their entry ranking for that particular course.
Course recommendations	Apply for Cert II in Computer building or A+ is another great course which will let you know all about the hardware of a computer. Both these courses are offered at TAFE aka Polytechnic West, hope I helped.
Selection of study options	University is always worth it. For both the academic and social aspects... assuming you study on campus.
Selection of courses	The science degree may be more relevant for fabrication, processing etc which is all well and good but probably not much use in Australia.

Table 6.7: A sample of the text conversations (post content) categorized as "Answers/Providing information" behaviour by the proposed algorithm from the higher education data set. Text segments representing the Answer Person behaviour has been highlighted in grey. The column on the left present a categorization of the information (through lenses of focus areas, products and processes) presented in the posts.

6.4) Discussion of implications

This chapter introduced a novel technique for automation of online social roles found in online communities' research based on a novel application of the honeycomb social media framework, machine learning, NLP and a hierarchy of social roles.

Also, the chapter presented experimental evidence using a demonstration of the proposed techniques on a case study; higher education domain by successfully extracting domain specific social roles from a public forum dataset of over 1.2 million social media conversations. The results demonstrated that this type of social media data can be highly effective in monitoring stakeholder behavior.

This extends the contributions presented in chapters 3, 4 and 5 by empowering organizations further in their attempt to serve their stakeholders by taking a step further from understanding the *Perceived Organization* (contribution from chapter 4) and understanding the perceived organizational brand (contribution from chapter 5) to understanding 'who' and 'why' is responsible for creation and diffusion of this information that creates the *Perceived Organization* and the perceived organizational brand. This enables organizations to identify influencers in online communities who push information related to the organization into the digital environment via social media from the real/physical world.

Six social roles were introduced from the large body of research on online social roles for use in the higher education sector to monitor stakeholder opinions, emotions, interactions and behaviors on public online communities.

The deep learning technique used an LSTM and yielded high accuracy (AUC), due to extra knowledge gained from the word embedding layer. This has been shown to be a highly useful technique for higher education Institutes to monitor their stakeholders (students, parents, alumni, staff etc.) in near real-time and provide valuable insights for university management (Wijenayake et al., 2020).

This method will prove a valuable asset in varied business sectors to monitor customers, as well as in healthcare to monitor patient recovery and progress, as well as in defense sectors to monitor civilians' behavior (Wijenayake et al., 2020).

6.5) Chapter Summary

This chapter presented a methodological approach for automated detection of social roles in online communities. Furthermore, the chapter introduced online social roles from a new lens/ perspective; as a means of adding a layer of rules to big data to reduce the complexity, by enabling social media conversations to be viewed through the lenses of social roles. The study also contributes to the growing body of literature on online social roles by introducing a standardization for scattered roles from a machine learning perspective and by introducing six roles for the higher education sector to monitor stakeholder opinions and behaviors.

The automation of structural analysis based social roles and a methodology to standardize social roles from a combined social science, psychological and technical perspectives is recommended as future work (Wijenayake et al., 2020).

7. Conclusion

"Concern for man and his fate must always form the chief interest of all technical endeavours. Never forget this in the midst of your diagrams and equations."

~ Albert Einstein

This thesis presented an exploration into developing techniques for leveraging social media to enhance communication and understanding between organizations and stakeholders. This chapter concludes the thesis by presenting a synopsis of the research endeavour. A summary of research contributions is presented together with their significance followed by a discourse on how the presented achievements address each of the research questions. Finally, future research directions arising from the presented research is suggested.

7.1) Summary of Contributions of this thesis/ Research Synopsis

Enhanced communication and understanding between organizations and their stakeholders are key to creating value for stakeholders, advancing organizations to serve and satisfy their customers better on a deeper level and in turn preserve the existence of business. The motivation for this research ascended from the observation of the evolution of interactions between organizations and their stakeholders which has shifted from pure face-to face interactions to a completely digital co-existence where the digital environment presents a digital form of the organization and the individual and thereby the main rendezvous point for interactions, understanding and relationships. One of the most significant aspects of this modern digital phenomenon is social media platforms. Currently, stakeholders are expressing themselves more frequently, in-depth and in-detail through these platforms.

The literature review and initial investigations into social media platforms revealed that social media data consists of a vast repertoire of information on stakeholder opinions, emotions and experiences related to organizations and present a huge opportunity for organizations to understand their stakeholder in depth and in real-time if leveraged successfully. Also, the literature review revealed that a diverse array of contemporary technologies exists in the areas of Artificial Intelligence and Machine Learning capable of extracting value from the unstructured data held by this social media. Furthermore, the literature review and the initial investigation into contemporary organizations uncovered that while organizations have embraced modern technologies, their ability to harness the value presented in social media data to understand their stakeholders on a deeper level is still in a preliminary stage and the opportunities presented by the presence of this data and technology for organizations to understand their stakeholders on a deeper level and keep them satisfied in the modern competitive business environment; has not yet been fully investigated. Furthermore, the literature review revealed that social media platforms in the category of public online discussion forums are a type of social media heavily underutilized in research and by organizations while possessing a stream of promising data related to the above mentioned stakeholders, organizations and their interactions. Moreover, the literature review revealed that the extent to which technology and the associated social media analytics can positively impact organizations as well as society in the long term is still unknown.

The exploration into techniques that can be used to generate deeper insights from social media related to stakeholders, organizations, and their relationships, yielded several significant theoretical and empirical contributions.

This thesis presented a range of novel contributions to address the discovered research gaps, ranging from a series of theoretical and empirical research contributions presented in the following sections.

The main theoretical and empirical contributions of this research can be divided into three categories as follows. Contributions related to:

1. Leveraging social media to understand stakeholders on a deeper level based on behavioural information and individuality captured from social media data
2. Leveraging social media to understand the organization perceived by the stakeholder
3. Leveraging social media to capture knowledge and insights for organizational strategy, performance, and management.

7.1.1) Contributions related to understanding stakeholders by leveraging social media

Significant contributions of this thesis related to understanding stakeholders through social media can be summarized as follows.

- An extensive literature review was conducted on existing research, theories, models, and frameworks related to representing an individual as a digital stakeholder for organizations and a novel definition for Digital Stakeholder is presented.

