SENTIMENT ANALYSIS AND INFORMATION DIFFUSION IN SOCIAL MEDIA: A STUDY ON MALAYSIA'S UNIVERSITY

BY

TAN SZE MEI

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2020

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title:

SENTIMENT ANALYSIS AND INFORMATION DIFFUSION IN SOCIAL MEDIA: A STUDY ON MALAYSIA'S UNIVERSITY

Academic Session: **JAN 2020**

I **TAN SZE MEI**

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:
BLOCK 11-14-09,
LTG MACALLUM 2,
10300 GEORGE TOWN, PENANG.

Dr. Pradeep a/l Isawasan
Supervisor's name

Date: 24/4/2020

Date: 24/4/2020

SENTIMENT ANALYSIS AND INFORMATION DIFFUSION IN SOCIAL MEDIA: A STUDY ON MALAYSIA'S UNIVERSITY

BY

TAN SZE MEI

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2020

DECLARATION OF ORIGINALITY

I declare that this report entitled "**SENTIMENT ANALYSIS AND INFORMATION DIFFUSION IN SOCIAL MEDIA: A STUDY ON MALAYSIA'S UNIVERSITY**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : TAN SZE MEI

Date : 24/4/2020

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisor, Dr. Pradeep who has given me this bright opportunity to engage in this data analytics project. It is my first step to explore the data analytical field. A million thanks to you.

To a very special person in my life - my guardian, for her patience, unconditional support and love, and for standing by my side during hard times. Finally, I must say thanks to my friends for their love, support and continuous encouragement throughout the course.

ABSTRACT

As for now, social media is being part of an individual life. Even academic institution tends to shift their marketing channel to social media. Social media usage is increasing tremendously and there provide opportunities to investigate how information in social media diffuse among the users as online networking sites can mirror the structure of disconnected human culture. As social media motivates two-way communication, a snippet of data can be traded or diffused between people in interpersonal organizations. From this dissemination procedure, bunches of inactive data can be mined. It can be utilized to study the relationship between information diffusion process and sentiment of posts. Consequently, data is collected from Facebook higher educational institution page in Malaysia and being analysed. The motivation behind this paper is to help on branding higher education in Malaysia.

A conceptual framework which is inspired by the research of de Vries et al. (2012) is proposed to investigate the correlation between 6 influential factors and the number of likes, comments and shares. Among the 6 factors, there is 2 factors which are newly added in this research: sentiment of the post and use of hashtag. The result shown that use of hashtag is positive correlated with information diffusion while positive posts have shown a weaker positive relationship with information diffusion. The negative posts have shown a weak negative relationship with information diffusion. Among the 6 factors, entertaining factor has the highest coefficient for the number of likes and comments while information content and use of hashtag have the highest coefficient for number of shares.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

CHAPTER 1 PROJECT BACKGROUND

In this chapter, the, problem statement, motivation, project scope, research question, project objectives, proposed approach, contributions and background of the research is presented and discussed.

1.1 Problem Statement

A brand and a perception that can be believed and accepted that will separate it from others is needed by every academic institution (Parameswaran and Glowacka, 1995; Santovec, 2007). However, **Malaysia's tertiary education does not be considered with regards to business and financial principle by various scholastics** (Pringle & Huisman, 2011). Intrinsically, tertiary education has transformed from being viewed as a 'open' good to being viewed as a 'private' conventional (Dill, 2003; Huisman & Currie, 2004; Jongbloed, 2003; Naidoo, Shankar &Veer, 2011; Pringle & Huisman, 2011; Pringle & Naidoo, 2016).

An effective brand is about significantly more than making a discrete physical existence in the market. It is stated that the brand must meet customers' mental needs through the qualities which they come to acknowledge the brand exemplifies (Temple, 2006). However, **less attention is directed at branding higher education institution in Malaysia**. Investigating the effect and power of internet-based application as a canal for marking colleges is yet ailing in Malaysia.

1.2 Motivation

The main motivation of this research is to handle the specialty market of university. Besides, it can benefit policy makers and administrators of the university to manage the university branding in the context of student expectations which contribute to university ranking. Furthermore, this can encourage the university to stand out their brand conveying their brand message effectively and soon building customer trust. These are critical because a brand is equivalent to a promise that an institution must deliver (Nandan, 2005).

1.3 Project Scope

The scope of the research will be delivering one model to asses Malaysian perception to university brand and analysis and visualization of collected data. The data sources will be from Facebook. The data period is from 19th January 2020 to 19th February 2020.

The university located in Malaysia is chosen only as the research focus on higher education in Malaysia.

In the analysis part, sentiment analysis will be carried out to understand Malaysian's reaction to their higher education-related post in social media. The posts which are in the languages other than English such as Malay and Chinese is eliminated. The conceptual model proposed included sentiment as one of the influential factor and the use of hashtag to explore their relationship with information diffusion. The relationship between them is then tested with Pearson Correlation analysis.

1.4 Research Question

The following research questions are considered:

RQ1 : What are the keywords which related to the higher education in Malaysia that is important to branding?

RQ2 : How Malaysian react and their emotions to Malaysia's higher education-related post in social media?

RQ3 : How the information related to higher education-related spread among social media users in Malaysia?

1.5 Project Objectives

The research is aimed to **identify the keyword that Malaysian use to search for higher education in Malaysia**. The list of keywords from the findings can help on the branding strategy. It provides meaningful analytics that synthesize an accurate description of keywords regarding higher education in Malaysia. The research is to **identify the distribution of the Malaysian's sentiment or emotions based on the keywords of higher education.** It also aims **to develop an information diffusion model for Malaysia's university branding.**

Therefore, in summary the research objectives are listed as following:

- To identify the keyword that Malaysian use to search for Malaysia's higher education-related tweets
- To identify the distribution of the Malaysian's sentiment or emotions to the higher education-related tweets
- To develop an information diffusion model for Malaysia's university branding

1.6 Proposed Approach

The following approach shown in ***Figure 1.1*** is projected according to Cross-Industry Standard Process for Data Mining (CRISP-DM) and Social Media Analytics Framework which will be further discuss in ***Section 4.4***.

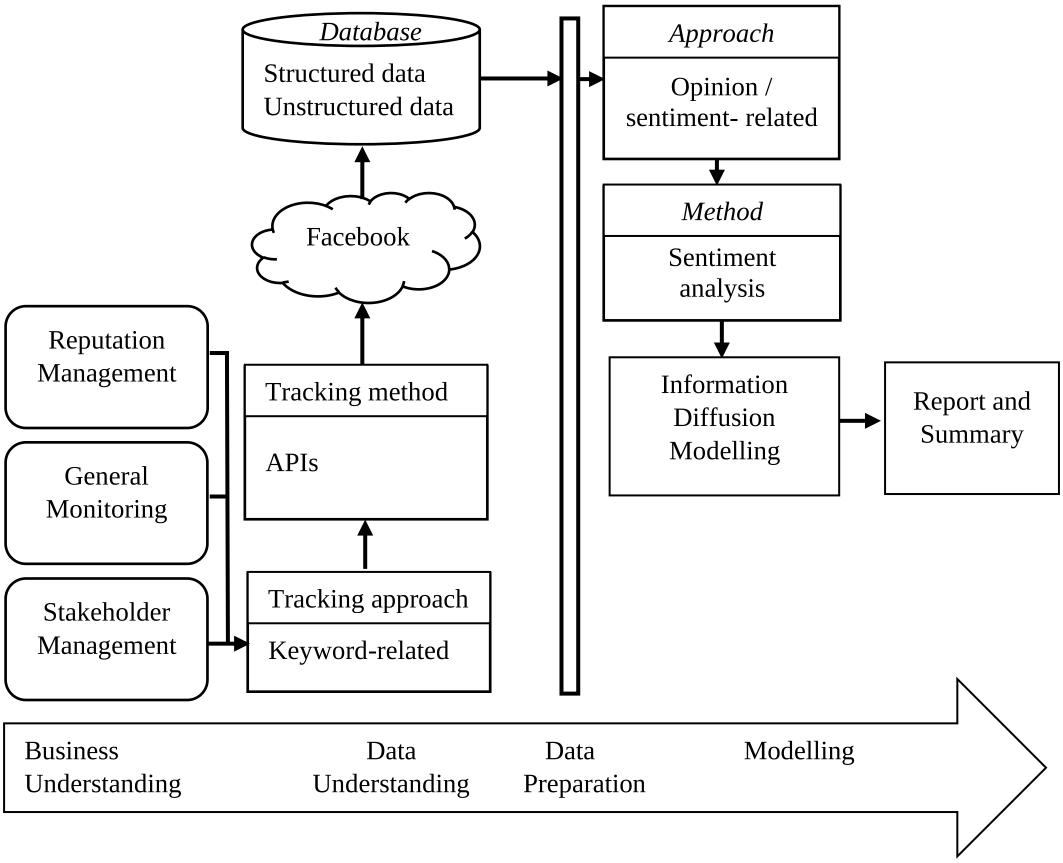


Figure 1.1*: Proposed approach overview*

1.7 Impact, Significance and Contribution

This research will help the university administrator to measure social media sentiments of higher education-related tweets that will directly help in their branding strategies. Another contribution of this research is the keywords identified for university branding to help on the enhancement of their branding strategies.

1.8 Background

Barnes (1954) has initiated the term "social network sites" (SNS). Social nets began with email and are presently generally utilized applications. New platforms are increasing tremendously with the advancement of social nets. The manners by which individuals get data have changed. Before, people were the recipients of information however at this point they are its dynamic distributers and communicators.

In the age of internet-based life such as utilizing of social media, people can get an informal student perspective on a college, which they would not discover on official college website pages. International Student Survey report (2017) shows that the noteworthy job internet-based social applications plays for some understudies picking up a university. The establishment of Social Media Marketing is one of the persistently progression throughout the entire existence of trade. Today, affiliations are using web-based social networking application to change client's conduct and to win their dedication. These correspondences assist publicists with choosing customer needs and realize what their market may look like. According to Chui and Manyika (2012), Rockendorf (2011), Forbes & Vespoli, (2013) Social Media Marketing can have positive effect on consumer buying decision making.

At the point when a piece of data streams beginning with one individual or system then onto the following in a framework, an information diffusion process has happened. Many studies have been putting effort on separating data dissemination, with most studies investigating which parts impact data dispersion, which data diffuses most quickly, and how data is dissipated (Christakis & Fowler, 2007; Zhang & Wu, 2012). These inquiries are addressed utilizing information diffusion models and different techniques, which assume a significant job in understanding the diffusion phenomenon. Nobody has the foggiest idea why the information streams to this course in social media, even though the upsides of a social network in information diffusion have seen. On the off chance that, utilizing information diffusion models, the significant clients and the components are impacting the information diffusion process can be figure out which helps in better understanding of such phenomenon.

1.9 Report Organization

This report covers 6 chapters in total. The first chapter is a brief introduction about the research such as the problem, motivation, proposed solution, objectives, scope and contribution of this research while the second chapter included the review of existing research towards the problem and related backgrounds. Chapter 3 presents the model developed while chapter 4 describes the methodology used for the method proposed in this project. In chapter 5, findings and results are presented. Lastly, this research is concluded in chapter 6.

Table 1.1*: Overview of the report.*

Chapter 1	• Problem Statement • Motivation • Project Scope • Research Question • Project Objectives • Proposed Approach • Impact, Significance and Contribution • Background
Chapter 2	• Literature Review Methodology • Social Media Landscape • Social Media Analytics • Social Media and Branding • Information Diffusion • Summary
Chapter 3	• Brand Fan Pages and Post Popularity • Influential Factors and Indicators of Brand Post • A Developed Conceptual Framework of University's Brand Post Popularity and Hypotheses
Chapter 4	• Case Study • CRoss Industry Standard Process for Data Mining Methodology • Social Media Analytics Framework • Proposed Approach Overview • Business Understanding • Implementation Issues and Challenges
Chapter 5	• Distribution of the samples • Relationship between the 6 Influential Factors and Information Diffusion
Chapter 6	• Conclusion • Validity and Generalizability • Limitations of Research Design • Future Work

CHAPTER 2 LITERATURE REVIEW

In this chapter, social media landscape, social media analytics and social media and branding are reviewed and discussed.

2.1 Literature Review Methodology

Studies often starts by choosing a regular issue and based on personal interest. Selecting an own-suited topic motivates passion towards working out the research as research is a long-term process. According to step 1 in ***Figure 2.1***, topic is chosen as **"Information Diffusion" and "Sentiment Analysis"** which is in the field of data mining. Secondly, since a writing survey must present a coherently contended case established on a far-reaching comprehension of the present condition of information, at that point the rules and instruments for building a casual contention must be utilized. A believable case produces decisions coming about because of a legitimate introduction of supporting proof. The apparatuses for proof structure, contention improvement, and coherent thinking are the structure squares used to put forth a valid defence. In this step, data is collected from Twitter and analysis is carried out to build the findings of this research.

Thirdly, search the literature across different databases to collect the related information. Databases used are **Scopus, IEEE Xplore, ScienceDirect and Google Scholar** accessed via UTAR web login. A keyword search is used as the query to collect related papers: **"information diffusion + brand", "emotion + sentiment analysis + branding"**. The searching period last over **3 months from 2nd August 2019 until 6th November 2019.** In step 4, reading and surveying the literature helps strengthening the understanding about related topics.

Table 2.1: *Amount of Literature studied.*

Categories of Literature	Amount of Paper
Quartile 1	25
Quartile 2	11
Quartile 3	7
Quartile 4	3
Conference Paper	13
Total	59

The literature is stored and managed by using Mendeley Desktop. Throughout reading and analysing the literature, strong foundation of understanding is built, and insights are discovered. Next, building argument and making critique of the literature. Lastly, review is drafted and revised few times finally established on this research.

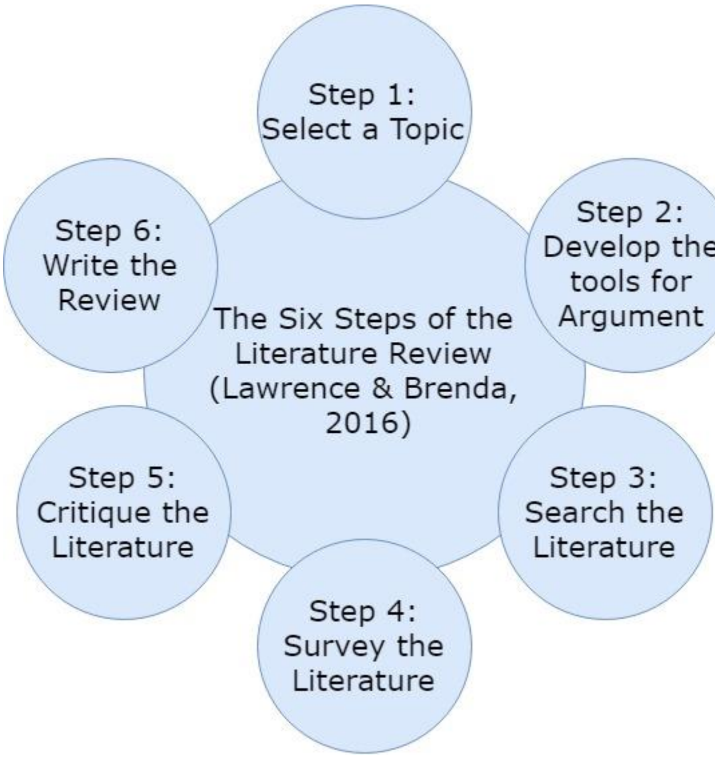


Figure 2.1*: Steps of literature review.*

2.2 Social Media Landscape

Social media is getting more advanced comparing to older days. It has become the domain for sharing data in governmental issues (Stieglitz and Dang-Xuan, 2013), stimulation (Shen, Chuan, and Cheng, 2016), emergency the executives (Stieglitz, Bunker, Mirbabaie, and Ehnis, 2017) and corporate (Beier and Wagner, 2016). The purpose behind the prevalence of internet-based networking application is the chance to get or then again make and offer open messages at small expenses and universally. Social media which has a huge transform becoming more advanced has prompted an expanding gathering of information.

Social media has numerous definitions in the literature while a few definitions originates from bloggers, intellectual and self-depicted specialist in this field. Policy makers have battled with not just how to benefit from social media strategies particularly profiting maturing society prosperity yet even how to compose social media so that it will be significant to their strategy creation. Kaplan & Haenlein (2010) characterize it as ‘‘assortment of online applications that operate with respect to the innovative establishments of Web 2.0 and that permit the formation and trade of customer generated content’’. In a general way, it covers web journals, Facebook profile pages, internet-based games and even wiki's, for example, Wikipedia.

Kaplan & Haenlein (2010) clustered social media according to a list of theories, including social nearness, media wealth, self-presentation and self- divulgence as a hypothetical establishment for a typology. There are six kinds of innovations being characterized as per degree of social existence or media extravagance (low, medium and high) and self-performance or self-divulgence (low and high). The six technologies include:

1. blogs,
2. social networking sites (e.g. Twitter, Instagram, WhatsApp),
3. virtual social worlds (e.g. Jump VR),
4. cooperative contents (e.g. Wikipedia),
5. content communities (e.g. Dailymotion) and
6. open online games (e.g. DotA, Mobile Legends).