- The concept of Digital stakeholder was defined and established from an organization perspective, utilizing organizational data and social media data through an exploration into the nature of human cognition, the cotemporary organization, and social media.

- A novel, generic, holistic/comprehensive, conceptual Digital Stakeholder model was constructed and proposed using existing research and models.
- A novel framework to model the proposed Digital Stakeholder's journey through the organization by using the proposed conceptual Digital Stakeholder model was constructed and proposed.
- A set of novel matrices were presented as a conceptual theoretical contribution for mapping theory from psychology to features of social media.
- A generic predictive model for monitoring of the life cycle of the Digital Stakeholder using the proposed framework was presented.
- A demonstration was offered on how the proposed generic framework together with the proposed conceptual Digital Stakeholder model and the proposed predictive model can be adopted to a specific organization domain using a case study application. The higher education domain was used as the case study and a holistic model of a university student was developed as a Digital Stakeholder Avatar and the student journey was presented using the proposed framework.

- An extensive literature review was conducted into existing research and methods of detection of online social roles played by the Digital Stakeholder in online communities from an opinion providing and information creation and diffusion perspective that impacts the perceived image and brand of organizations.

- A generic methodological approach was proposed to automate the detection of online social roles from online communities using a typology of social roles identified from theory and concepts related to online social roles found in communication and media literature, together with a novel application of the honeycomb social media framework and a combination of machine learning techniques of Natural Language Processing (NLP) deep learning and word embedding.
- Six social roles were identified using the literature review into the existing repertoire of online social roles and proposed for the case study domain; the higher education sector to monitor stakeholder opinions, emotions, interactions and behaviors on public online discussion forums.
- A demonstration of the effectiveness of the proposed method for automated detection of online social roles of the Digital Stakeholder is presented through an application on the case study domain; higher education, using a public social media dataset of over one million text conversations from public online discussion forums.

The significance of the contributions

The presented contributions address a significant gap in research literature. The lack of a comprehensive, holistic, adoptable model capable of capturing the diverse and dynamic range of traits, characteristics and preferences that make up an individual has been successfully addressed using the proposed conceptual model and framework. The proposed conceptual model is generic and can be utilized by any organization in any business domain to model and understand their varied stakeholders.

The contributions on online social roles of the Digital Stakeholder addresses another significant research gap; the lack of a methodology or framework to automate the extraction of online social roles. Theory and applications of online social roles span decades of research in media and communication literature and investigation into machine learning for identification of online social roles have been suggested but not investigated comprehensively. Hence, the presented contributions generate a significant and essential missing piece into this discipline of research. With the current boom of social media resulting in social data streams of high volume and velocity, it is no longer feasible to extract and monitor online social roles using the existing, traditional manual coding methods that have been widely used in existing research. Furthermore, this thesis introduces online social roles from a new lens/perspective; as a means of adding a layer of rules to unstructured social media data to reduce the complexity, by enabling social media conversations to be viewed through the lenses of social roles. Also, this methodology presents the first novel application of the widely accepted honeycomb social media framework to discover online social roles from social media. Moreover, the application contributions demonstrate that this type of social media data can be highly effective in monitoring stakeholder behavior in public online communities external to the organization. The contributions have been published in (Wijenayake et al., 2020).

7.1.2) Contributions related to understanding the organization by leveraging social media

Significant contributions of this thesis related to understanding perceptions of organizations held by stakeholders using social media can be summarized as follows.

- An extensive investigation was conducted into existing research, methods, and frameworks available to transform unstructured social media in text form into organizational insights related to the Perceived Organization (stakeholders' perception about an organization).

- An extensive literature review was conducted on techniques available for structuring and analysing unstructured textual social media data and existing technologies of machine learning and data analytics that have been used by business organizations to capture value from textual big data.
- A novel framework named Social Media to Organizational Insights Transfiguration Framework was proposed for capturing the Perceived Organization by using social media through a method utilizing the honeycomb social media framework, text mining techniques, Natural Language Processing (NLP), exploratory and visual analytics and feature engineering.
- A novel transformation technique was proposed to remodel raw unstructured social media text data in public online discussion forums into structured data with the extracted knowledge components related to organizational aspects embedded using digital traces of stakeholder experiences, opinions and emotions. This technique supports analytics and visalization to be performed on unstructured textual social media data from public online discussion forums.
- A novel technique called ESV was proposed to quantify and measure qualitative components discovered from textual social media using Emotion, Strength and Variety.
- A novel technique for extraction of emotions for the higher education domain from public online discussion forums was proposed.

- A suite of algorithms to compare organizations and products is presented including a novel algorithm to rank universities from social media data in public online discussion forums using a combination of the above proposed techniques and discovered processes.
- A demonstration of the effectiveness of the proposed framework and the proposed techniques is presented using a case study organization and a public social media data set of over one million textual discussions from public online discussion forums.

The significance of the contributions

The proposed framework present novel techniques that enable organizations to leverage social media and monitor the organizational perception from varied lenses of products, processes and focus areas and understand the positioning of the organization among its competitors. The contributions demonstrate developing of generic techniques to generate a layer of actionable insights aligned with organizational strategy to improve the perceived stakeholder experience. Furthermore, the contributions present a novel application of the widely accepted honeycomb social media framework to provide structure to unstructured social media text and identify aspects of organizations (products, processes, focus areas and competitor organizations) embedded in social media.

7.1.3) Contributions related to leveraging social media to capture knowledge and insights for organizational strategy, performance and management

Significant contributions of this thesis related to capturing knowledge for organizational strategy, performance and management using social media can be summarized as follows.

- An extensive investigation was conducted into existing research and methods used by organizations to model organizational brand personality (the organization's brand perceptions held by stakeholders) and their efficacy.
- A novel method was proposed to extract an organization's brand from social media by introducing a technique to map components of social media into accepted features of brand, presented in marketing literature, by extending the previously presented Social Media to Organizational Insights Transfiguration Framework and by using techniques of emotion extraction, text mining, NLP, deep learning and word embedding .
- The proposed technique presented a method of automating monitoring of stakeholders' perception of organizational brand image, brand consistency and harmony of brand architecture in near real-time, at different granularities across a spectrum of organizational constituents facilitating benchmarking against competitors.
- A demonstration of the efficacy of the proposed method is presented by using a case study organization and a case study business domain; Australian universities and the higher education industry of Australia using a public social media dataset of over one million text conversations from public online discussion forums.

Significance of the contributions

The contribution on brand personality addresses a significant research gap in marketing literature by presenting the first research study to provide an automated method to extract brand personality using machine learning and also the very first study where the possibility and opportunities of leveraging social media to extract brand personality using brand personality scales have been proposed and demonstrated. The proposed method presents a highly versatile, novel technique suitable for contemporary organizations by proposing and demonstrating alternative methods that can be utilized by any business organization from low to high use of machine processing and automation to suit the technological capacity and cultural dynamics of diverse organizations. Furthermore, the contributions demonstrate how heavily underutilized, unstructured, high volume, publicly available and continuously updated big data sources can simply and effectively be utilized to understand the ethos, pathos and logos of an organization's brand Image, and realize customer needs of the organization to serve their customers better. The contributions have been published in (Wijenayake et al., 2021).

7.2) Addressing the Research Questions

This section presents how the above summarized contributions of this research have addressed the research questions presented in the introduction chapter. Four research questions were formulated and presented in chapter 1 of the thesis. These research questions have been addressed throughout the preceding chapters of the thesis and have yielded the research contributions briefed in section 7.1 above.

1) *To what depth can knowledge and value be captured from the digital environment and social media to understand communication and interactions between organizations and individuals?*

This research question investigates the extend of depth in comprehension achievable in relation to understanding communication between organizations and individuals by using the digital environment with a special focus on social media. In order to address this research question, an investigation was carried out into existing research on capturing the Perceived

Organization through the digital environment. Consequently, challenges faced by organizations related to using big data to generate value were identified and a research study was conducted to develop and trial novel techniques, models, and frameworks to overcome the challenges which resulted in the contributions briefed in section 7.1.2 above and presented in chapter 4 of the thesis.

2) *How can a holistic and generic framework of a digital stakeholder be developed from an organizational perspective using organizational data and the digital environment?*

This research question is formulated to investigate how existing theories, models and frameworks can be utilized to generate a comprehensive and generic framework of an individual as a stakeholder from an organization's perspective by using the combination of organizational data and social media data available in the digital environment. To address the research question a novel, generic, holistic/comprehensive framework to model a Digital Stakeholder and their journey through the organizations by leveraging organizational data sources as well as social media data was developed and demonstrated as summarized in section 7.1.1 above and presented in chapter 3 of the thesis.

3) *How can an organization utilize digital stakeholder opinions present in social media to develop a model of the external viewpoint of themselves and their organizational brand, to generate insights to provide a better stakeholder experience?*

This research question is articulated to investigate how theories, concepts and models present in marketing literature on organizational brand can be utilized to generate value for organizations to improve their stakeholder experience by using social media and machine learning capabilities. To address the research question an investigation was carried out into existing methods and techniques utilized for generating brand (summarized in section 2.2). A significant research gap related to existing methods used for generating brand personalities were identified and consequently a research investigation was carried out to address the research gap by developing and trialling novel machine learning based methods to automate the extraction of brand personality which resulted in the contributions summarized in section 7.1.3 above and presented in chapter 5 of this thesis.

4) *What key roles do digital stakeholders play in online communities from an opinion providing and information creation and diffusion perspective and how can a generic framework of automation be provided to observe these online stakeholder roles?*

This research question is articulated to further investigate the Digital Stakeholder defined and modelled by addressing the research question 2, from a novel perspective of online social roles that are responsible for creation and diffusion of information related to organizations in social media communities. To address this research question, an extensive investigation was conducted to explore theories and techniques on online social roles in media and communication literature. The investigation resulted in discovering a significant research gap in media and communication literature which was addressed by developing and trialling a generic methodological approach for automated detection of online social roles briefed in section 7.1.1 above and presented in chapter 6 of this thesis.

7.3) Limitations and future directions

This section discusses any limitations faced in the current research endeavour and possible future research directions unveiled by the research and contributions presented in this thesis.

Understanding the ever changing and evolving needs and wants of individuals remains an arduous task. The novel frameworks, models and techniques presented in this thesis were trialled on the higher education domain as the case study for experiments. However, since these proposed approaches are generic and promising as evident through the presented results, there exist much potential to apply the proposed frameworks, models and techniques on other domains and empirically demonstrate their applicability, value as well as any limitations in relation to varied business sectors. These business cases have not been covered in the scope of this research thesis.

Furthermore, this research mainly used heavily underutilized social media in the category of public online discussion forums for experimentation and demonstration. There is opportunity to experiment the applicability and efficacy of the proposed frameworks, models, and techniques on other categories of textual social media such as posts/comments from groups and pages on Facebook, comments from online blogs, comments collected from Twitter exchanges etc.

Moreover, the presented research presents much opportunity and grounds to extend and expand some of the novel frameworks and models proposed and demonstrated herein.

One possible future application is the proposed *Social Media to Organizational Insights Transfiguration Framework* and it's six phases (presented in section 4.2) applied to differing organizations in other industries and demonstrating the extraction of domain specific processes, products and focus areas and the possibilities of value creation through generation of domain specific organizational insights.

There also exist possibility to extend the understanding of the expansive range of possibilities held by leveraging public online discussion forums as a category of social media platforms heavily underutilized by research and organizations. There are copious possibilities to explore this further by utilizing the results of the application of the honeycomb social media framework presented in section 4.2 of this thesis. During this research endeavour the exploration of varied public online discussion forums revealed them to be rich in emotions, opinions and experiences of individuals related to not simply higher education (which was the empirical focus of this study) but also many other domains and services such as healthcare, banking, technology, service providers, house hold brands, governing bodies etc.

This thesis defined and established the concept of a Digital Stakeholder which opens doors to future research and development of theories and models by application, adoption and diversification of the work presented herein.

The social media empowered *conceptual holistic Digital Stakeholder* model presented in section 3.2 can be extended further with new research, as new research techniques capable of extracting additional specific individual traits through novel applications of social media are unveiled. With the current boom and significance of social media, more research in this research area are likely to follow opening up opportunities for further expansion and adaption of this model to provide a more comprehensive model of an individual as a Digital Stakeholder.