Social networking sites, where more often denoted as social media, is viewed as high on self-introduction or self-divulgence, described as any ‘‘sort of social collaboration individuals wish to control the impressions other individuals type of them’’ (Kaplan & Haenlein, 2010). It is named medium on social nearness or media lavishness, characterized as ‘‘the auditory, pictorial, and bodily contact that can be accomplished – that permits to rise between two correspondence accomplices’’ (Kaplan & Haenlein, 2010).

Social media platform offers numerous potential outcomes of data presentations, including geolocations, sounds, recordings, pictures, and textual data. For the most part, this data usually managed by categorised it into structure and unstructured data (Baars and Kemper, 2008). The follower connection is an instance of structured information while geolocation is an example of unstructured information.

2.2.1 Social Networking Sites (SNS)

As indicated by boyd & Ellison (2007) social network sites is as internet-based services that enable people to

1. articulate a rundown of different clients with whom they share an association,
2. view and navigate their rundown of associations and those made by others inside the framework, and
3. develop an open or semi-public profile inside a limited framework.

Social media essentially should mean any online medium that permits client impart and intermingle socially. Social Networking Sites embodies Apps, sites or internet-based pages that enable social media happenings as defined above.

Social media is an umbrella term and SNS, email, texting, blogging is on the whole sorts of social media (boyd & Ellison., 2007). Despite the fact that the use of web-based social networking like email, texting and blogging is broad, yet it just offers one-to-one or small group-oriented communication (boyd & Ellison., 2007). Besides, SNS permit huge scale communication and support with various groups of clients. A portion of the instances of SNS are Facebook, Twitter, Instagram, LinkedIn, WhatsApp and others. ***Figure 2.2*** below provides a simple illustration of social media map.

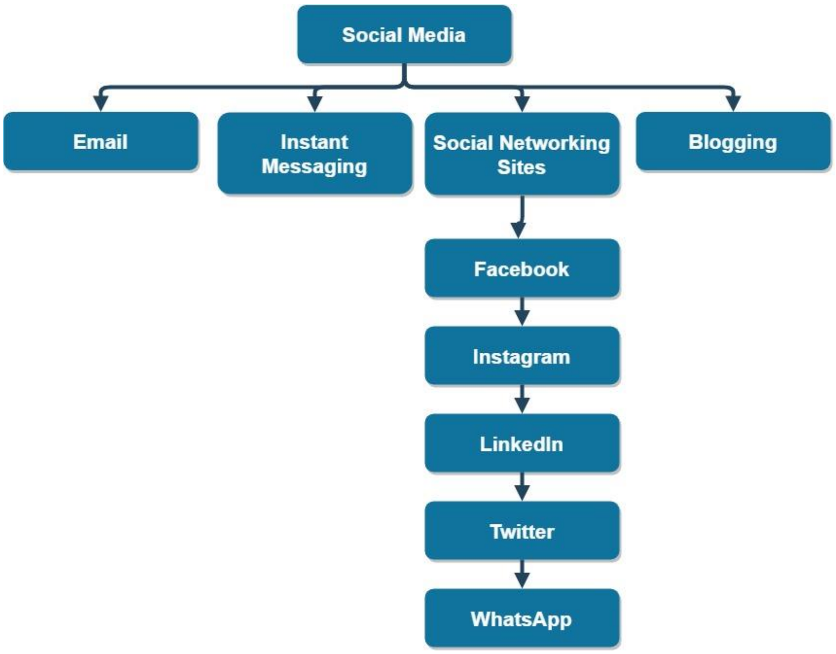


***Figure 2.2**: Social media map*

2.2.2 Twitter Use

The usage of Twitter has gain focus in different backgrounds by recent research. A study was led by Kwak et al. (2010) to investigate Twitter's topological qualities and uncover its capacity as another mode of information distribution. The topological and topographical possessions of Twitter's social net is analysed by Java et al. (2007) from Twitter's open course of events. There is various utilization recognised such as day by day babble, discussions, sharing data/URLs, and revealing news.

Recent studies have shown that Twitter is used mostly to notify others and to convey what needs be. For instances, the messages gathered through the open course of events is coded manually to inspect the content of 3,379 tweets. It is discovered that one fifth of the clients in their study uploaded post involving themselves or their contemplations, instead of sharing general news (Naaman et al., 2010). Distinctive interactive parts of retweeting on Twitter is discovered by Boyd et al. (2010). For instance, how creation, attribution, and informative devotion are consulted in assorted ways tending to the practice of retweeting. In addition, retweeting is expected to be engaging a particular group of spectators, to remark on somebody's tweet, to freely concur with somebody, or to save tweets for forthcoming individual access.

The role of influencing politics has been played by social media in the United States and round the global (Farrell & Drezner, 2008; Tumasjan et al., 2011; Wattal et al., 2011). For instance, studies revealed that representative from U.S. Assembly use Twitter as a channel to share their thoughts and activities, especially as a part of their political campaigns (Golbeck et al., 2010). Meanwhile, there is also other studies have conducted against citizens in the political context round the world such as Germany (Tumasjan, 2011), Sweden (Larsson & Moe, 2012), Iran (Gaffney, 2010), and the United States (Conover et al., 2011; Wattal et al., 2010).

The research resulted that Twitter is a power tool for sharing political-related tweets. Twitter has become a platform for talkers who previously subsidiary with jutting positions in regular media or overall political contention. However, the research only focusses on political tweets leaving little research directly investigating the tweets for branding university purpose.

Twitter is the one of the massive social media data resources. The data from Twitter can be accessed via tools and via Application programming interface (API) (Batrinca & Treleaven, 2014). Data can be accessed via tools as Twitter provide controlled access to their social media data via dedicated tools, both to facilitate easy interrogation and to stop users tracking all the data from the repository. Meanwhile, data can be accessed via APIs as Twitter data repositories providing programmable HTTP-based access to the data via APIs (Batrinca & Treleaven, 2014). In this research, Twitter Search API is used.

2.3 Social Media Analytics

The term social media is referred as "Internet-based applications that work with respect to the ideological and mechanical establishments of Web 2.0", where Web 2.0 implies that "content what's more, applications are never again made and distributed by people, however, rather are constantly changed by all clients in a participatory and community oriented style" (Kaplan and Haenlein, 2010). As a result of the wide-ranging meaning of social media, its application purposes are diverse.

In spite of the enormous assortment of platform, a few attributes are common to a significant number of them. As a result of the measure of the content delivered every day and the quantity of dynamic clients on the platform, associations are inspired to comprehend which issues and patterns develop to recognize risk and chances in the interaction and infer helpful suggestions. It is additionally applicable for associations to comprehend who makes the content and which entertainers are the most persuasive drivers in the communication other than the measure of content. The business organizations and non-benefit associations try to gather the information created by the user so as to pick up bits of understandings into mass communication. The information is regularly assorted with apparatuses which speak with the particular API of the social networking platform, on the off chance that one exists, and skulk the data.

A lot of attention is directed to the term “Social Media Analytics”. It is characterized as “a rising interdisciplinary study field that emphasize on joining, broadening, and adjusting strategies for examination of social media data” (Zeng, et al., 2010). While the point of view on the framework is one significant viewpoint, yet there is another angle is the point of view on the clients who generate the content. Research that receives this point of view investigates various role in the communication and the impacts an individual role can have on the communication and the diffusion of information (Stieglitz et al., 2017). For instance, influencers or opinion pioneers can be recognized through a social network analysis, and by looking at their follower arrange, one can uncover the range of such an individual (Mirbabaie, Ehnis, Stieglitz, and Bunker, 2014; Mirbabaie and Zapatka, 2017). Furthermore, the behaviour of the roles is analysed in request to comprehend the reasons of a key role in the network and the impacts it has on the general network (Mirbabaie et al., 2014). Mirbabaie et al (2014) led a research on understanding of the roles that open occasions which is labour day members play in their utilization of social media. It has limitation that the event is comparatively strategic, well-controlled which may not be able to be efficiently scaled up to a less controlled and more chaotic event (such as a crisis event) and too general which is not specific to education. This study has been focused more on public events in Germany but not Malaysia. Therefore, it remains a little research investigating the roles that twitter user engage with tweets related to higher education. One research objective might be to identify and analyse how the information related to higher education in Malaysia diffuse in Twitter.

2.4 Social Media and Branding

Business companies are now shifting its brand-related events to the channel of social media which is accepting noteworthy attention (L. de Vries, Gensler, & Leeflang, 2012). Correspondingly, universities are progressively changing their marketing strategies by publishing its programme information to social media platform. (Belanger et al., 2014).

Social media has been offering a major function which is generating and sharing opinions online (Kaplan & Haenlein, 2010), and, not at all like customary interactive channels, the time taken for the process is shorten a lot. Not only that, its span conceivably worldwide (Hakala et al., 2017). Consequently, captivating social media platform require manpower from management whom are wise-thinking and smart maintaining public attention on its channel ensuring brand messages convey correctly and excel in crisis management. Social media motivates engagement of two parties including discussion and participation in a two-way manner (Constantinides & Zinck Stagno, 2011). As indicated by a work by Barnes and Lescault (2011), social media platform that stand out the most with 98% reception being utilized by colleges and universities in the US is Facebook. Accordingly, Twitter ranked in second place after Facebook. Researchers can utilize social media platform to explore important insights as it is low-cost and swift in information sharing which would then contribute to strategies and approaches of tertiary education (Davis et al.,2012).

The expanding noticeable quality of social media in the promoting efforts of advanced academic organizations, including colleges in Canada concentrating on the utilization of internet-based socializing application as a channel to investigate the college reputation guarantee (Belanger et al., 2014). Rutter, Roper, and Lettice (2016) clarified that the association between the intensity of social network and branding is additionally bolstered as they contend that "Twitter followers acting as the middleperson of the brand quality and notoriety of the university reputation… and empowering relationship with purchasers who guarantee the brand is essential to the productive usage of web-based social networking". At last, Davis et al. (2012) has led a study revealing that 'not many of the studies [reviewed] utilised information that produced from SMT [social media technology] sites.'

Pringle and Fritz (2018) has led explore on researching the effect and impact of social media as a canal for marking 3 distinct universities in Ontario, Canada. The research uncovered that the subject distributes on social networking sites at three higher education institutions while generally reliable, is full of shades of grey adjusting regular bran informing with organizational standards and qualities (Rutter, Lettice, and Nadeau, 2017). As the research focuses on United States, less attention is directed at branding higher education in Malaysia.

2.5 Information Diffusion

The corporate and advertising writing tending to information dissemination as far as internet-based informal exchange and epidemiologic promoting (Godes & Mayzlin, 2004; Leskovec et al., 2006). Information diffusion research has progressively directed consideration toward various social media platform, for example, SNS (Sun et al.,2009), weblogs (Chai et al., 2012; Goetz et al., 2009; Wattal et al., 2010), picture-sharing portals (Cha et al., 2008), just as online communities (Chen & Whinston, 2011; Garg & Telang, 2011). Specifically, an enormous number of studies have concentrated on Twitter (Cha et al., 2010) as it gives an express method to stamp the diffusion of information as retweets. Relations in the social nets have acting an essential character in the dispersion of information with the evidence of retweets sign.

Suh et al. (2010) constructed a predictive retweet model in an enormous scale investigation of 74 million tweets, and recognized a few variables influencing the amount of retweets a Twitter message gets, including URL posting and hashtag incorporation just as the quantity of followers and the age of clients' records. Macskassy and Michelson (2011) displayed diverse retweet models assessed on a Twitter dataset comprising of more than 768,000 tweets assembled from checking more than 30,000 clients for a time of one month. They found that context explicit retweet models with respect to homophily or similitude are better at clarifying the observed retweet behaviour than general or network-based models.

In any case, as far as anyone is concerned, a likely connection among sentiment and information diffusion on Twitter has yet to explore. Few researches related to information propagation has concentrated uniquely on retweet amount as a part of it. Regardless, another vital part of information diffusion still is the rapidity at which information disperses through systems (Yang & Counts, 2010). Hence, whether and how sentiment of Twitter messages associates with the retweet behaviour regarding retweet amount along with retweet swiftness as a parameter to improve information diffusion in internet-based social communities has leaving a little research area to investigate.

Based on ***Figure 2.3***, the literature identified can be characterized into two classes which is explanatory models and predictive models. This research explores, investigate and examined the most generally utilized fundamental models and their accessible models in these two classes. The information diffusion model reviewed are portrayed in ***Figure 2.4***.

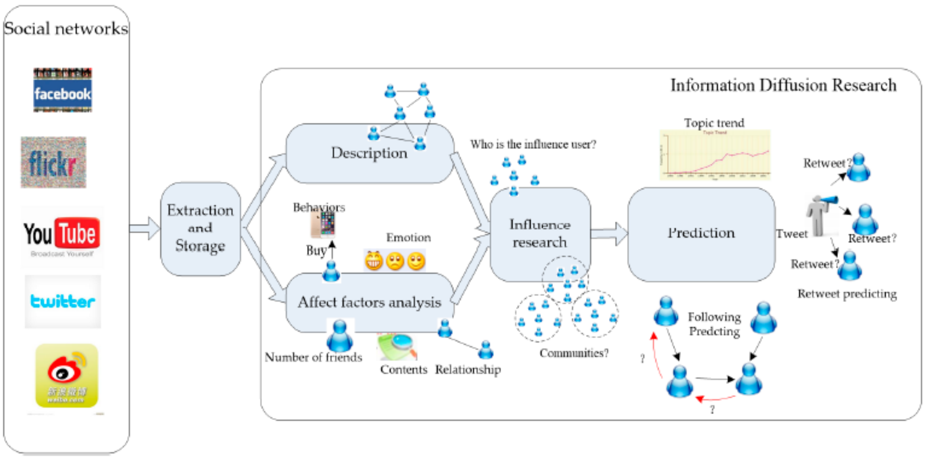


Figure 2.3: *Research roadmap on Information Diffusion (Mei, et al., 2017).*

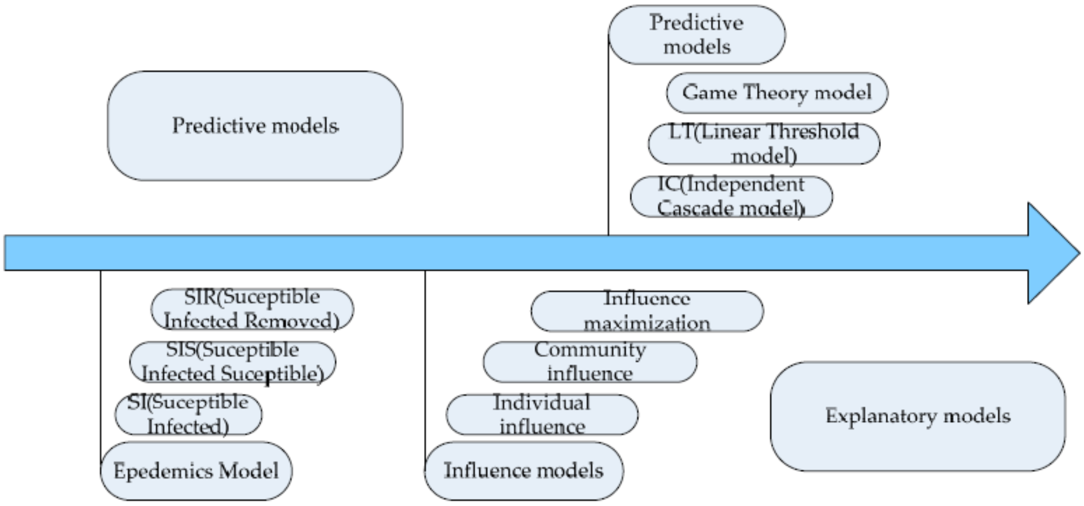


Figure 2.4*: Categorization of Information Diffusion Models (Mei, et al., 2017).*

2.5.1 Explanatory model

Information is spread by method for communications between various people in the public eye. These people can be viewed as nodes in social networks. A node in social network is a conceptual portrayal of a client in "genuine" society. The collaborations between two clients might be viewed as relations, which are embodied edges running between two nodes in social network. Along these lines, a genuine social cluster can be mapped by an enormous social network and a snippet of data can be scattered by these nodes inside it. This brings up numerous issues about the information diffusion process, for example, what are the fundamental factors that influence information diffusion? Which node has the most impact? For what reason does the information diffuse the manner in which it does? For instance, a few nodes will not acknowledge information, some won't spread data, and some both acknowledge and spread data (Han & Niu, 2013). Various clustering of nodes additionally has various qualities: some are homogenous, and some are heterogeneous (Ou et al., 2017).

Epidemics model

The information diffusion can be considered similarly as an epidemic distribution process. In epidemics transmission, there are two clients tainted with pathogens and clients who are vulnerable to the pathogens. The infection can spread from the tainted clients to defenceless clients. Furthermore, data can be diffused from communicators to beneficiaries likewise. To explore data dispersion, it bodes well to gain from the essential scourge models. In the compartment model of scourges, the essential models are SI (Susceptible Infected) model, SIS (Susceptible Infected Susceptible) model, SIR

(Susceptible Infected Removed) model and SIRS (Susceptible Infected Removed Susceptible) model.

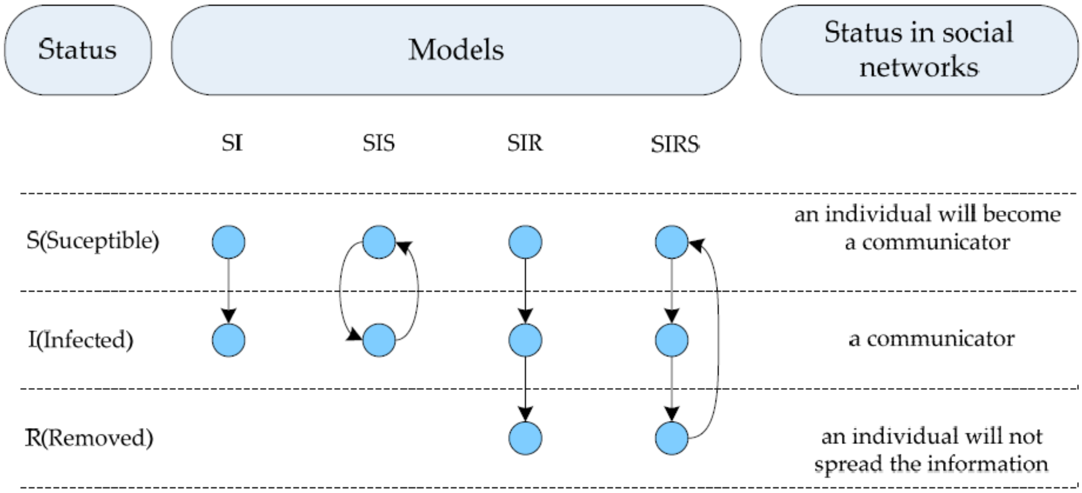


Figure 2.5: *Comparison of four basic epidemic models. (Mei, et al., 2017)*

Influence model

Influence examination is key in interpersonal networks. Data dissemination dependent on impact is separated into three classifications: singular influence, community influence and impact maximization.

1. **Individual influence**

   Singular impact alludes to conclusion pioneers related studies. Conclusion pioneers are the hubs who can assume a job as a scaffold of data dissemination. They impact other clients in an interpersonal organization.

Researcher	Network Structure	User Interactions	User Attributes		Method	Quantitative Criterion	Applications
			User behaviors	Other features			
Chenxu (2015)	√	-	-	-	social network analysis	out-degree	identify opinion leaders and prediction
Bo (2014)	-	√	√	centrality	competency	activists, centrality and intermediary	identify opinion leaders and influence maximization
Jiaxin (2014)	√	-	√	access time	social network analysis	capability of diffusion	influence predicting
Xianhui (2015)	√	√	√	topic and weight	page-rank	coverage and coreratio	mining topic opinion leader
Ullah (2017)	√	√	√	neighbors-of-neighbors	social network analysis	activists	identify influential nodes

Figure 2.6: *Comparison of the individual influence methods.*

2. Community influence

A group of people is a gathering of individuals with some regular characteristics. In informal organizations, people will frame different networks based on interests. A group of the people is a subcategory of the system wherein the clients are thickly associated and have comparable qualities. For instance, they share same hobby, or their examination territory is comparable. In spite of the fact that the structure of informal organizations will change after some time, networks remain moderately steady. The principle challenge is the manner by which to recognize those networks that include high impact inside an informal community.

Model	Links	Attributes or Contents	Sentiment	Method	Quantitative Criterion
PCL-DC (2014)	√	√	-	probability	-
SA-Cluster-Inc (2010)	√	prolific and topic	-	cluster	density and entropy function
CODICIL (2012)	√	stemmed words, title and context, tags	-	cluster	quality function
sentiment-topic based (2014)	√	user, text	√	probability	sentiment-topic similarity
SVO (2015)	√	interests	√	cluster	homophily
interest and trust based (2017)	√	interest, trust	-	both	quality function

Figure 2.7*: Comparison of the community influence methods.*

3. Influence maximization

Recent studies have been centred around both the individual level and network level. The regular object of these two levels is to discover seed hubs and expand their impact. Impact boost studies is consistently information and model driven. In a model-driven calculation, a known impact dispersion model is given at first, at that point a specific heuristic calculation can be utilized to pick seed hubs. It isn't versatile for some system topology. Be that as it may, the investigation of the informal communities depends on genuine social network data in information driven models. A definitive model will at that point be accomplished by method for a learning procedure. Accordingly, these models are versatile.

Model	Find Seeds	Techniques for Choosing Seed Nodes	Data/Model Driven	Multi-Round	Multi Innovations/Items/ Information	Application
OIM (2015)	√	explore-exploit, heuristic	model	√	-	individual influence maximization
Adaptively Seeding (2015)	√	friendship paradox	data	-	-	
CASINO (2014)	√	conformity aware is mentioned	data	√	-	
Optimal percolation (2015)	√	the important of weak nodes	data	-	-	
STORM (2015)	√	maximization the total gain	data	√	√	competitive influence maximization
GETREAL (2015)	√	game theory	model	-	√	

Figure 2.8*: Comparison of influence maximization method.*

2.5.2 Predictive model

When a bit of significant data is distributed by an individual in social network, the data will be spread rapidly all through the social network. Particularly on account of "awful" data, an administration will need to know how a circumstance will create: having the option to anticipate how data will spread all through the system later will be helpful. Predictive models are utilized to anticipate the future information diffusion process in social network dependent on specific elements. These models are additionally regularly utilized for influence maximization. They are the IC model, the LT model, and the Game Theory model.

Table 2.2*: Comparison of Independent Cascade Model (ICM), Linear Threshold Model (LTM) and Game Theory Model (Mei, et al., 2017).*

Model	Basic Model			Research Views	Application
	IC	LT	GT		
EM (Saito et al., 2008)	√			The probability for information diffusion events	Forecast of dissemination likelihood
ASIM (Arora et al.,2015)	√			Join running-time with memory-consumption	Influence maximization
TIC, TLT (Barbieri, 2012)	√	√		Topic-aware	Forecast of subject distribution
DRUC (Lagnier et al.,2013)		√		Information content and user profile	Affect factors discovery
Heuristic and Greedy		√		Influence of nodes and the node's	Choose the greatest

(Chen & Wang, 2012)				initiation threshold	influence nodes
Microscopic (Hang et al.,2014)			√	Relationship and cost	Forecast of the information dissects
Evolutionary game (Wang et al., 2015)			√	Individual information behaviour in micro level	Forecast of information diffusion in dynamic network
Game Coalitional (Liu et al,2014)			√	Structure of social network and communicating features	Relationship prediction

2.6 Summary

Social media covers blogs, social networking sites, virtual social worlds, collaborative contents, content communities and open online games. From there, a vast quantity of data is available online and there occurs a need to analyse the large volumes of user-generated content to gain meaningful insights into the diffusion of information, opinions and sentiments about branding activities.

CHAPTER 3 CONCEPTUAL MODEL

This chapter will first present and discuss definition of popularity of university's brand post; second introduce a conceptual model inspired by (de Vries, et al., 2012) and explain why it is regarded as a proper model for this project. The model is referenced to develop an information diffusion model for Malaysia's university branding and identify particular factor for the popularity of social media posts from Malaysia's university.

3.1 Brand Fan Pages and Post Popularity

Social Networking Sites have become very famous in just only a couple of years. Facebook, for example, have pulled in over 2.5 billion active users as of April 2019. Individuals can follow a brand page (becoming fan of brands on these dedicated pages). Brand fans can voice out their opinions about the brand on these committed pages which will invoke much more interaction and customer engagements (Kozinets 1999).

Company must make a good use of social network which act as an information diffusion platform for branding their products which will encourage interaction from customers by liking or commenting on them. Everyone has different measurements and interpretations on "Popularity of Online Content". In any case, the common recognition is that a legitimate meaning of popularity needs to mirror the speed and volume of information diffuse into the people (Zadeh & Sharda, 2014). Drawing upon past academic works, this research will build up a complete definition of popularity which fit for this purpose on the domain of university.

Based on the existing literature, there are two aspects to define popularity which are level of customer engagement and the nature of brand posts. Customer engagement including liking, reacting, commenting and sharing which plays a crucial role in social media marketing (Muntinga, Moorman and Smit, 2011; Alton Chua, and Snehasish Banerjee, 2015). At the meantime, there is other works discussed about the nature of brand posts which will affect the popularity of brand post. Brand post content is often categorized into information type of content and brand post characteristics. There are informational posts and entertaining posts for the primary category; whilst for subsequent category, vividness and interactivity are the most regularly cited factors by the researches. The most broadly referred among different models is the model which proposed by de Vries et al. (2012), where they consolidated both of these two perspectives together to evaluate brand post popularity.

De Vries et al. (2012) have chosen 6 potential persuasive components to look at their relationship with brand post popularity which is reflected by two pointers – the number of likes, the number of comments. Aside from 4 unmistakable elements which consists of vividness, interactivity, informational and entertaining content, they have additionally incorporated two factors: position of the post and valence of comments into the relationship assessment. Recent studies on advertising shown that position of advertisement is crucial for click-through rates. Advertisement placed on topmost page attracts more clicks (Rutz andTrusov, 2011). The valence of comment alludes to the mentality direction of the crowd—in light of the pennant publicizing writing and the verbal correspondence writing. A **Conceptual Framework of Brand Post Popularity** as de Vries et al. have developed is shown in ***Figure 3.1***.

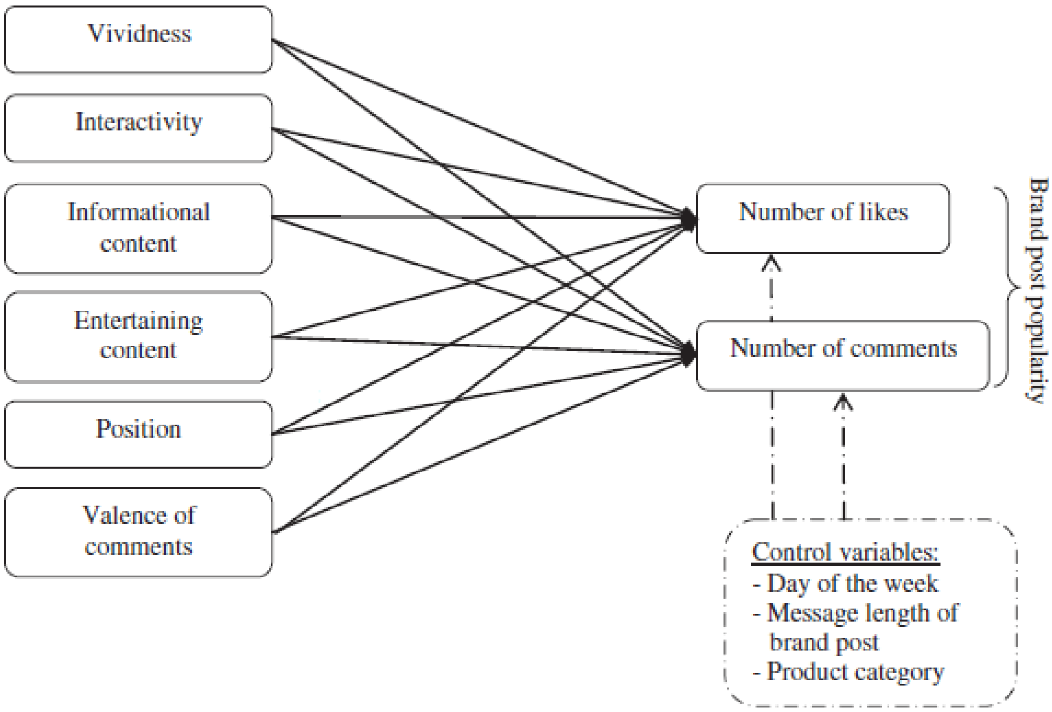


Figure 3.1*: Conceptual Framework proposed by de Vries et al. (2012).*

In this study which conducted by de Vries et al. (2012), there are 355 brand posts from 11 international brands across six product categories being analysed to determine the possible drivers for brand post popularity. The developed conceptual model is based upon the findings from the banner and advertising literature as well as the word-of-mouth theories. By selecting the brand post which posted during weekdays, word count of the post content and the product category, they argued that vividness, interactivity, the content of the brand post (information, entertainment), the top position of a brand post, and the valence of comments on a brand post are related to brand post popularity.

According to de Vries et al. (2012), they were the first to study the influential factors of brand posts at a social networking sites and this work has been widely cited which is around 1655 citations as of now by other researchers. WeiXian (2016) has adopted de Vries et al. framework in his study to investigate the popularity of brand post on Sina Weibo and found out that the factor of entertaining content may act differently in Western countries and China. In Western countries, entertaining content does not play a crucial role in enhancing the number of likes and comments whilst in China, entertainment is affecting the brand post popularity largely. To conclude, the generality of model developed by de Vries et al. (2012) is still quite limited as it only focuses on one social media site. Nevertheless, the factor of interactivity and vividness still important as well in the findings of WeiXian (2012). Hence, de Vries et al's study established a relatively high rate of reliability and suitable to be adopt as a conceptual framework. Being inspired by this widely cited referenced conceptual framework, this research will present and discuss the influential factors and indicators of brand post popularity for Malaysia's university on Facebook.

3.2 Influential Factors and Indicators of Brand Post

There are two of the usually referred to variables of brand post popularity in the existing writing are that of brand post characteristics and that of the content of the brand post. With respect to, the number of likes, comments and shares are regularly represented the signs of brand post popularity which is a process of information diffusion. In this segment, various investigations on brand post are introduced to survey different elements and pointers for brand posts from social media pages of Malaysia's universities, and both of these elements and pointers are breaking down to delineate those qualified ones that are for use with a relationship test.

3.2.1 Brand Post Popularity and its Indicator

Level of user engagement in brand posts can be observed from the indicators such as liking, commenting and sharing as they are important to a successful Social Media marketing (Muntinga, Moorman and Smit, 2011; Alton Chua, and Snehasish Banerjee, 2015). In de Vries et al.'s study, brand post popularity is measured by number of likes and comments in Twitter as they mentioned that liking and commenting on a brand post is thus similar to Word-of-Mouth communication. Followers of the brand can share their views freely by liking or commenting on a brand post. In the meantime, WeiXian (2016) has measured using only the number of reposts (share) and replies (comment) in Sina Weibo. The segment of "favourites" (like) in Sina Weibo have been avoided in light of the fact that as indicated by Chu and Sung (2011), Weibo clients have yet framed the propensity for utilizing the "thumbs up" button to communicate affection, so a post's "favourites" number may not honestly speak to the amount it is enjoyed by the crowds.

In this research, the amount of likes and number of comments which mutually included as well in de Vries's study, as well as an additional parameter - the number of shares, are collected to measure the brand popularity as well as information diffusion. As discussed in ***Section2.5***, relations in the social nets have acting an essential character in the dispersion of information with the evidence of retweets (share) sign. In order to study whether and how sentiment of Facebook posts associates with the sharing behaviour regarding the number of sharing, this parameter is added to improve information diffusion in Social Media network.

3.2.2 Vividness

In 1992, Steuer mentioned that vividness mirrors the extravagance of a brand post's proper characteristics; as such, it is the degree to which a brand post invigorates the various senses. The consideration of lively components such as (contrasting) colours, or pictures (Goodrich 2011) or videos can accomplish vividness. For instance, a video is more vivid than an image in light of the fact that a video invigorates sight, yet in hearing.

As indicated by Coyle and Thorson (2001), various degrees of vividness are positioned in this sense: a high level of vividness (sound present and movement present), medium level (either sound present or movement present) and low level (sound missing and animation missing). Fortin and Dholakia (2005) created the levels that are suitable for Social Networking Sites by characterizing the low level as pictorial content, with the medium level as an upcoming occasion (offline) declaration, and the elevated level as content containing video.

Research proven that vastly vivid posters are increasingly compelling regarding click-through rates (Lohtia, Donthu, and Hershberger 2003). In addition, higher degrees of vividness give off an impression of being best to attract user towards a site (Fortin and Dholakia 2005). This research proposes that higher vividness a brand post, the more popular the brand post.

Therefore, we formulate:

H1: When the level of vividness of a brand post increase, the popularity of the brand post will increase.

3.2.3 Interactivity

Liu and Shrum (2002) defined that interactivity is defined as "how much two or more community can follow up on one another, on the communication medium, and on the messages and the degree to which such impacts are synchronized". Interactivity is s described by two-way communication among organizations and clients, just as between clients themselves; put in an unexpected way, it describes many-to-many communication (Goldfarb and Tucker 2011; Hoffman and Novak 1996).

Brand post qualities vary in the degree of interactivity. For instance, text element not really interactive. Meanwhile, a website link is progressively communicating (Fortin and Dholakia 2005) subsequently brand fans can navigate themselves through the link to connect themselves to the website. Moreover, a post with question is highly interactive as it incubates brand fans to think and challenge the fans to tackle the question. As WeiXian's study (2016) and de Vries et al.'s study (2012) proven that interactivity plays an important role in both social media platform, Twitter and Sina Weibo, this research expect that higher degrees of interactivity will generate more likes and comments.

H2: The higher the level of interactivity of a brand post, the more popular the brand post.

3.2.4 Informational Content

Puto and Wells (1984) has defined “informational advertising” as one which furnishes customers with true, applicable brand information in a clear and legitimate way to such an extent that buyers have more noteworthy to survey the functional properties of products and services. Information-seeking is a significant explanation for people to utilize social networking sites (Lin and Lu 2011). Consumers may want to browse brand pages to refer other informal information such as consumers’ voices about the brand as social media users can voice out their opinions on the post. As such, this research proposes:

H3: The content of brand post which is informational widespread than the non-informational brand post.