Another possible future opportunity presented is the potential to apply the proposed framework of the *Digital Stakeholder's journey through the organization* presented in section 3.3 together with the *conceptual holistic Digital Stakeholder* model, on diverse organisations in dissimilar industries with differing availability of organizational and public data sources. Demonstrations on implications of application of the model and the framework on other categories of stakeholders (the thesis provided a demonstration on modelling a student in a higher education organization) such as a patient in a healthcare organization, a customer in a bank etc. would generate valuable empirical outcomes.

The proposed methodology to generate *Brand Personality* of an organization from social media presented in section 5.6 of this thesis could be applied and demonstrated on other business sectors such as banking and healthcare considering the availability of domain specific brand personality scales in marketing literature and the availability of a vast amount conversations in public online discussion forums related to these domains.

Another possible avenue for future research is presented by the technique for automation of online social roles presented in section 6.3 of this thesis. This method will prove valuable in varied business sectors to monitor online behavior of customers as well as in healthcare to monitor patient recovery and progress as well as in defense sectors to monitor civilians' behavior if the value can be empirically established through novel applications. Moreover, it was observed that online social roles are very ambiguous, context and platform (social media) dependent, not mutually exclusive and explained differently across literature (for example *gate keeper* and *mentor*). Hence, a clear requirement was observed for providing a universal standardization for online social roles.

7.4) Final words

This research recommends a novel approach for contemporary organizations to improve their agility and continuously serve and satisfy their stakeholders through a union of human, machine, traditional and social media competencies.

The research suggested novel methods of merging human, computer and Artificial Intelligence capabilities as well as technique to be used hand-in-hand with traditional methods for maximum impact; in order for business organizations to understand their stakeholders' needs , wants and deeply held perceptions about organizations and their products and services. Technical experts working together with the domain experts for informed decision-making and for enhanced, better performing, better serving organizational processes was proposed and envisioned. We propose this direction as the new paradigm for business organizations to approach the dynamic demands of organizational agility enforced with the imminent new era of Artificial Intelligence and digital transformation.

References

Aaker, D. A. (2012). *Building strong brands*. Simon and Schuster.

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research*, *34*(3), 347–356. https://doi.org/10.2307/3151897

Adikari, A., De Silva, D., Ranasinghe, W. K. B., Bandaragoda, T., Alahakoon, O., Persad, R., Lawrentschuk, N., Alahakoon, D., & Bolton, D. (2020). Can online support groups address psychological morbidity of cancer patients? An artificial intelligence based investigation of prostate cancer trajectories. *PLoS ONE*, *15*(3), 1–14. https://doi.org/10.1371/journal.pone.0229361

Alahakoon, D., & Wijenayake, P. (2018). Leveraging big data for organizational performance management and control. In *The Routledge Companion to Performance Management and Control*. https://doi.org/10.4324/9781315691374

Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, *26*(6), 1289–1295. https://doi.org/10.1016/j.chb.2010.03.018

Ammu, N., & Irfanuddin, M. (2013). Big data challenges. *International Journal of Advanced Trends in Computer Science and Engineering*, *2*(1), 613–615.

Asbjørn Følstad, & Knut Kvale. (2018). Customer journeys: a systematic literature review Asbjørn Følstad, Knut Kvale Customer Journeys: A Systematic Literature Review. *Journal of Service Theory and Practice*. https://doi.org/10.1108/JSTP-11-2014-0261

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of Facebook usage. *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'12*, *volume*, 24–32. https://doi.org/10.1145/2380718.2380722

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*(3), 372–374.

Bai, S., Yuan, S., Hao, B., & Zhu, T. (2014). Predicting personality traits of microblog users. *Web Intelligence and Agent Systems*, *12*(3), 249–265. https://doi.org/10.3233/WIA-140295

Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*, *43*(3), 465–482. https://doi.org/10.1007/s11186-014-9216-5

Balbi, S., Misuraca, M., & Scepi, G. (2018). Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management*, *54*(4), 674–685.

Bandaragoda, T. R., De Silva, D., & Alahakoon, D. (2017). Automatic event detection in microblogs using incremental machine learning. *Journal of the Association for Information Science and Technology*, *68*(10), 2394–2411.

Bandaragoda, T., Ranasinghe, W., Adikari, A., de Silva, D., Lawrentschuk, N., Alahakoon, D., Persad, R., & Bolton, D. (2018). The Patient-Reported Information Multidimensional Exploration (PRIME) framework for investigating emotions and other factors of prostate cancer patients with low intermediate risk based on online cancer support group discussions. *Annals of Surgical Oncology*, 1–9.

Baptista, J., Newell, S., & Currie, W. (2010). Paradoxical effects of institutionalisation on the strategic awareness of technology in organisations. *The Journal of Strategic Information Systems*, *19*(3), 171–183.

Baptista, J., Wilson, A. D., Galliers, R. D., & Bynghall, S. (2017). Social media and the emergence of reflexiveness as a new capability for open strategy. *Long Range Planning*, *50*(3), 322–336.

Baym, N. K. (2012). The Emergence of On-Line Community. *Cybersociety 2.0: Revisiting Computer-Mediated Communication and Community*, *May 2019*, 35–68. https://doi.org/10.4135/9781452243689.n2

Benamar, L., Balagué, C., & Ghassany, M. (2017). The Identification and Influence of Social Roles in a Social Media Product Community. *Journal of Computer-Mediated Communication*, *22*(6), 337–362. https://doi.org/10.1111/jcc4.12195

Benthaus, J., Risius, M., & Beck, R. (2016). Social media management strategies for organizational impression management and their effect on public perception. *Journal of Strategic Information Systems*, *25*(2), 127–139. https://doi.org/10.1016/j.jsis.2015.12.001

Bhattacharya, P., Phan, T. Q., Bai, X., & Airoldi, E. (2015). Investigating the impact of network effects on content generation: Evidence from a large online student network. *ArXiv Preprint ArXiv:1512.06469*.

Bolat, E., & O'Sullivan, H. (2017). Radicalising the marketing of higher education: learning from student-generated social media data. *Journal of Marketing Management*, *33*(9–

10), 742–763. https://doi.org/10.1080/0267257X.2017.1328458

Bose, I., & Mahapatra, R. K. (2001). Business data mining - A machine learning perspective. *Information and Management*, *39*(3), 211–225. https://doi.org/10.1016/S0378-7206(01)00091-X

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159.