3.2.5 Entertaining Content

The entertainment value of a social networking site is a significant factor for utilizing it as it was reported in 2019 that the average time spent on social networking was projected as 144 minutes per day. According to Zephoria (2020), the average time spent per Facebook visit is 20 minutes. Therefore, administrator of brand pages has to retain customer in such short period hence, entertaining is a good weapon to tackle it. A study on brand post popularity in Sina Weibo conducted by WeiXian (2016) has showed that entertainment factor is very popular and important to brand post popularity in China compared to Western countries. Hence, this research proposes:

H4: Entertaining brand posts are more popular than non-entertaining brand posts.

3.2.6 Sentiment of Brand Post

The perceptions of consumers to a value of product is very crucial, it represents the probability to endorse the product (Gruen, Osmonbekov, and Czaplewski 2006), and sales (Chevalier and Mayzlin 2006). A positive post may impress consumers with their brand image. Contrary, a negative post may ruin the brand image. In any case, as far as anyone is concerned, a likely connection among sentiment and information diffusion on brand post has yet to explore. To explore the relationship between sentiment of brand post and popularity in Facebook, this research proposes:

H5a: The positive posts are positively related to a brand post popularity.

H5b: The negative posts are negatively related to a brand post popularity.

3.2.7 Hashtag use

Hashtag use referring to prefix a keyword by a # symbol, a Twitter hashtag fills in as a base on user-proposed tagging convention. Hashtag use has become a one of a kind labelling show to assist with post content with specific occasions or settings. Hashtags are valuable with regards to gatherings or occasions, given that the hashtags have been declared or advertised. As of late, Major League Baseball proceeded with the utilization of hashtags to advance different activities during their All-Star occasions (Blaszka, 2012). A study conducted by Blaszka et al. (2012) inspected "#WorldSeries", which was explicitly made by Major League Baseball. There are insights found regarding gratifications of consumers through the assessment of this hashtag in which significant game properties can better draw in their fan base during sporting games. A Twitter hashtag archive is the result of an aggregate exertion that the posts can be totalled into a single stream with the regular #hashtag. Hence, this research proposes:

H6: Brand post with hashtag popular than those brand post without hashtag.

3.3 A Developed Conceptual Framework of University's Brand Post Popularity

In the light of research conducted by de Vries et al. (2012), a conceptual framework of brand post popularity with case study of Malaysia's university on Facebook is presented in ***Figure 3.2***. The "position" and "valence of comments" elements in the original model have been replaced by "sentiment of post" and "hashtag use". In this model, there are three indicators of brand post popularity – the number of likes, the number of comments and the number of shares. A Pearson Correlation test will be performed on the conceptual framework proposed.

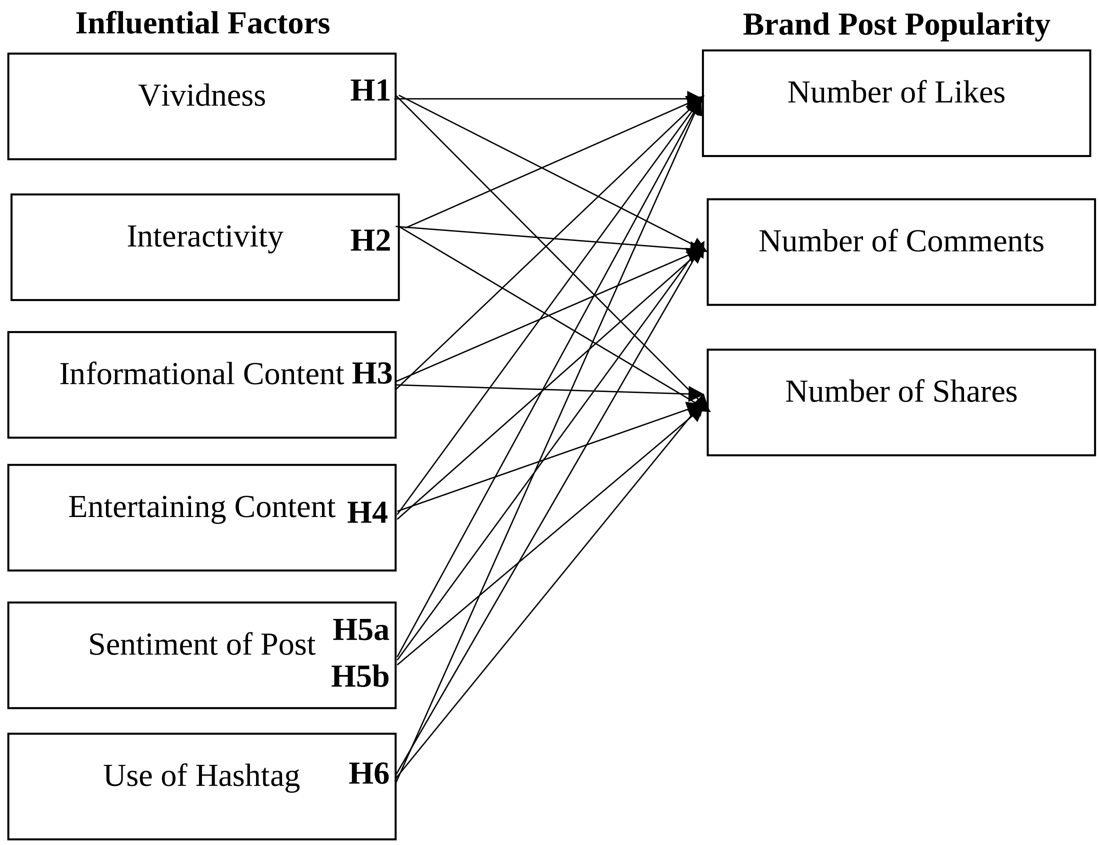


***Figure 3.2**: Conceptual Framework Proposed.*

CHAPTER 4 METHODOLOGY

This chapter will firstly present how the case study of university is deployed in this research; and second, the methodology adopted from CRISP-DM (Wirth & Hipp, n.d.) and Social Media Analytics Framework (Stieglitz, et al., 2018); and third, the data sampling and collection; and lastly, the method of Pearson Correlation Analysis is introduced. Besides, the operationalizations for each factor is presented and discussed.

4.1 Case Study

According to Creswell (2002), case study is defined as a problem to be considered, which will uncover a top to bottom comprehension of a "case" or bounded system, which includes understanding an occasion, activity, procedure, or one or more individuals.

This research seeks to study how information posted by higher education institution in Malaysia spread among social media users. In other words, how Malaysia's university perform in social media marketing is the findings of this research. This research tends to examine the behaviour of social media user associated with taking part on a brand's Social Networking sites. Moreover, from the point of view of brand, there is likewise the topic of how they can present an open door for these online practices to support popularity and effectively actualize social media marketing.

4.2 CRoss Industry Standard Process for Data Mining Methodology

Cross-Industry Standard Process for Data Mining (CRISP-DM) is used for the methodology in this research. The CRISP-DM project proposed a comprehensive process model for carrying out data mining projects which is suitable for this research. The process model is independent of both the industry sector and the technology used (Wirth & Hipp, n.d.). CRISP-DM aid to project planning and management and encourage best practices. It is initially launched in late 1996 by Daimler, NCR, and OHRA and later CRISP-DM 1.0 was published in 1999. CRISP-DM breaks the data mining into six major phases which are business understanding, data understanding, data preparation, modelling, evaluation and deployment. The lifecycle of these six phases is shown in ***Figure 4.1***.

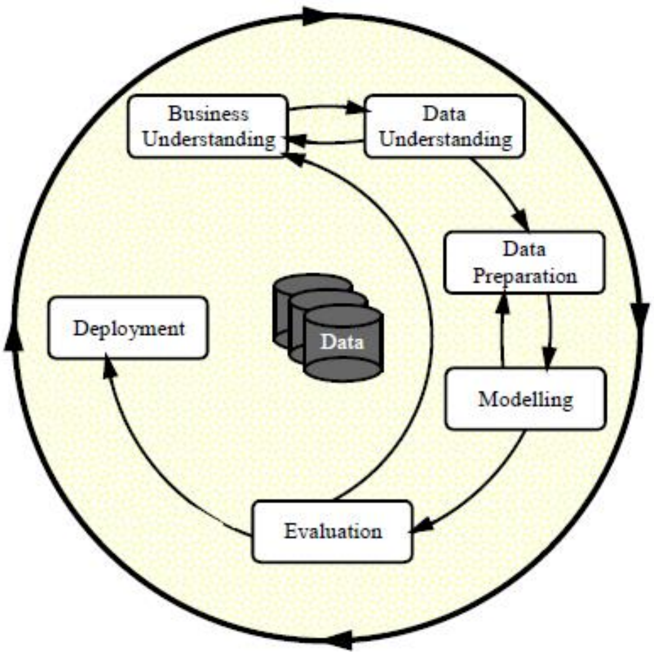


Figure 4.1*: CRISP-DM process model for data mining (Wirth & Hipp, n.d.)*

According to KDNuggests polls results as shown below in ***Table 4.1,*** the most popular methodology for data mining process is CRISP-DM, followed by SEMMA and KDD (Piatetsky, 2019).

Table 4.1*: KDNuggets Poll on Data Mining Methodology results*

Poll Years	**2002**	**2004**	**2007**	**2014**
CRISP-DM	51%	42%	42%	43%
SEMMA	12%	10%	13%	8.5%
KDD Process			7%	7.5%
My organizations’	7%	6%	5%	3.5%
My own	23%	28%	19%	27.5%
None	4%	7%	5%	0%

Table 4.1 has showed that the usage of other methodologies is steadily decreasing while the usage of CRISP-DM has achieved a level.

4.3 Social Media Analytics Framework

Stieglitz et al (2018) has identified the challenges in social media analytics and proposed framework as shown in ***Figure 4.2*** to conquer the challenges faced. The challenges in social media are conceptualising by using four-step social media framework. The phases of social media analytics are disclosure, assortment and groundwork as Chinnov et al. (2015) mentioned with the volume of accessible word-based data, the requirement for programmed strategies of topic discovery in the Internet develops exponentially.

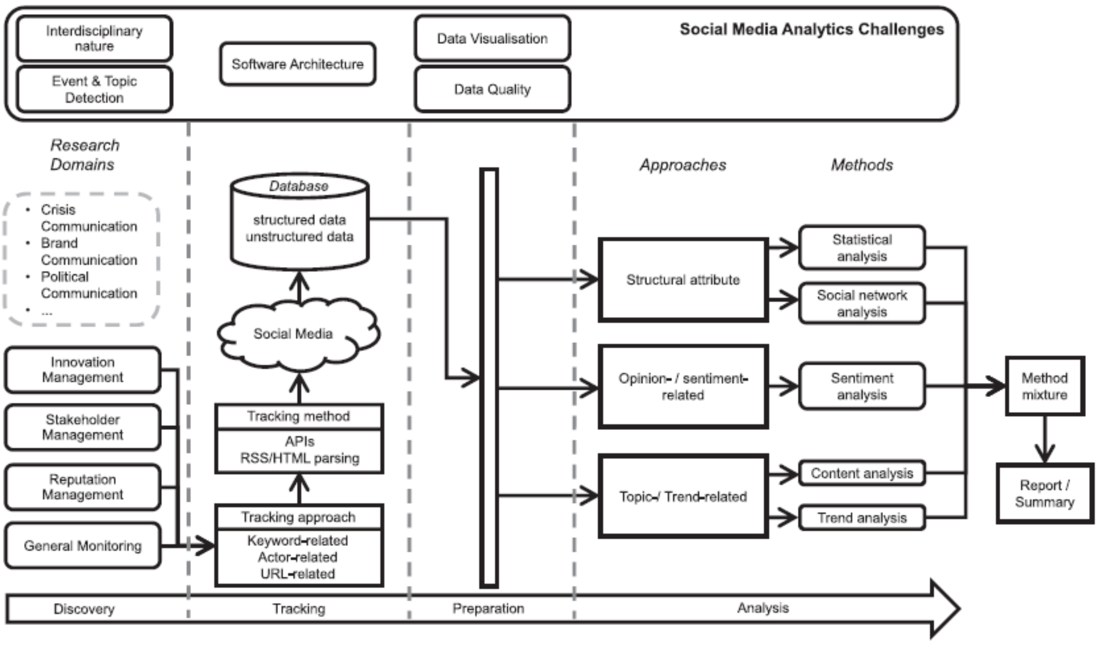


Figure 4.2*: Social media analytics framework* (Stieglitz, et al., 2018)

4.4 Proposed Approach Overview

CRISP-DM methodology is widely used according to the poll result in ***Table 4.1*** and social media analytics framework proposed by Stieglitz (2018) in ***Figure 4.2*** is suitable for the case of social media. However, CRISP-DM is relatively general and social media analytics framework proposed by Stieglitz (2018) is not very specific on this project problem as the data analysis in this project only focus on opinion / sentiment-related approach instead of structural attribute approach and topic / trend-related approach. Besides, tracking approach is only focus on keyword-related rather than actor-related and URL-related. Therefore, a social media analytics framework as shown in ***Figure 4.3*** is proposed based on social media analytics framework (Stieglitz, et al., 2018) after fine tuning and CRISP-DM (Wirth & Hipp, n.d.) to specific on the project problem. The social media analytics framework proposed is divided into the following subsections to provide further explanation.

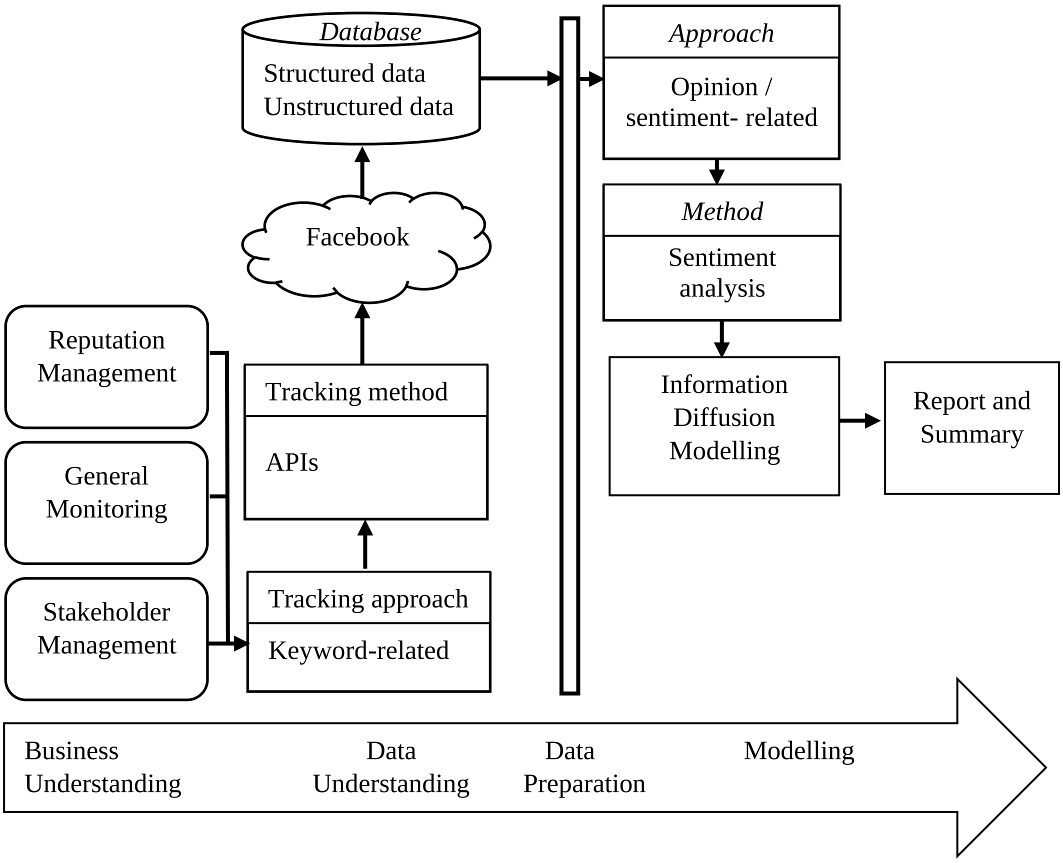


***Figure 4.3**: Approach which adopted from CRISP-DM and Social media analytics framework.*

4.5 Business Understanding

The research is initiated with business understanding. The main business objective identified is branding Malaysia's university. Universities and colleges around the world have started a quest for what people perceived them as. They are doing this in order to stand out from others and be the extraordinary one to appeal intake students and staffs (Chapleo, 2004; Hemsley-Brown & Goonawardana, 2007). Tertiary education are now alert of the connection between what they "represent" regarding qualities and attributes, and how they are seen due to the jargons that are emerging such as branding, identity and reputation (Waeraas & N.Solbakk, 2009). Branding is important for the university to stand out among other universities in the market and to provide the characteristics of brand distinctiveness to examine the perceived brand personality of the university by consumers.

The branding has an impact on the university's reputation and to have a positive influence on university ranking. These rankings have become an industry worth millions of dollars by the publications providing the rankings and millions of dollars spent by universities trying to burnish their image and enhance their position in these rankings (L. Bunzel, 2007). Consumers pay attention to better brands inaugurated by these rankings. The better brands gain in quality of student and raise the overall academic standing of a university (L. Bunzel, 2007).

The organizational structure is determined through stakeholder management. Reputation management is involved to study how people are talking about the university such as trustworthiness. General monitoring of social media post in Malaysia's university is computing the grade of impact the university possessing in the social media. The success criteria are measured by if there is a rise in the influence of Malaysia's university in social media. The business knowledge is then converted to a data mining problem definition and a preliminary project plan is designed to accomplish the objectives.

4.6 Data Understanding

This stage starts with data collection and get acquainted with the data identifying the data quality problem to discover insights into the data. Some intriguing subsets can be distinguished to form hypothesis for hidden information. After this stage, data is fully understood and described. The close link between Data Understanding and Business Understanding refers to the formulation of data mining problem and the project plan require at least some understanding of the available data.

4.6.1 Data Sampling and Collection

The data of 6 university brands that were actively posting content on Facebook from January 17th, 2020 to February 17th, 2020 is collected. The universities involved in this research comprises of University of Tunku Abdul Rahman (UTAR), University of Malaya (UM), National University of Malaysia (UKM), Universiti Teknologi Petronas (UTP), University of Science, Malaysia (USM), Universiti Malaysia Perlis (UniMAP). The number of likes, comments and shares on a brand post, as well as sentiment of brand post, hashtag use, and other brand post feature are collected through a total of 320 brand posts. The sample rows before data cleaning are shown in ***Figure 4.4***.

	University	Text
0	University of Malaya (UM)	Media statement - in connection with media rep...
1	University of Malaya (UM)	Dear Campus Community, Let's get to know our l...
2	University of Malaya (UM)	13th residence college of university of malaya...
3	University of Malaya (UM)	Appointment of ytm tunku Abdul Rahman As Chanc...
4	University of Malaya (UM)	Dear Campus Community, Wishing you a year of p...
5	University of Malaya (UM)	Congratulations to Dr. Iskandar Bin Abdullah f...
6	University of Malaya (UM)	Vice Chancellor new year message ceremony 2020...
7	University of Tunku Abdul Rahman (UTAR)	Register Now!\n\nExplore UTAR from wherever yo...
8	University of Tunku Abdul Rahman (UTAR)	Technical Seminar and Workshops Series by the ...
9	University of Tunku Abdul Rahman (UTAR)	MOA Signing for Work-based Learning Education ...

Figure 4.4*: First 10 rows of brand post content data.*

4.7 Data Preparation

Data cleaning is done at this stage. The regular expression such as '\br' is replaced with one blank space and the symbols and punctuation are removed for further analysis. Top 10 rows of data after data cleaning is shown in ***Figure 4.5***.

	University	Text
0	University of Malaya (UM)	Media statement - in connection with media rep...
1	University of Malaya (UM)	Dear Campus Community, Let's get to know our l...
2	University of Malaya (UM)	13th residence college of university of malaya...
3	University of Malaya (UM)	Appointment of ytm tunku Abdul Rahman As Chanc...
4	University of Malaya (UM)	Dear Campus Community, Wishing you a year of p...
5	University of Malaya (UM)	Congratulations to Dr. Iskandar Bin Abdullah f...
6	University of Malaya (UM)	Vice Chancellor new year message ceremony 2020...
7	University of Tunku Abdul Rahman (UTAR)	Register Now! Explore UTAR from wherever you a...
8	University of Tunku Abdul Rahman (UTAR)	Technical Seminar and Workshops Series by the ...
9	University of Tunku Abdul Rahman (UTAR)	MOA Signing for Work-based Learning Education ...

Figure 4.5*: Screenshot of data after cleaning.*

Tools to use

The software that has been chosen for developing this research are:

1. Anaconda Navigator

   Anaconda Navigator is a desktop application that launch applications and easily manage conda packages, environments, and channels without using command-line commands. It is available for Windows, macOS, and Linux. The following applications are available by default in Navigator:

   - RStudio
   - Anaconda Prompt (Windows only)
   - Anaconda PowerShell (Windows only)
   - Spyder
   - PyCharm
   - JupyterLab
   - Jupyter Notebook
   - VSCode
   - Glueviz
   - Orange 3 App

2. Jupyter Notebook

   The Jupyter Notebook is an open-source web application that allows user to create and share documents that contain live code include data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning. Jupyter supports over 40 programming languages, including Python, R, Julia, and Scala.

3. Pandas

   Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

4. Re

   A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

4.8 Modelling

In this research, a conceptual model is proposed to achieve data mining goals. Sentiment value is analysed using the text attribute to see is there any relationship between the sentiment value and information diffusion. In addition, one parameter is added which is use of hashtag in order to explore does the post contains hashtag get more popular. Besides, there are 4 parameters tested in Twitter (de Vries, et al., 2012) and Sina Weibo (WeiXian, 2016) in existing literature are used in this research as well to see does it work in Facebook. The measurement of 6 variables will be presented in ***Table 4.2***. Few indicators are selected to measure information diffusion which are: number of likes, number of comments, number of shares.

4.8.1 Operationalization of the Variables

In this research, brand post popularity is measured by the number of likes, the number of comments and the number of shares on a brand post. There are four levels (no, low, medium, high) for both vividness and interactivity which defined by previous study (Coyle and Thorson 2001; Fortin and Dholakia 2005) is illustrated in ***Table 4.2***.

Table 4.2*: Operationalizations of Vivid and Interactive Brand Post Characteristics (Coyle and Thorson 2001; Fortin and Dholakia 2005).*

Level	Vividness	Interactivity
Low	Pictorial (photo or image)	1. Link to a website (mainly to news sites or blogs, but never to the company website) 2. Voting (brand fans are able to vote for alternatives (e.g., which taste or design they think is best))
Medium	Event (application at the brand page and announces an upcoming (offline) event of the brand)	1. Call to act (urges brand fans to do something (e.g., go to certain website, liking, or commenting) 2. Contest (brand fans are requested to do something (e.g., Tweet or like a website) for which they can win prizes)

High	Video (mainly videos from YouTube)	1. Question 2. Quiz (similar to question, but now brand fans can win prizes)

Regarding the content of brand post, according to the definitions of informational content and entertaining content mentioned in ***Section 3.2***, university announcement, student affairs, conference, competition is considered informational. Contrary, entertaining content is not important announcement, fun, exciting such as anecdotes. The posts which are void of content above are considered neither informational nor entertaining. The parameter - use of hashtag is a binary variable, it is considered as yes (1) if there is a prefix word with a "#" symbol, else it is a no (0).

In the meantime, sentiment value of each post is getting through sentiment analysis library in Python. Social media usage is growing rapidly day by day. People often share their thoughts or feelings through posting on social media platform. Opinion is important to university management when comes to decision making. The vibe of predominant slant (constructive or adverse sentiments) or suppositions given by individuals about the college is important to the board of university. Individuals' conclusions as far as perspectives, demeanours, evaluations and feelings towards substances, occasions and their characteristics have been gaining attention for sentiment assessment or opinion mining in a progressively careful manner. In recent years, sentiment analysis or opinion mining has emerged as a distinct method to study people's opinions in terms of views, attitudes, appraisals and emotions towards entities, events and their attributes in a more thorough way (Liu 2010; Pang and Lee 2008). For labelling another 5 variables, numbers are used to indicate each level as shown in Table. A sample data after labelling is shown in ***Figure 4.6***.

Table 4.3*: Labelling variable with number.*

Vividness	1: No Vividness 2: Low Level of Vividness 3: Medium Level of Vividness 4: High Level of Vividness
Interactivity	1: No Interactivity

	2: Low Level of Interactivity 3: Medium Level of Interactivity 4: High Level of Interactivity
Informational Content	0: Not Informational Content 1: Informational Content
Entertaining Content	0: Not Entertaining Content 1: Entertaining Content
Use of Hashtag	0: Not contain keyword prefixed by “#” symbol 1: Contains keyword prefixed by “#” symbol

	University	Text	Vividness	Interactivity	Informational Content	Entertaining Content	Hashtag	Sentiment	Like	Comment	Share
0	University of Malaya (UM)	Media statement - in connection with media rep...	2	1	1	0	0	0.000000	256	142	166
1	University of Malaya (UM)	Dear Campus Community, Let's get to know our l...	2	1	1	0	0	0.656250	212	5	78
2	University of Malaya (UM)	13th residence college of university of malaya...	2	1	1	0	0	0.034028	549	26	47
3	University of Malaya (UM)	Appointment of ytm tunku Abdul Rahman As Chanc...	2	1	1	0	0	0.084091	220	0	18
4	University of Malaya (UM)	Dear Campus Community, Wishing you a year of p...	2	1	0	0	0	0.722917	273	18	15
5	University of Malaya (UM)	Congratulations to Dr. Iskandar Bin Abdullah f...	2	1	1	0	0	0.246667	712	113	19
6	University of Malaya (UM)	Vice Chancellor new year message ceremony 2020...	4	1	1	0	0	0.136364	393	41	91
7	University of Tunku Abdul Rahman (UTAR)	Register Now! Explore UTAR from wherever you a...	3	3	1	0	0	0.085227	69	0	5
8	University of Tunku Abdul Rahman (UTAR)	Technical Seminar and Workshops Series by the ...	3	1	0	0	1	-0.040000	6	0	1
9	University of Tunku Abdul Rahman (UTAR)	MOA Signing for Work-based Learning Education ...	2	2	1	0	1	0.000000	23	0	2

Figure 4.6*: Sample data after labelled.*

Tools to use

The software that has been chosen for developing this research are:

1. Anaconda Navigator (refer to Tools to use in ***Section 4.7***)
2. Jupyter Notebook (refer to Tools to use in ***Section 4.7***)
3. NumPy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

4. TextBlob

   TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

4.9 Evaluation

Prior to continuing deployment of the model, it is essential to evaluate the model whether it achieve the success criteria of data mining and business. At this stage, the model will be evaluated thoroughly, and the steps executed to construct the model are reviewed. Evaluation is to determine the next step whether to continue to deployment phase or replacing, refining model. In this research, Pearson Correlation analysis is chosen to measure the correlation between the influential factors and the indicators.

4.9.1 Correlation Analysis

Correlation can be a proper strategy to apply to explore user behaviours; for instance, in this research, why people likes, comments on, or share a brand post correlate with the six influential factors. As per Sriram (2002), the most regular estimation of correlation in statistics is the Pearson correlation, which can uncover direct connection between two variables and the coefficient is denoted by “r” in the formulation. However, Pearson correlation is very sensitive to outliers so there is one additional correlation analysis method included in this research: spearman correlation. Spearman correlation is calculated in the way same as Pearson correlation, but utilizing rank instead of data points. It is suitable for data with ordinal values. Spearman’s correlation is same with the interpretation to that “r” of Pearson. The nearer the “r” value is to 1 or -1, the stronger the correlation. There are 3 types of correlation:

1. Positive Correlation – r is positive value; the responding variable likewise tends to increment.
2. Negative Correlation – r is negative value; the responding variable tends to diminish.

3. No Correlation – r is a value of 0; the responding variable neither increase nor decrease.

The strength of correlation is described in ***Table 4.4***. The correlation is studied between the popularity of Malaysia's posts on Facebook and the six influential factors which include vividness, interactivity, informational content, entertaining content, sentiment of the post, and use of hashtag.

Table 4.4*:Description of the strength of the correlation.*

Value	Description of the strength of correlation
.00 - .19	Very Weak
.20 - .39	Weak
.40 - .59	Moderate
.60 - .79	Strong
.79 - 1.0	Very Strong

Tools to use

The software that has been chosen for developing this research are:

1. Anaconda Navigator (refer to Tools to use in ***Section 4.7***)
2. Jupyter Notebook (refer to Tools to use in ***Section 4.7***)
3. SciPy.stats

   SciPy.stats is a package of SciPy in Python. This module contains a large number of probability distributions as well as a growing library of statistical functions.

4.10 Deployment

The results from the models and findings are summarized in ***CHAPTER 5***.

4.11 Implementation Issues and Challenges

As more and more rich social media, popular online social networking sites, and various kinds of social network analysing and mining techniques are available, privacy in social networks becomes a serious concern (Backstrom et al, 2007; Kleinberg, 2007; Srivastava et al, 2008). Due to the privacy concern, Facebook's developer documentation states that "while developer are testing his or her app and before submitting it for review, the app can access content only on a Page for which the following is true: The person who holds the admin role for the Page also holds a role in the app as app admin, developer, or tester. If the developer wants the app to be able to access public content on other Pages, developer must submit this feature for review." In order to get the access, it required business official document and additional contract signing. It is not allowed for academic research purpose.

4.12 Timeline

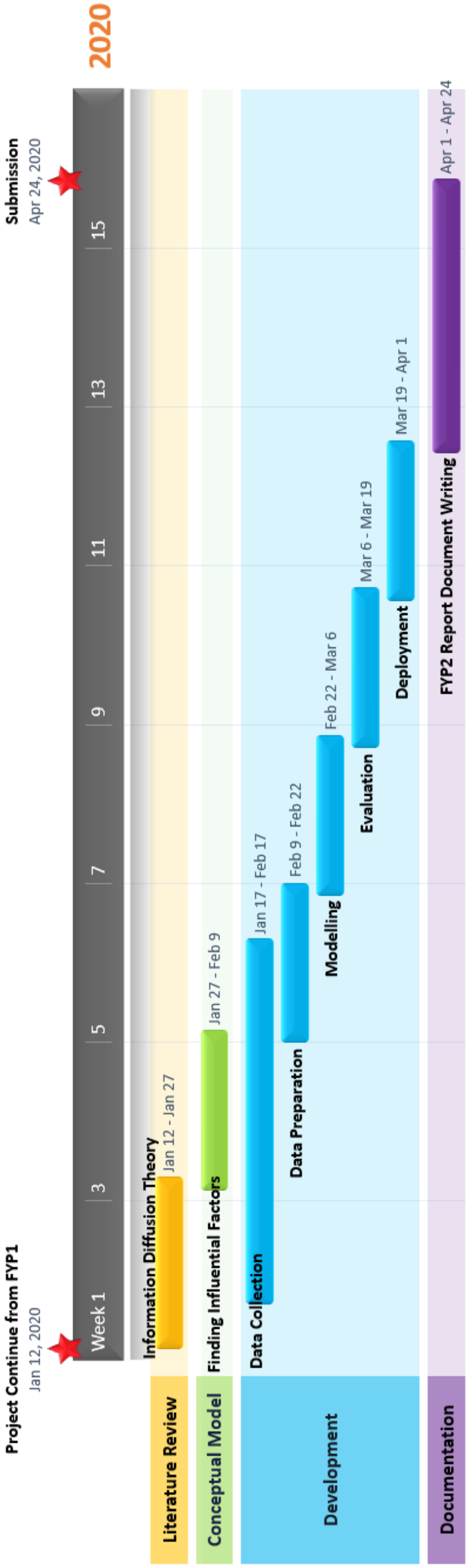


***Figure 4.7**: Project timeline for this research.*

CHAPTER 5 RESULT AND ANALYSIS

This chapter covers some findings of the research.

5.1 Distribution of the Collected Data

The average the number of Facebook post per brand is 53.333 (SD=33.866) while the number of Facebook follower of all university brand is 183,000 per brand (SD=105,046). The follower count and number of post from respective university are shown in ***Table 5.1***. On average, there are 98.994 likes per brand post (SD= 187.806); 10.678 comments per brand post (SD=50.952); 18.725 shares per brand post (SD=60.382). ***Table 5.2*** has shown mean of likes, comments and shares per university brand.

Table 5.1*: Number of Post and Number of Follower per Brand.*

University	N	Follower Count (,000)
UTAR	65	49
UTP	29	289
UM	7	251
UKM	84	273
UniMAP	104	29
USM	31	207

Table 5.2*: Average number of Likes, Comments and Shares of posts per University.*

University	Likes		Comments		Shares	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
UTAR	26.523	61.487	17.400	98.012	9.385	42.455
UTP	36.414	26.914	1.207	2.369	2.552	4.882
UM	373.571	177.004	49.286	51.597	62.000	51.597
UKM	84.429	94.182	5.155	14.501	20.119	52.194
UniMAP	46.452	72.118	1.894	8.179	5.135	10.341
USM	463.226	364.314	41.161	58.900	85.484	139.918

5.2 Vividness and Information Diffusion

Frequency analysis is conducted on the levels of vividness. How different levels of vividness perform on liking, commenting and sharing is discussed and correlation is computed as well to study the relationship between vividness and information diffusion. The analysis result is presented in the following subsections.

5.2.1 The Frequency of Levels of Vividness

In order to have an overall impression on university's performance on brand post's vividness, a frequency analysis of different levels of vividness is firstly conducted. According to ***Table 5.3***, the most frequent one is low level of vividness which show that most of the brand post contains images which occupy 68% of total posts. There is only 5 of the brand post which occupy only 1% of total posts does not include at least an image to attract the social media user. The second highest frequency (74) category is medium level which announces an upcoming offline event. There is only 6% of total posts are belongs to high level vividness include a video in the brand post. ***Figure 5.1*** has shown that majority of university's posts in this sample are low in vividness. They only post with only one or more pictures.