Brech, F. M., Messer, U., Vander Schee, B. A., Rauschnabel, P. A., & Ivens, B. S. (2017). Engaging fans and the community in social media: Interaction with institutions of higher education on Facebook. *Journal of Marketing for Higher Education*, *27*(1), 112–130.

Breuel, T. M. (2017). High performance text recognition using a hybrid convolutional-lstm implementation. *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference On*, *1*, 11–16.

Brown, T. J., & Dacin, P. A. (1997). The company and the product: Corporate associations and consumer product responses. *Journal of Marketing*, *61*(1), 68–84. https://doi.org/10.2307/1252190

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97–102. https://doi.org/10.1016/j.paid.2014.01.016

Callarisa, L., García, J. S., Cardiff, J., & Roshchina, A. (2012). Harnessing social media platforms to measure customer-based hotel brand equity. *Tourism Management*

Perspectives, *4*, 73–79. https://doi.org/10.1016/j.tmp.2012.04.005

Cantador, I., Fernández-Tob\'\ias, I., & Bellog\'\in, A. (2013). Relating personality types with user preferences in multiple entertainment domains. *CEUR Workshop Proceedings*.

Carpentier, M., Van Hoye, G., & Weijters, B. (2019). Attracting applicants through the organization's social media page: Signaling employer brand personality. *Journal of Vocational Behavior*, *115*, 103326.

Celiktutan, O., & Gunes, H. (2017). Automatic Prediction of Impressions in Time and across Varying Context: Personality, Attractiveness and Likeability. *IEEE Transactions on Affective Computing*, *8*(1), 29–42. https://doi.org/10.1109/TAFFC.2015.2513401

Celli, F., & Rossi, L. (2012). The role of emotional stability in Twitter conversations. *Proceedings of the Workshop on Semantic Analysis in Social Media*, 10–17.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 1165–1188.

Choi, J., Yoon, J., Chung, J., Coh, B. Y., & Lee, J. M. (2020). Social media analytics and business intelligence research: A systematic review. *Information Processing and Management*, *57*(6), 102279. https://doi.org/10.1016/j.ipm.2020.102279

Chowdhary, K. R. (2020). Natural language processing. In *Fundamentals of Artificial Intelligence* (pp. 603–649). Springer.

Chung, S., & Park, J. (2017). The influence of brand personality and relative brand identification on brand loyalty in the European mobile phone market. *Canadian Journal of Administrative Sciences*, *34*(1), 47–62. https://doi.org/10.1002/cjas.1321

Clifton, R., & Simmons, J. (2003). Brands and branding (pp. 13--14). *London: The Economist* *and Profile Books. Google Scholar*.

Costa Jr, P. T., & McCrae, R. R. (2008). *The Revised NEO Personality Inventory (NEO-PI-R).* Sage Publications, Inc.

Davenport, T. (2014). *Big data at work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press.

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. *International Institute for Analytics*, *3*, 1–31.

De Silva, D., Ranasinghe, W., Bandaragoda, T., Adikari, A., Mills, N., Iddamalgoda, L., Alahakoon, D., Lawrentschuk, N., Persad, R., Osipov, E., Gray, R., & Bolton, D. (2018). Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. *PLoS ONE*, *13*(10), 1–19. https://doi.org/10.1371/journal.pone.0205855

de Valck, K., van Bruggen, G. H., & Wierenga, B. (2009). Virtual communities: A marketing perspective. *Decision Support Systems*, *47*(3), 185–203. https://doi.org/10.1016/j.dss.2009.02.008

Dhar, S., & Mazumdar, S. (2014). Challenges and best practices for enterprise adoption of big data technologies. *Technology Management Conference (ITMC), 2014 IEEE International*, 1–4.

Dowling, G. R. (1986). Managing your corporate images. *Industrial Marketing Management*, *15*(2), 109–115.

Drucker, P. F. (1974). Management: Tasks, responsibilities, practices. In *Management: Tasks, Responsibilities, Practices*. https://doi.org/10.4324/9780080939063

Duesterhaus, A., & Duesterhaus, M. (2014). Attributes of successful university brands in the USA. *Journal of Brand Strategy*, *3*(2), 169–183.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Evans, J. A., & Aceves, P. (2016). Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology*, *42*(1), 21–50. https://doi.org/10.1146/annurev-soc-081715-074206

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., Davalos, S., Moens, M. F., & De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, *26*(2–3), 109–142. https://doi.org/10.1007/s11257-016-9171-0

Farnadi, G., Zoghbi, S., Moens, M.-F., & De Cock, M. (2013). Recognising personality traits using facebook status updates. *Proceedings of the Workshop on Computational Personality Recognition (WCPR13) at the 7th International AAAI Conference on Weblogs and Social Media (ICWSM13)*, 14–18.

Ferwerda, B., Schedl, M., & Tkalcic, M. (2016). Personality traits and the relationship with (non-) disclosure behavior on facebook. *Proceedings of the 25th International Conference Companion on World Wide Web*, 565–568.

Fisher, D., Smith, M., & Welser, H. T. (2006). You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, *3*, 59b-59b. https://doi.org/10.1109/HICSS.2006.536

Freeman, R. E. (2011). Managing for Stakeholders. *SSRN Electronic Journal*, 1–22. https://doi.org/10.2139/ssrn.1186402

Friedman, A. L., & Miles, S. (2002). Developing stakeholder theory. *Journal of Management Studies*, *39*(1), 1–21. https://doi.org/10.1111/1467-6486.00280

Füller, J., Hutter, K., Hautz, J., & Matzler, K. (2014). User roles and contributions in innovation-contest communities. *Journal of Management Information Systems*, *31*(1), 273–308. https://doi.org/10.2753/MIS0742-1222310111

Garanti, Z., & Kissi, P. S. (2019). The effects of social media brand personality on brand loyalty in the Latvian banking industry. *International Journal of Bank Marketing*.

Garcia, D., & Sikström, S. (2014). The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*, *67*, 69–74. https://doi.org/10.1016/j.paid.2013.10.001

Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, *20*(1), 112–141. https://doi.org/10.1080/10864415.2016.1061792

George, G., Haas, M. R., & Pentland, A. (2014). *Big data and management*. Academy of Management Briarcliff Manor, NY.

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. University of Chicago Press.

Golbeck, J. A. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research*, *2*(1), 2.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 253–262).

Goodboy, A. K., & Martin, M. M. (2015). The personality profile of a cyberbully: Examining the Dark Triad. *Computers in Human Behavior*, *49*, 1–4. https://doi.org/10.1016/j.chb.2015.02.052

Gordon, R., Zainuddin, N., & Magee, C. (2016). Unlocking the potential of branding in social marketing services: utilising brand personality and brand personality appeal. *Journal of Services Marketing*, *30*(1), 48–62. https://doi.org/10.1108/JSM-02-2015-0105

Graves, A., Mohamed, A., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *Acoustics, Speech and Signal Processing (Icassp), 2013 Ieee International Conference On*, 6645–6649.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, *17*(4), 397–431.

Hall, R. (1993). A framework linking intangible resources and capabiliites to sustainable competitive advantage. *Strategic Management Journal*, *14*(8), 607–618.

Hankinson, G. (2012). The measurement of brand orientation, its performance impact, and the role of leadership in the context of destination branding: An exploratory study. *Journal of Marketing Management*, *28*(7–8), 974–999. https://doi.org/10.1080/0267257X.2011.565727

Hill, C. W., & Jones, T. M. (1992). Stakeholder-Agency Theory Charles. *Journal of Management Studies*, *29*(2), 131–154.

Himelboim, I., Gleave, E., & Smith, M. (2009). Discussion catalysts in online political discussions: Content importers and conversation starters. *Journal of Computer-Mediated Communication*, *14*(4), 771–789. https://doi.org/10.1111/j.1083-6101.2009.01470.x

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hu, Y., Xu, A., Hong, Y., Gal, D., Sinha, V., & Akkiraju, R. (2019). Generating Business Intelligence Through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content. *Journal of Management Information Systems*, *36*(3), 893–930. https://doi.org/10.1080/07421222.2019.1628908

Huang, J., Baptista, J., & Galliers, R. D. (2013). Reconceptualizing rhetorical practices in organizations: The impact of social media on internal communications. *Information & Management*, *50*(2–3), 112–124.

Huang, J., Baptista, J., & Newell, S. (2015). Communicational ambidexterity as a new capability to manage social media communication within organizations. *The Journal of Strategic Information Systems*, *24*(2), 49–64.

Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. (2016). The influence of social media interactions on consumer-brand relationships: A three-country study of brand perceptions and marketing behaviors. *International Journal of Research in Marketing*, *33*(1), 27–41. https://doi.org/10.1016/j.ijresmar.2015.06.004

Hultman, M., Skarmeas, D., Oghazi, P., & Beheshti, H. M. (2015). Achieving tourist loyalty through destination personality, satisfaction, and identification. *Journal of Business*

Research, *68*(11), 2227–2231. https://doi.org/10.1016/j.jbusres.2015.06.002

Iacobelli, F., & Culotta, A. (2013). Too neurotic, not too friendly: structured personality classification on textual data. *Seventh International AAAI Conference on Weblogs and Social Media*.

International Telecommunication Union. (2019). Measuring digital development Facts and figures 2019. In *ITUPublications*. https://www.itu.int/en/mediacentre/Documents/MediaRelations/ITU Facts and Figures 2019 - Embargoed 5 November 1200 CET.pdf

Internet live stats. (n.d.). Retrieved May 16, 2020, from https://www.internetlivestats.com/

Jahnke, I. (2010). Dynamics of social roles in a knowledge management community. *Computers in Human Behavior*, *26*(4), 533–546. https://doi.org/10.1016/j.chb.2009.08.010

John, O. P., Srivastava, S., & others. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, *2*(1999), 102–138.

Johnson, S. L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research*, *26*(1), 165–187. https://doi.org/10.1287/isre.2014.0562

Jordan, M. I., & Mitchell, T. M. (2015). *Machine learning: Trends, perspectives, and prospects*. *349*(6245).

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. *2013 Sixth International Conference on Contemporary Computing (IC3)*, 404–409.

Kemp, S. (2020). *DIGITAL 2020: GLOBAL DIGITAL OVERVIEW*. https://datareportal.com/reports/digital-2020-global-digital-overview

Khanna, M., Jacob, I., & Yadav, N. (2014). Identifying and analyzing touchpoints for building a higher education brand. *Journal of Marketing for Higher Education*, *24*(1), 122–143. https://doi.org/10.1080/08841241.2014.920460

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. https://doi.org/10.1016/j.bushor.2011.01.005

Kietzmann, J. H., Silvestre, B. S., McCarthy, I. P., & Pitt, L. F. (2012). Unpacking the social media phenomenon: towards a research agenda. *Journal of Public Affairs*, *12*(2), 109–119.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kumar, A., & Sangwan, S. R. (2019). Rumor detection using machine learning techniques on social media. In *Lecture Notes in Networks and Systems* (Vol. 56). Springer Singapore. https://doi.org/10.1007/978-981-13-2354-6_23

Kunte, A., & Panicker, S. (2020). Personality Prediction of Social Network Users Using Ensemble and XGBoost. In *Progress in Computing, Analytics and Networking* (pp. 133–140). Springer.

Lambiotte, R., & Kosinski, M. (2014). Tracking the digital footprints of personality.

Proceedings of the IEEE, *102*(12), 1934–1939.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *International Conference on Machine Learning*, 1188–1196.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

Lee, D., & Liu, H. (2015). LIKE and recommendation in social media. *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*, 1539–1540. https://doi.org/10.1145/2740908.2741981

Li, J., Ritter, A., & Hovy, E. (2014). Weakly supervised user profile extraction from twitter. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 165–174.

Li, Z., Fan, Y., Jiang, B., Lei, T., & Liu, W. (2019). A survey on sentiment analysis and opinion mining for social multimedia. *Multimedia Tools and Applications*, *78*(6), 6939–6967. https://doi.org/10.1007/s11042-018-6445-z

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Liu, P., Tov, W., Kosinski, M., Stillwell, D. J., & Qiu, L. (2015). Do Facebook Status Updates Reflect Subjective Well-Being? *Cyberpsychology, Behavior, and Social Networking*, *18*(7), 373–379. https://doi.org/10.1089/cyber.2015.0022

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26.