Table 5.3*: The Frequency of Different Levels of Vividness.*

Vividness Level	Frequency	Percentage	Cumulative Percentage
No	5	1.5625	1.5625
Low	219	68.4375	70
Medium	74	23.125	93.125
High	22	6.875	100
Total	320	100.0	

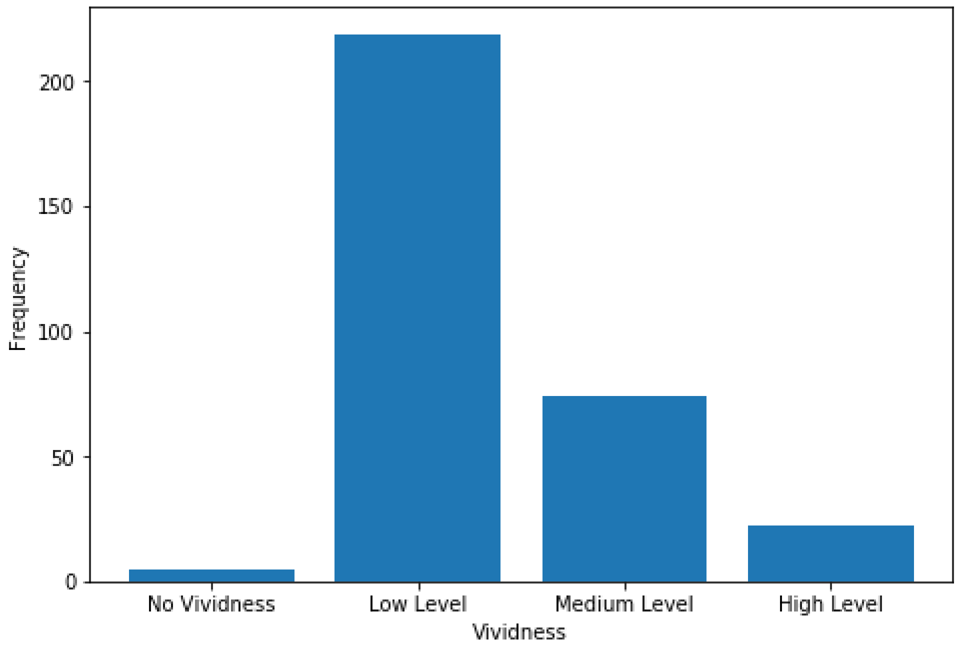


Figure 5.1: *The frequency of different levels of vividness.*

5.2.2 Performances of Different Levels of Vividness on Liking, Commenting and Sharing

In order to study how each level of vividness reacting on number of likes, comments and shares, the mean and standard deviation of likes, comments and shares are computed and shown in ***Table 5.4***. The result has shown that the post which has no vividness has the lowest mean in the number of likes among four levels of vividness. Meanwhile, the post which contains video element that high in vividness has the highest mean in the number of likes and number of shares among four levels of vividness as video is an attractive element. Besides, the category of low vividness has the lowest mean value in the number of comments while the highest mean value in the number of comments is the group of no vividness. This may because image presentation is very clear and easy to understand which less invoke comments. The mean value of number of comments are very close among the level of no vividness, medium vividness and high vividness. In summary, it is observed that high vividness helped in boosting the number of likes and number of shares.

Table 5.4: *Average number of likes, comments and shares of different levels of vividness's posts.*

Vividness	Likes	Comments	Shares

	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
No Vividness	66.800	74.799	20.800	41.102	11.800	14.260
Low Vividness	91.374	165.094	6.822	35.194	10.466	29.669
Medium Vividness	88.081	215.582	18.959	84.126	27.122	90.886
High Vividness	218.864	261.789	18.909	27.791	74.273	111.476

5.2.3 The Correlation between Vividness and the Popularity

It is not enough to judge the correlation between the vividness and popularity by the value of mean and standard deviation. Scatter plot is used to plot data points on a horizontal and a vertical axis. Three visualizations of scatter plot, ***Figure 5.2*** to ***Figure 5.4***are plotted to show how much one variable is affected by another. In ***Figure 5.2***, there is one data point having over 1400 likes that is situated away from other data points. Meanwhile in ***Figure 5.3*** and ***Figure 5.4***, there are few data points which recorded over 400 comments and shares differing from other data points.

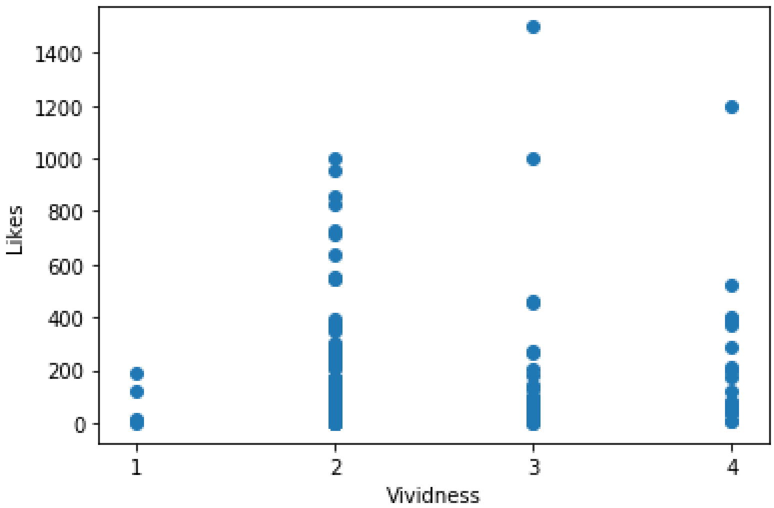


Figure 5.2*: Scatter plot of vividness and likes.*

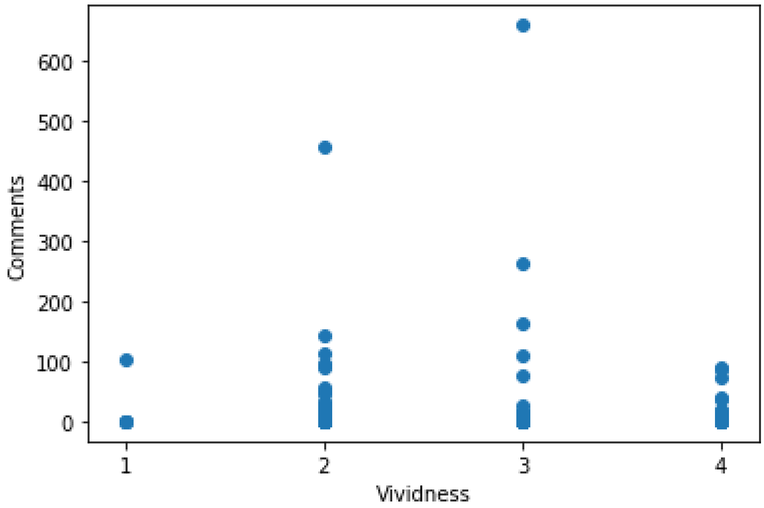


Figure 5.3*: Scatter plot of vividness and comments.*

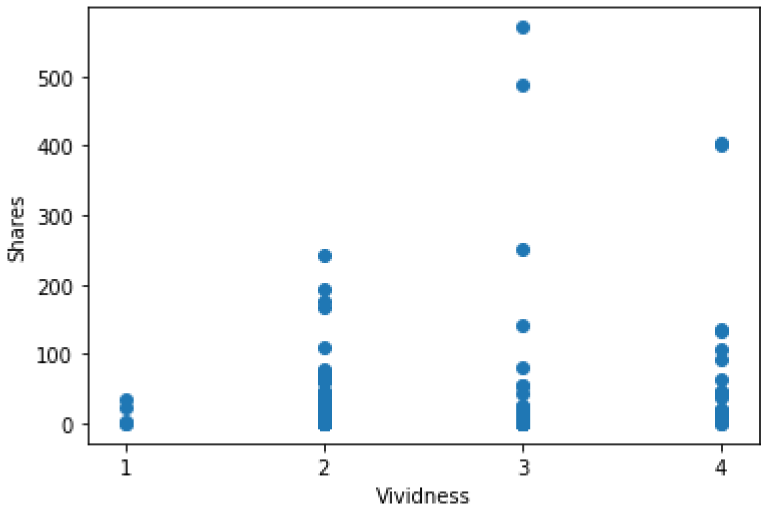


Figure 5.4*: Scatter plot of vividness and shares.*

Besides, Pearson correlation efficient and spearman correlation efficient are computed and shown in ***Table 5.5***. It is found that Pearson correlation coefficient and spearman correlation efficient are larger than 0 indicating a positive relationship but it is very weak in the category of like and comment, meanwhile it is slightly higher in the category of share.

Table 5.5*: The correlation between vividness and popularity.*

	Like	Comment	Share

Pearson Correlation	0.122	0.090	0.255
Spearman Correlation	0.093	0.137	0.252

5.3 Interactivity and Information Diffusion

Frequency analysis is conducted on the levels of interactivity. How different levels of interactivity perform on liking, commenting and sharing is discussed and correlation is computed as well to study the relationship between interactivity and information diffusion. The analysis result is presented in the following subsections.

5.3.1 The Frequency of Levels of Interactivity

According to ***Table 5.6***, the post with no interactivity stand out the most, which occupied 78.75 percent of all posts followed by low interactivity that have 37 posts which include link to a website. There are only 4 posts with highest interactivity which posted question or quiz to seek answer from users. The frequency of different levels of interactivity is plotted in a bar graph and displayed in ***Figure 5.5***.

Table 5.6*: The Frequency of Different Levels of Interactivity.*

Interactivity Level	Frequency	Percentage	Cumulative Percentage
No	252	78.75	78.75
Low	37	11.5625	90.3125
Medium	27	8.4375	98.75
High	4	1.25	100
Total	320	100.0	

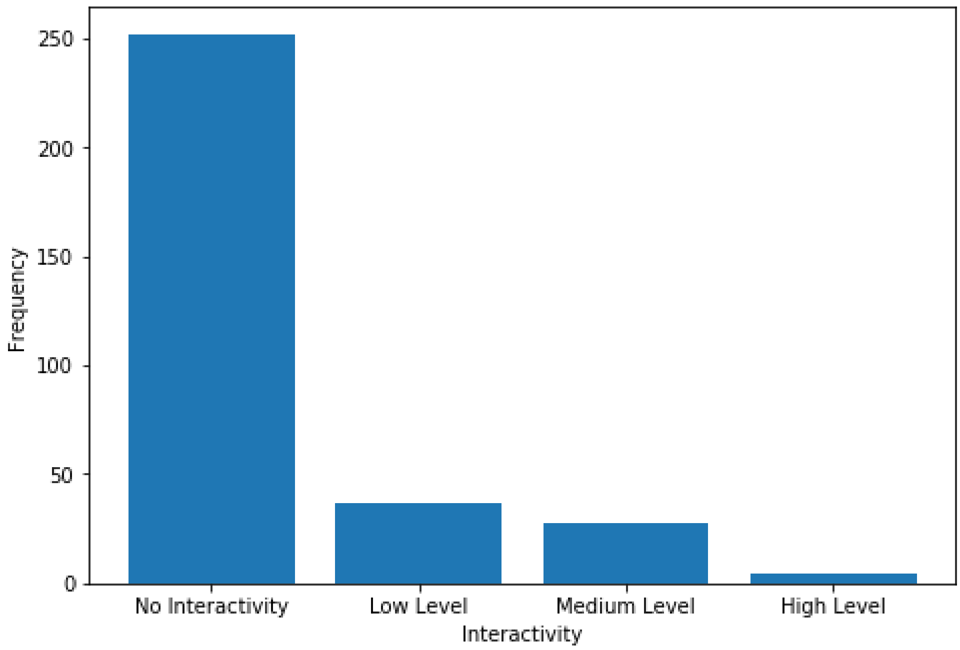


Figure 5.5: *The frequency of different levels of interactivity.*

5.3.2 Performances of Different Levels of Interactivity on Liking, Commenting and Sharing

The mean and standard deviation on number of likes, comments and shares according to different levels of interactivity are computed and illustrated in ***Table 5.7***. High interactivity post performed the best among other levels as it has the highest mean value on amount of likes, comments and shares. Meanwhile, the medium interactivity level has the lowest mean value in likes, comments and shares.

Table 5.7: *Average number of likes, comments and shares of different levels of interactivity's posts.*

Interactivity	Likes		Comments		Shares	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
No Interactivity	99.329	198.019	11.044	55.718	16.254	55.242
Low Interactivity	95.811	163.275	6.486	26.697	35.189	98.406

Medium Interactivity	80.444	94.872	4.037	6.161	15.333	22.665
High Interactivity	232.500	54.225	71.250	33.409	45.000	52.187

5.3.3 The Correlation between Interactivity and the Popularity

Three visualizations of scatter plot which plotted likes, comments and shares according to levels of interactivity is shown in ***Figure 5.6***, ***Figure 5.7*** and ***Figure 5.8*** respectively. There are few data points having over 1200 likes, 400 comments and 500 shares differing from other data points.

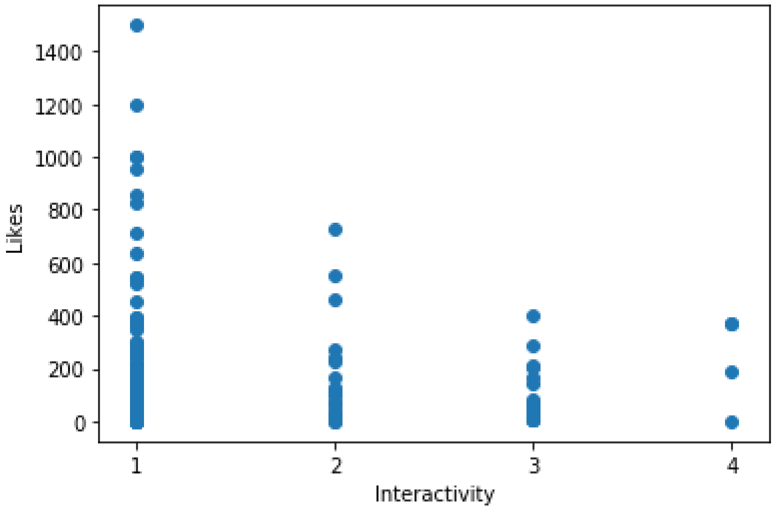


Figure 5.6*: Scatter plot of interactivity and likes.*

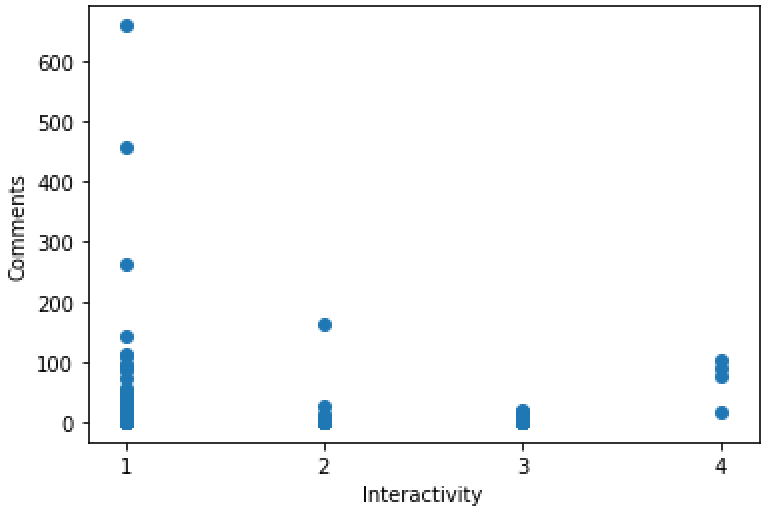


Figure 5.7: *Scatter plot of interactivity and comments.*

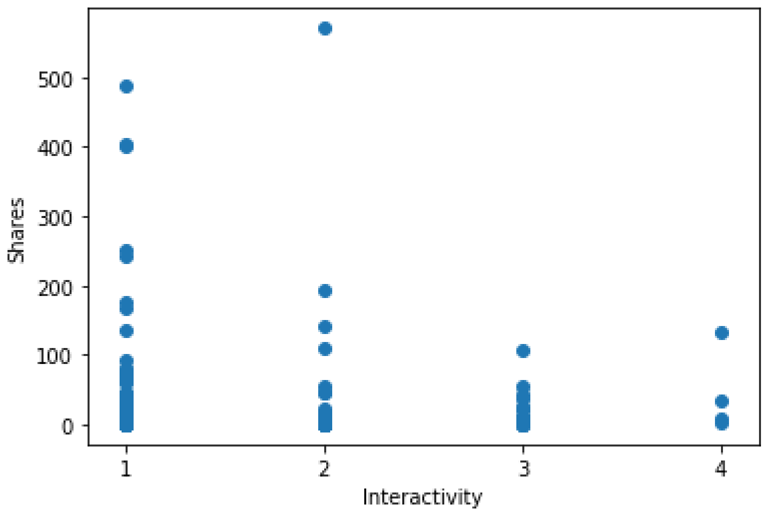


Figure 5.8: *Scatter plot of interactivity and shares.*

From ***Table 5.8***, it is shown that interactivity has no relationship with the number of likes as the coefficient is around zero. Meanwhile, the spearman correlation coefficients of comment and share are larger than 0.01. The result indicates that there is a weak strength between interactivity and number of comments, shares.

Table 5.8: *The correlation between interactivity and popularity.*

	Like	Comment	Share
Pearson Correlation	0.012	0.019	0.056
Spearman Correlation	0.016	0.112	0.194

5.4 Informational Content and Information Diffusion

Frequency analysis is conducted on the informational and non-informational content. How informational and non-informational content perform on liking, commenting and sharing is discussed and correlation is computed as well to study the relationship between informational content and information diffusion. The analysis result is presented in the following subsections.

5.4.1 The Frequency of Informational Content and Non-Informational Content

According to ***Table 5.9***, informational content has occupied 60 percent of total posts while non-informational content has 127 posts (40 percent). The frequency informational and non-informational posts is plotted in a bar graph and displayed in ***Figure 5.9***.

Table 5.9: *The Frequency of informational and non-informational posts.*

Informational Content	Frequency	Percentage	Cumulative Percentage
No	127	39.6875	39.6875
Yes	193	60.3125	100
Total	320	100.0	

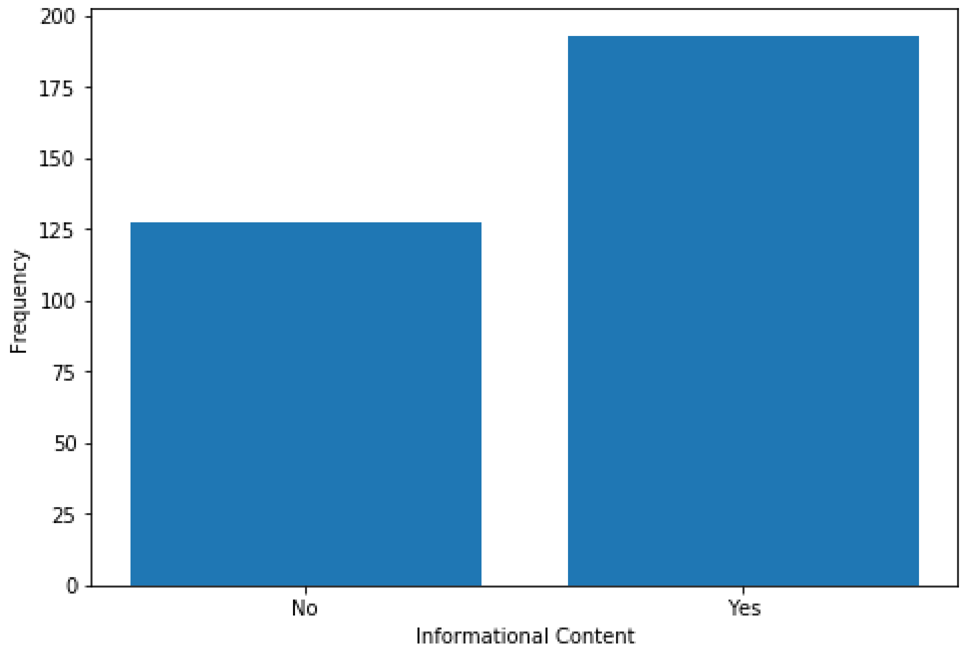


Figure 5.9: *The frequency of informational and non-informational posts.*

5.4.2 Performances of Informational Content and Non-Informational Content on Liking, Commenting and Sharing

The informational content performed better through the mean value of likes, comments and shares compared to non-informational content. ***Table 5.10*** has shown the average number of likes, comments and shares of informational and non-informational posts.

***Table 5.10**: Average number of likes, comments and shares of informational and non-informational posts.*

Informational Content	Likes		Comments		Shares	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
No	89.339	199.437	5.346	19.033	10.472	40.091
Yes	105.347	179.459	14.187	63.521	24.155	70.094

5.4.3 The Correlation between Informational Content and the Popularity

Scatter plots of likes, comments and shares according to informational and non-informational content are shown below. There are few data point which has over 1200 likes in ***Figure 5.10***, 300 comments in ***Figure 5.11*** and 300 shares in ***Figure 5.12*** lies outside the overall data points pattern.

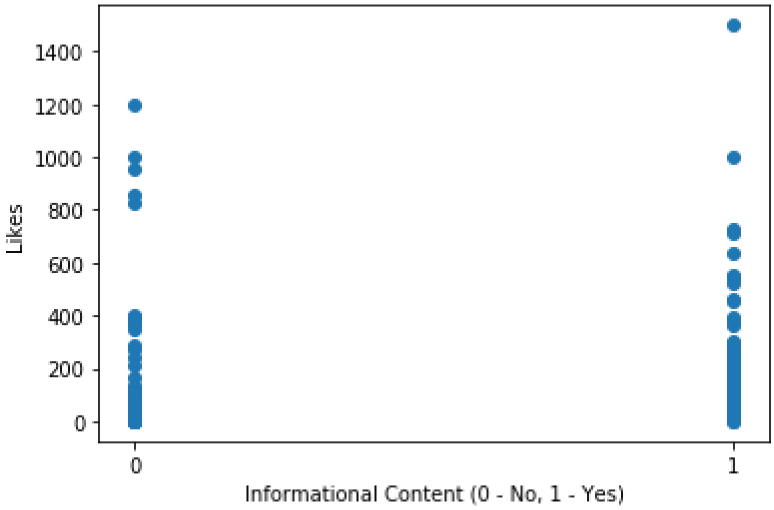


***Figure 5.10**: Scatter plot of informational content and likes.*

Figure 5.11: *Scatter plot of informational content and comments.*

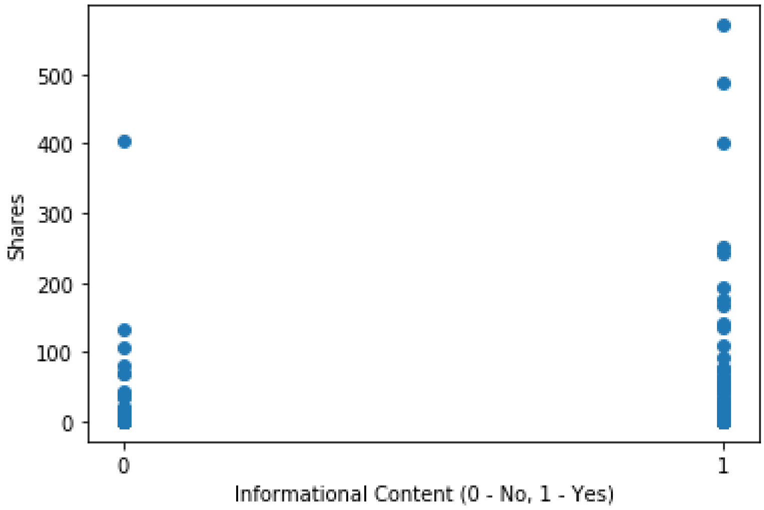


Figure 5.12: *Scatter plot of informational content and shares.*

As Pearson correlation is very sensitive to the points that lies outside the pattern of the data, it has shown an extremely low coefficient according to ***Table 5.11***. However, the value of spearman correlation coefficient has achieved more than 0.01, which is 0.208, 0.186 and 0.266 for likes, comments and shares respectively. This result has suggested that informational content get relatively higher popularity than non-informational content.

Table 5.11: *The correlation between informational content and popularity.*

	Like	Comment	Share

Pearson Correlation	0.042	0.085	0.111
Spearman Correlation	0.208	0.186	0.266

5.5 Entertaining Content and Information Diffusion

Frequency analysis is conducted on the entertaining and non- entertaining content. How entertaining and non- entertaining content perform on liking, commenting and sharing is discussed and correlation is computed as well to study the relationship between entertaining content and information diffusion. The analysis result is presented in the following subsections.

5.5.1 The Frequency of Entertaining Content and Not-Entertaining Content

According to ***Table 5.12***, not-entertaining content has stand out the most which it occupied 95 percent of total posts while entertaining content has only 5 posts. The frequency of entertaining and not-entertaining posts is plotted in a bar graph and displayed in ***Figure 5.13***.

Table 5.12*: The Frequency of entertaining and not-entertaining posts.*

Entertaining Content	Frequency	Percentage	Cumulative Percentage
No	304	95	95
Yes	16	5	100
Total	320	100	

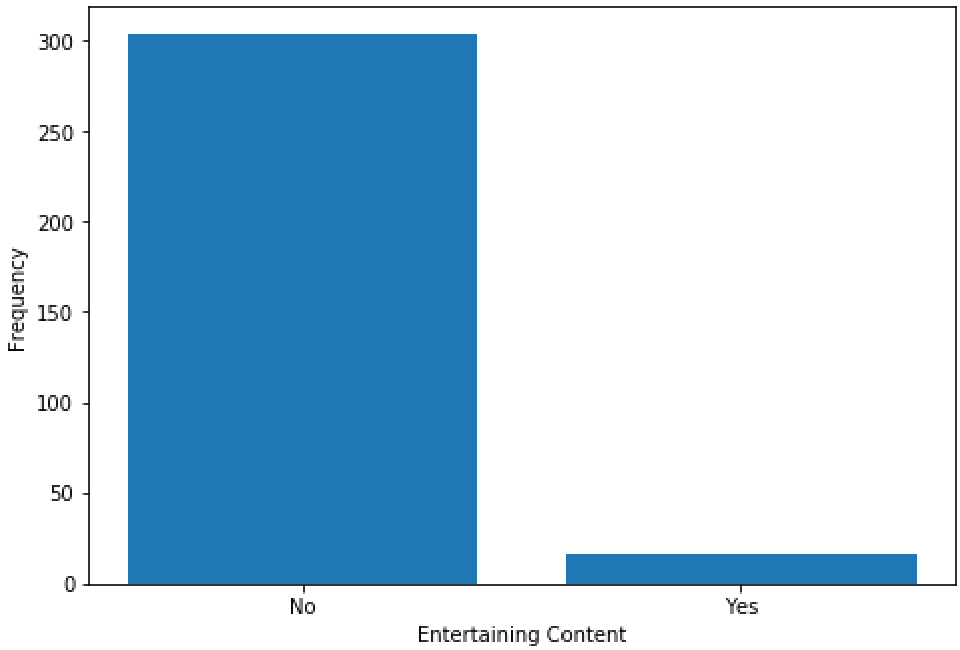


Figure 5.13*: The frequency of entertaining and not-entertaining posts.*

5.5.2 Performances of Entertaining Content and Not-Entertaining Content on Liking, Commenting and Sharing

Table 5.13 has shown the average number of likes, comments and shares of entertaining and not-entertaining posts. The entertaining content performed better through the mean value of likes, comments and shares compared to not-entertaining content.

Table 5.13*: Average number of likes, comments and shares of entertaining and not-entertaining posts.*

Entertaining Content	Likes		Comments		Shares	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
No	83.602	157.500	9.467	51.123	16.510	56.951
Yes	391.438	379.582	33.688	41.318	60.812	97.115

5.5.3 The Correlation between Entertaining Content and the Popularity

There are three visualizations of scatter plots of likes, comments and shares according to entertaining and not-entertaining posts are shown below. There are few data point which has over 1200 likes in ***Figure 5.14***, 300 comments in ***Figure 5.15*** and 300 shares in ***Figure 5.16*** lies outside the overall data points pattern.

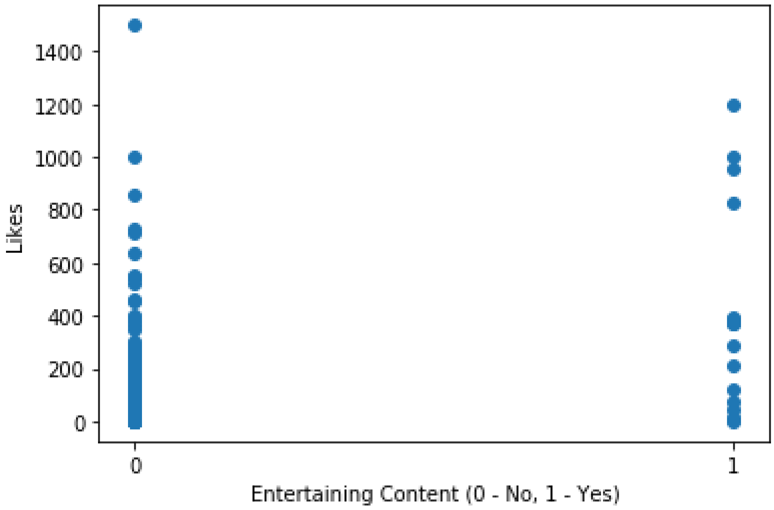


Figure 5.14*: Scatter plot of entertaining content and likes.*

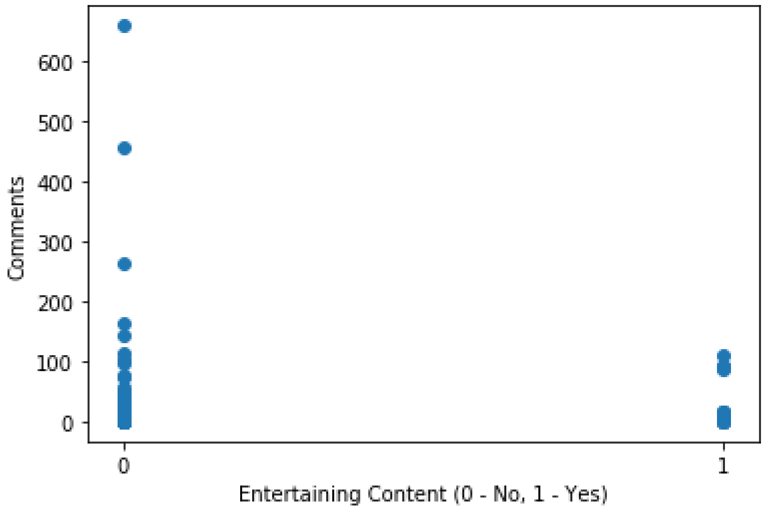


Figure 5.15*: Scatter plot of entertaining content and comments.*

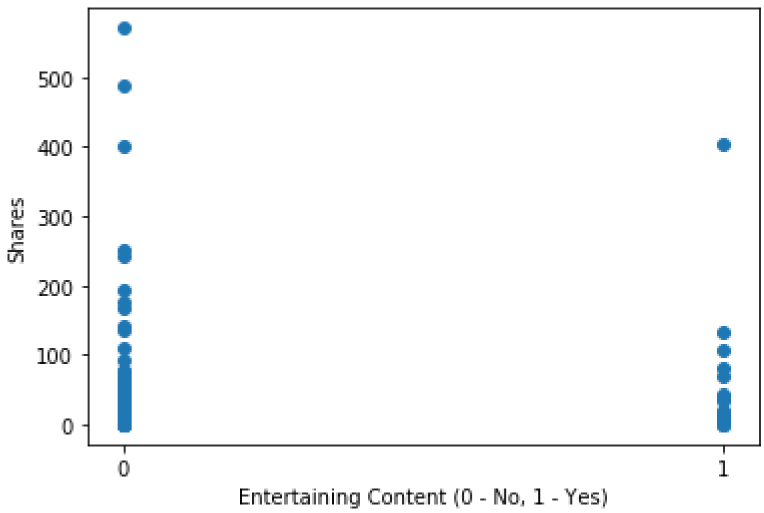


Figure 5.16*: Scatter plot of entertaining content and shares.*

According to ***Table 5.14***, both correlation's coefficients are larger than 0 in like, comment and share. Among that, Pearson correlation coefficient has reached the value of 0.357 which is the highest coefficient value among other influential factors until now. The results have proven that entertaining content boost number of likes, comments and shares.

Table 5.14*: The correlation between entertaining content and popularity.*

	Like	Comment	Share
Pearson Correlation	0.357	0.104	0.160
Spearman Correlation	0.199	0.230	0.192

5.6 Sentiment Value and Information Diffusion

Frequency analysis is conducted on different sentiments of post. How various sentiments perform on liking, commenting and sharing is discussed and correlation is computed as well to study the relationship between sentiment and information diffusion. The analysis result is presented in the following subsections.

5.6.1 The Frequency of Levels of Sentiment Value

According to ***Table 5.15***, positive content has occupied half (53 percent) of total posts while negative content has only 35 posts. The frequency of positive, negative and neutral posts is plotted in a bar graph and displayed in ***Figure 5.17***.

Table 5.15*: The Frequency of 3 levels of sentiment value.*

Sentiment	Frequency	Percentage	Cumulative Percentage
Positive	170	53.125	53.125
Neutral	115	35.9375	89.0625
Negative	35	10.9375	100
Total	320	100	

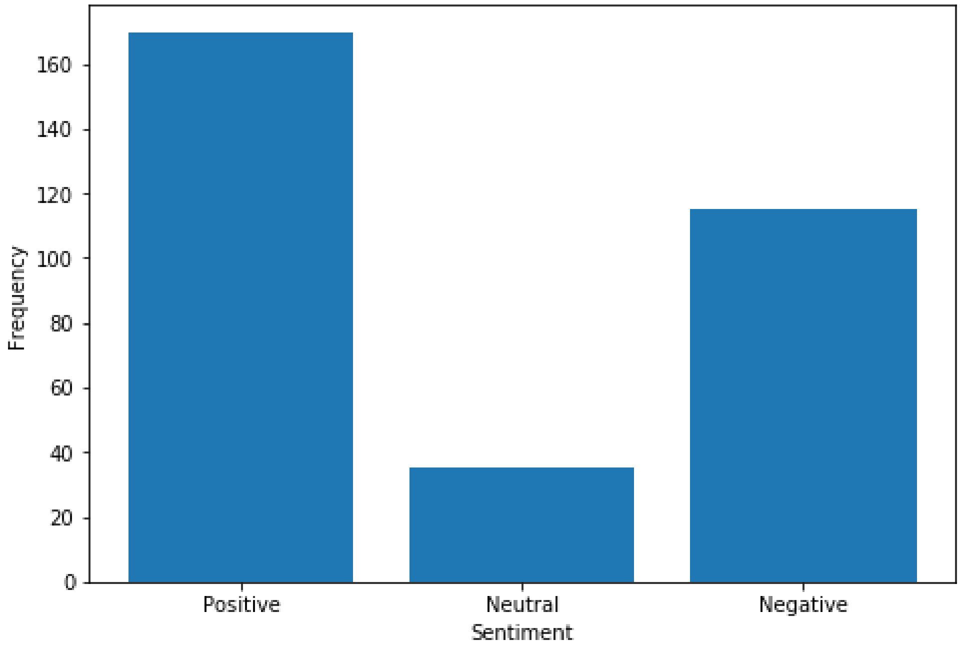


Figure 5.17*: The frequency of three levels of sentiment.*

5.6.2 Performances of Different Levels of Sentiment Value on Liking, Commenting and Sharing

Table 5.16 has shown the average number of likes, comments and shares of different sentiments. The negative post has the highest mean value of likes, comments and shares compared to positive and neutral content. Meanwhile, the sentiment having lowest mean value of likes, comments and shares is neutral.