Lozano, J. H., Gordillo, F., & Pérez, M. A. (2014). Impulsivity, intelligence, and academic performance: Testing the interaction hypothesis. *Personality and Individual Differences*, *61–62*, 63–68. https://doi.org/10.1016/j.paid.2014.01.013

Lu, R.-S., Tsao, H.-Y., Lin, H.-C. K., Ma, Y.-C., & Chuang, C.-T. (2019). Sentiment Analysis of Brand Personality Positioning Through Text Mining. *Journal of Information Technology Research (JITR)*, *12*(3), 93–103.

Marouf, A. A., Hasan, M. K., & Mahmud, H. (2020). Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media. *IEEE Transactions on Computational Social Systems*, *7*(3), 587–599. https://doi.org/10.1109/TCSS.2020.2966910

Martineau, P. (1958). *The personality of the retail store*.

Matzler, K., Strobl, A., Stokburger-Sauer, N., Bobovnicky, A., & Bauer, F. (2016). Brand personality and culture: The role of cultural differences on the impact of brand personality perceptions on tourists' visit intentions. *Tourism Management*, *52*, 507–520. https://doi.org/10.1016/j.tourman.2015.07.017

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

Mcalexander, J. H., Schouten, J. W., & Koenig, H. F. (2002). Marketplace Communities A Broader View of Brand Community. *Journal of Marketing*, *66*(January), 38–54.

Meyer, C., Schwager, A., & others. (2007). Understanding customer experience. *Harvard Business Review*, *85*(2), 116.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, *22*(4), 853–886. https://doi.org/10.5465/AMR.1997.9711022105

Mithas, S., Tafti, A., & Mitchell, W. (2013). How a firm’s competitive environment and digital strategic posture influence digital business strategy. *MIS Quarterly*, 511–536.

Mooi, E., & Sarstedt, M. (n.d.). *A Concise Guide to Market Research The Process, Data, and Methods Using IBM SPSS Statistics. 2011*. Berlin: Springer.

Morgan, B. (2019). *100 Stats On Digital Transformation And Customer Experience*. https://www.forbes.com/sites/blakemorgan/2019/12/16/100-stats-on-digital-transformation-and-customer-experience/

Naidoo, V., & Hollebeek, L. D. (2016). Higher education brand alliances: Investigating consumers’ dual-degree purchase intentions. *Journal of Business Research*, *69*(8), 3113–3121. https://doi.org/10.1016/j.jbusres.2016.01.027

Naili, M., Chaibi, A. H., & Ben Ghezala, H. H. (2017). Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, *112*, 340–349. https://doi.org/10.1016/j.procs.2017.08.009

Newell, S., Marabelli, M., Martineau, P., Graves, A., Mohamed, A., Hinton, G., Saboo, A. R., Kumar, V., Ramani, G., Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F. E., Duesterhaus, A., Duesterhaus, M., Callarisa, L., García, J. S., … Giaglis, G. M. (2012). Unpacking the social media phenomenon: towards a research agenda. *Journal of Public Affairs*, *12*(2), 109–119. https://www.sciencedirect.com/science/article/pii/S0167923616301671

Nicholas, C. K., & Lee, A. S. (2017). *Voice of Customers : Text Analysis of Hotel Customer Reviews (Cleanliness , Overall Environment & Value for*. 104–111.

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of Computer and System Sciences*, *80*(1), 57–71.

Ouyang, F., & Chang, Y. H. (2019). The relationships between social participatory roles and cognitive engagement levels in online discussions. *British Journal of Educational Technology*, *50*(3), 1396–1414. https://doi.org/10.1111/bjet.12647

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.*, *57*, 401–421.

Patas, J. (2015). Developing Individual IT-Enabled Capabilities for Management Control Systems. In *Business Intelligence for New-Generation Managers* (pp. 51–66). Springer.

Pawlewicz, K. (n.d.). *A history of customer service*. Retrieved May 12, 2020, from https://blog.olark.com/the-history-of-customer-service

Plutchik, R. (2001). *The Nature of Emotions*. *89*(4), 344–350.

Pournarakis, D. E., Sotiropoulos, D. N., & Giaglis, G. M. (2017). A computational model for mining consumer perceptions in social media. *Decision Support Systems*, *93*, 98–110. https://doi.org/10.1016/J.DSS.2016.09.018

Preoţiuc-Pietro, D., Volkeva, S., Lampos, V., Bachrach, Y., & Aletras, N. (2015). Studying user income through language, behaviour and affect in social media. *PLoS ONE*, *10*(9), 1–17. https://doi.org/10.1371/journal.pone.0138717

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. *Proceedings of the Fifth ACM Conference on Recommender Systems*, 157–164.

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 180–185.

Rauschnabel, P. A., Krey, N., Babin, B. J., & Ivens, B. S. (2016). Brand management in higher education: The University Brand Personality Scale. *Journal of Business Research*, *69*(8), 3077–3086. https://doi.org/10.1016/j.jbusres.2016.01.023

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Ridings, C. M., & Gefen, D. (2004). Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, *10*(1), JCMC10110.

Riemer, K., Finke, J., & Hovorka, D. (2015). *Bridging or bonding: do individuals gain social capital from participation in enterprise social networks?*

Robertson, J., Ferguson, S. L., Eriksson, T., & Näppä, A. (2019). The brand personality dimensions of business-to-business firms: a content analysis of employer reviews on social media. *Journal of Business-to-Business Marketing*, *26*(2), 109–124. https://doi.org/10.1080/1051712X.2019.1603354

Rose, S., Hair, N., & Clark, M. (2011). Online customer experience: A review of the business-to-consumer online purchase context. *International Journal of Management Reviews*, *13*(1), 24–39.

Russom, P. (2011). Big Data Analytics - TDWI Best Practices Report Introduction to Big Data Analytics. *TDWI Best Practices Report, Fourth Quarter*, *19*(4), 1–34.

https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf

Russom, P., & others. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, *19*(4), 1–34.

Rutter, R., Roper, S., & Lettice, F. (2016). Social media interaction, the university brand and recruitment performance. *Journal of Business Research*, *69*(8), 3096–3104. https://doi.org/10.1016/j.jbusres.2016.01.025

Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior*, *27*(5), 1658–1664. https://doi.org/10.1016/j.chb.2011.02.004

Saboo, A. R., Kumar, V., & Ramani, G. (2016). Evaluating the impact of social media activities on human brand sales. *International Journal of Research in Marketing*, *33*(3), 524–541. https://doi.org/10.1016/j.ijresmar.2015.02.007

Sak, H., & Senior, A. W. (2017). *Processing acoustic sequences using long short-term memory (LSTM) neural networks that include recurrent projection layers*. Google Patents.

Schaeffer, S. E., & Rodriguez Sanchez, S. V. (2020). Forecasting client retention — A machine-learning approach. *Journal of Retailing and Consumer Services*, *52*(March 2019), 101918. https://doi.org/10.1016/j.jretconser.2019.101918

Schläfke, M., Silvi, R., & Möller, K. (2013). A framework for business analytics in performance management. *International Journal of Productivity and Performance Management*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & others. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS One*, *8*(9), e73791.

Schwartz, H. A., Sap, M., Kern, M. L., Eichstaedt, J. C., Kapelner, A., Agrawal, M., Blanco, E., Dziurzynski, L., Park, G., Stillwell, D., & others. (2016). Predicting individual well-being through the language of social media. *Biocomputing 2016: Proceedings of the Pacific Symposium*, 516–527.

Sewwandi, D., Perera, K., Sandaruwan, S., Lakchani, O., Nugaliyadde, A., & Thelijjagoda, S. (2017). Linguistic features based personality recognition using social media data. *2017 6th National Conference on Technology and Management (NCTM)*, 63–68.

Shah, D., Rust, R. T., Parasuraman, A., Staelin, R., & Day, G. S. (2006). The path to customer centricity. *Journal of Service Research*, *9*(2), 113–124. https://doi.org/10.1177/1094670506294666

Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2298–2304.

Skansi, S. (2018). *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Springer.

So, K. K. F., King, C., Hudson, S., & Meng, F. (2017). The missing link in building customer brand identification: The role of brand attractiveness. *Tourism Management*, *59*, 640–651. https://doi.org/10.1016/j.tourman.2016.09.013

Soong, H. C., Jalil, N. B. A., Kumar Ayyasamy, R., & Akbar, R. (2019). The essential of sentiment analysis and opinion mining in social media : Introduction and survey of the recent approaches and techniques. *ISCAIE 2019 - 2019 IEEE Symposium on Computer Applications and Industrial Electronics*, 272–277. https://doi.org/10.1109/ISCAIE.2019.8743799

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, *36*, 10–25. https://doi.org/10.1016/j.inffus.2016.10.004

Sung, Y., Choi, S. M., Ahn, H., & Song, Y. A. (2015). Dimensions of Luxury Brand Personality: Scale Development and Validation. *Psychology and Marketing*, *32*(1), 121–132. https://doi.org/10.1002/mar.20767

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the Association for Information Science and Technology*, *61*(12), 2544–2558.

Turner, T., Smith, M., Fisher, F., & Welser, H. (n.d.). T.(2005)," Picturing Usenet: mapping computermediated collective action. *Journal of Computer-Mediated Communication*, *10*(4), 0.

Unnisa, M., Ameen, A., & Raziuddin, S. (2016). Opinion Mining on Twitter Data using Unsupervised Learning Technique. *International Journal of Computer Applications*, *148*(12), 12–19. https://doi.org/10.5120/ijca2016911317

Urde, M. (2016). The brand core and its management over time. *Journal of Product & Brand Management*, *25*(1), 26–42.

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC*, *14*, 1–14.

Vora, H., Bhamare, M., & Kumar, D. K. A. (2020). Personality Prediction from Social Media Text: An Overview. *Int. J. Eng. Res*, *9*(05), 352–357.

Warshaw, J., Matthews, T., Whittaker, S., Kau, C., Bengualid, M., & Smith, B. A. (2015). Can an Algorithm Know the "Real You"? Understanding People's Reactions to Hyper-Personal Analytics Systems. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 797–806. https://doi.org/10.1145/2702123.2702274

Welser, H. T., Gleave, E., & Fisher, D. (2007). Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure*, *8*(2), 1–32.

Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, *26*(1–2), 1–12.

Wijenayake, P., Alahakoon, D., De Silva, D., & Kirigeeganage, S. (2021). Deep LSTM for Generating Brand Personalities Using Social Media: A Case Study from Higher Education Institutions. *International Journal of Computer and Communication Engineering*, *10*(1), 17–27. https://doi.org/10.17706/IJCCE.2021.10.1.17-27

Wijenayake, P., De Silva, D., Alahakoon, D., & Kirigeeganage, S. (2020). Automated detection of social roles in online communities using deep learning. *ACM International Conference Proceeding Series*, 63–68. https://doi.org/10.1145/3378936.3378973

Wilkins, S., Butt, M. M., & Heffernan, T. (2017). International brand alliances and co-branding: antecedents of cognitive dissonance and student satisfaction with cobranded higher education programs. *Journal of Marketing for Higher Education*, 1–19.

Xiao, S., Wei, C.-P., & Dong, M. (2016). Crowd intelligence: Analyzing online product reviews for preference measurement. *Information & Management*, *53*(2), 169–182.

Xu, A., Liu, H., Gou, L., Akkiraju, R., Mahmud, J., Sinha, V., Hu, Y., & Qiao, M. (2016). Predicting Perceived Brand Personality with Social Media. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*, *Icwsm*, 436–445.

Xu, J., Marshall, B., Kaza, S., & Chen, H. (2004). Analyzing and Visualizing Criminal Network Dynamics: A Case Study. In H. Chen, R. Moore, D. D. Zeng, & J. Leavitt (Eds.), *Intelligence and Security Informatics* (pp. 359–377). Springer Berlin Heidelberg.

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). *Comparative Study of CNN and RNN for Natural Language Processing*. http://arxiv.org/abs/1702.01923

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

Zenker, S., Braun, E., & Petersen, S. (2017). Branding the destination versus the place: The effects of brand complexity and identification for residents and visitors. *Tourism Management*, *58*, 15–27. https://doi.org/10.1016/j.tourman.2016.10.008