Table 5.16*: Average number of likes, comments and shares of different sentiment of posts.*

Sentiment	Likes		Comments		Shares	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Positive	91.282	162.530	9.894	53.946	14.482	51.498
Neutral	46.371	54.191	1.800	6.028	7.829	16.106
Negative	126.409	237.740	14.539	53.582	28.313	77.404

5.6.3 The Correlation between Sentiment Value and the Popularity

Positive Posts

There are three visualizations of scatter plots of likes, comments and shares of positive posts are shown below. There are few data point which has over 600 comments in ***Figure 5.19*** and 500 shares in ***Figure 5.20*** differing from other data points.

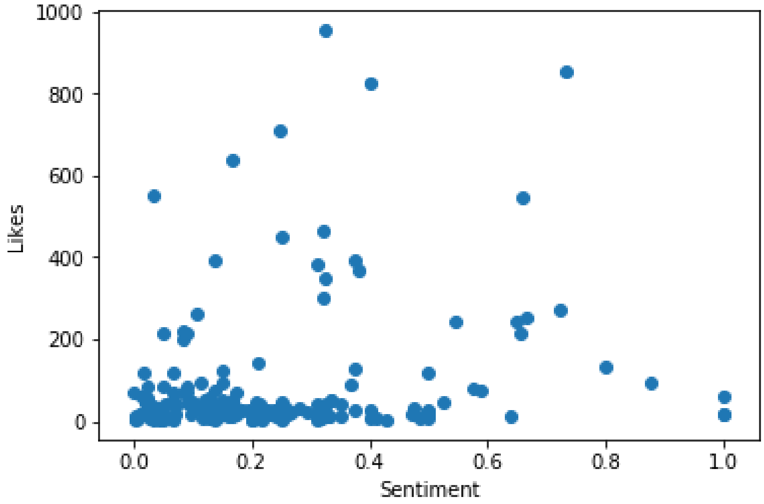


Figure 5.18*: Scatter plot of positive posts and likes.*

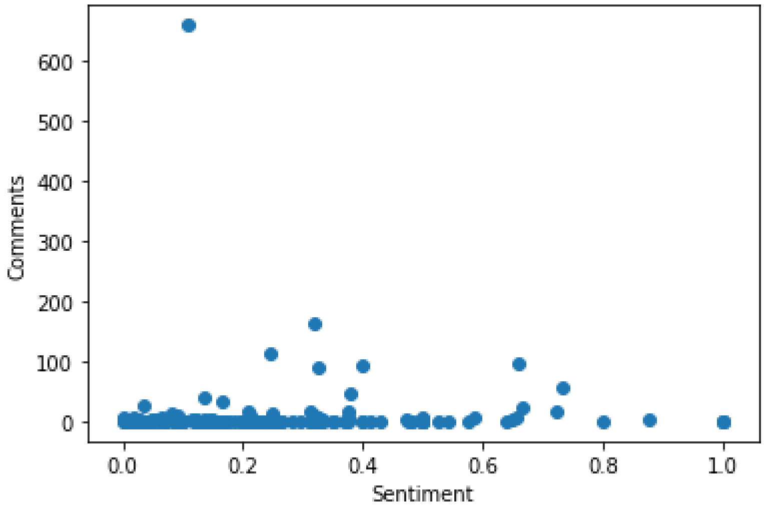


Figure 5.19*: Scatter plot of positive posts and comments.*

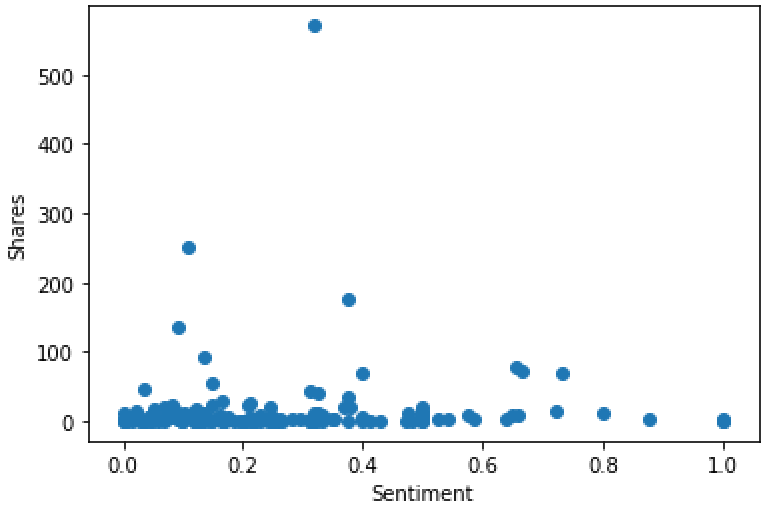


Figure 5.20: *Scatter plot of positive posts and shares.*

Pearson correlation and spearman correlation are computed and shown in ***Table 5.17***. The result has shown that positive posts is positively correlated to number of likes and a weak positive correlation to number of comments as well. However, it has no relationship between positive post and number of shares.

Table 5.17: *The correlation between positive posts and popularity.*

	Like	Comment	Share
Pearson Correlation	0.200	0.057	0.011
Spearman Correlation	0.108	0.104	0.032

Negative Posts

There are three visualizations of scatter plots of likes, comments and shares of negative posts are shown below. There are few data point which has over 30 comments in ***Figure 5.22*** and 40 shares in ***Figure 5.23*** differing from other data points.

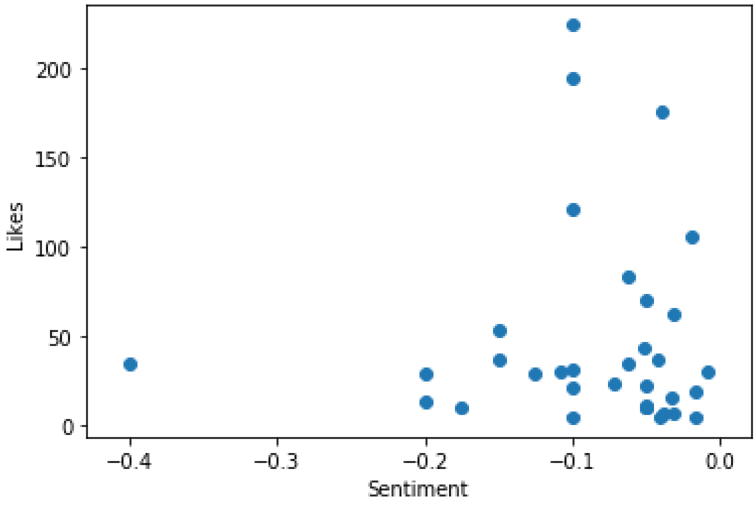


Figure 5.21: *Scatter plot of negative posts and likes.*

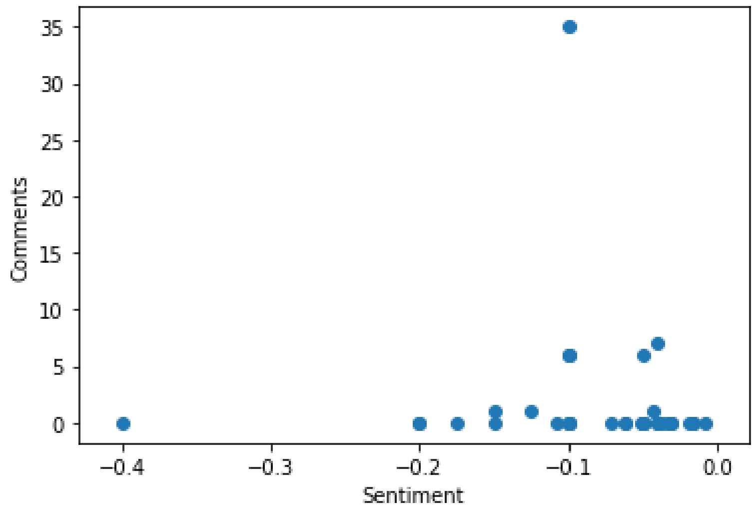


Figure 5.22: *Scatter plot of negative posts and comments.*

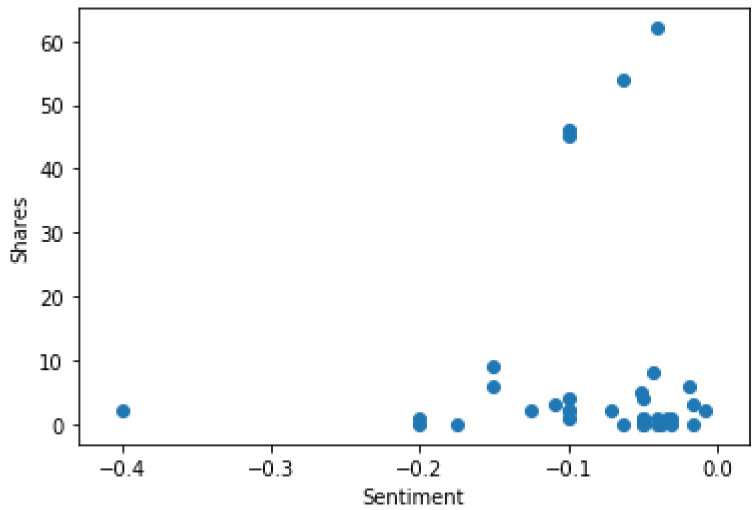


Figure 5.23*: Scatter plot of negative posts and shares.*

According to the result in ***Table 5.18***, Pearson correlation analysis has shown there is no relationship between negative post and number of likes, comments nor share. However, spearman correlation has shown a very weak negative relationship between negative posts and number of likes, comments and shares which is -0.187, -0.192 and -0.144 respectively.

Table 5.18*: The correlation between negative posts and popularity.*

	Like	Comment	Share
Pearson Correlation	-0.034	-0.079	0.010
Spearman Correlation	-0.187	-0.192	-0.144

5.7 Use of Hashtag and Information Diffusion

Frequency analysis is conducted on the posts with hashtag and posts without hashtag. How posts with hashtag and posts without hashtag perform on liking, commenting and sharing is discussed and correlation is computed as well to study the relationship between use of hashtag and information diffusion. The analysis result is presented in the following subsections.

5.7.1 The Frequency of Content with Hashtag and Content without Hashtag

According to ***Table 5.19***, there is almost two-thirds of posts include hashtag in their content. There is 94 out of 320 posts does not include a hashtag. A bar plot of frequency of use of hashtag is shown in ***Figure 5.24***.

Table 5.19*: The Frequency of posts containing hashtag and not-containing hashtag.*

Use of Hashtag	Frequency	Percentage	Cumulative Percentage
No	94	29.375	29.375
Yes	226	70.625	100
Total	320	100	

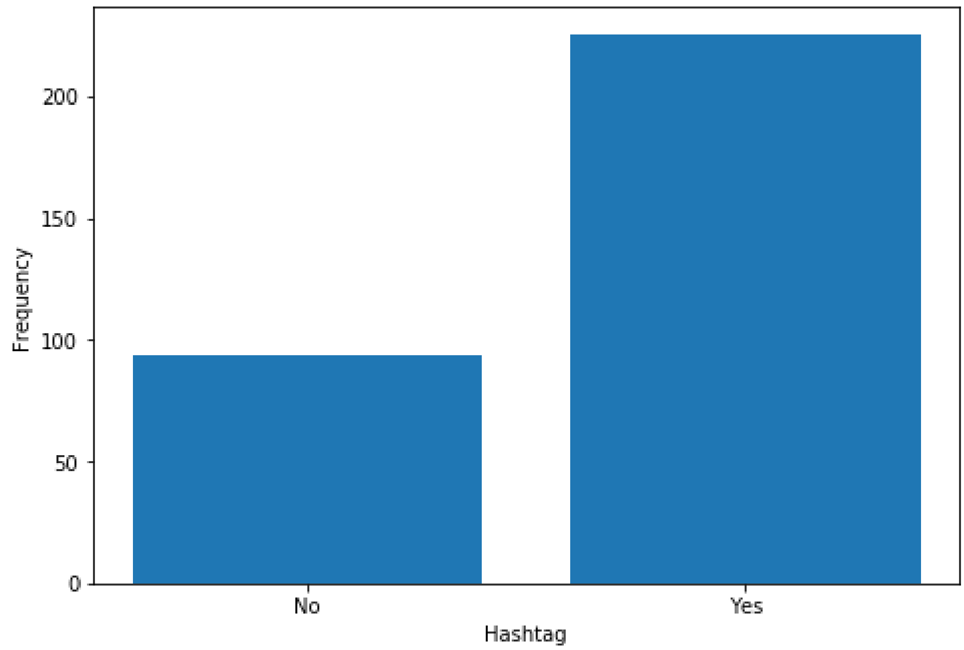


Figure 5.24*: The frequency of posts containing hashtag and not-containing hashtag.*

5.7.2 Performances of Content with Hashtag and Content without Hashtag on Liking, Commenting and Sharing

According to ***Table 5.20***, the post with hashtag included has higher mean value on total number of like than the post without hashtag. However, in the category of comment, the content without hashtag tends to be higher in average compared to the content without hashtag. Meanwhile, there is very less in the difference between mean of content that involving hashtag and mean of content that not involving hashtag.

Table 5.20*: Average number of likes, comments and shares of posts containing hashtag and not-containing hashtag.*

Use of Hashtag	Likes		Comments		Shares	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
No	73.309	157.114	19.543	85.582	20.011	70.559
Yes	109.677	198.207	6.991	24.150	18.190	55.595

5.7.3 The Correlation between Use of Hashtag and the Popularity

Scatter plot of number of likes, comments and shares according to use of hashtag are shown respectively in ***Figure 5.25***, ***Figure 5.26*** and ***Figure 5.27***. In ***Figure 5.25***, one data point which having over 1400 likes differing from other points. There are three data points in ***Figure 5.26*** which have over 200 comments are situated away from the other data points. In ***Figure 5.27***, there are three data points having over 300 shares are very much larger than the nearest data points.

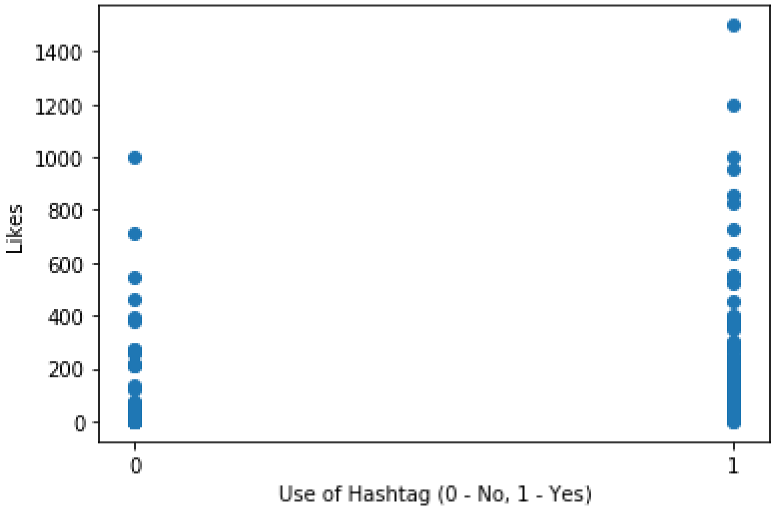


Figure 5.25*: Scatter plot of use of hashtag and likes.*

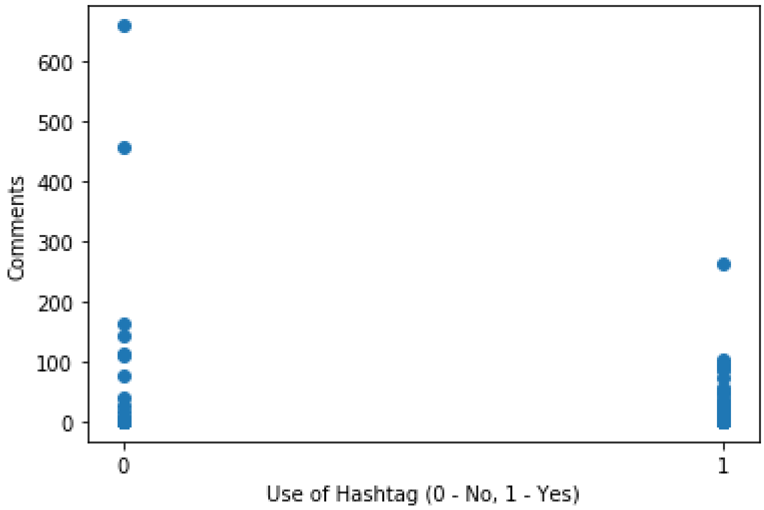


Figure 5.26*: Scatter plot of use of hashtag and comments.*

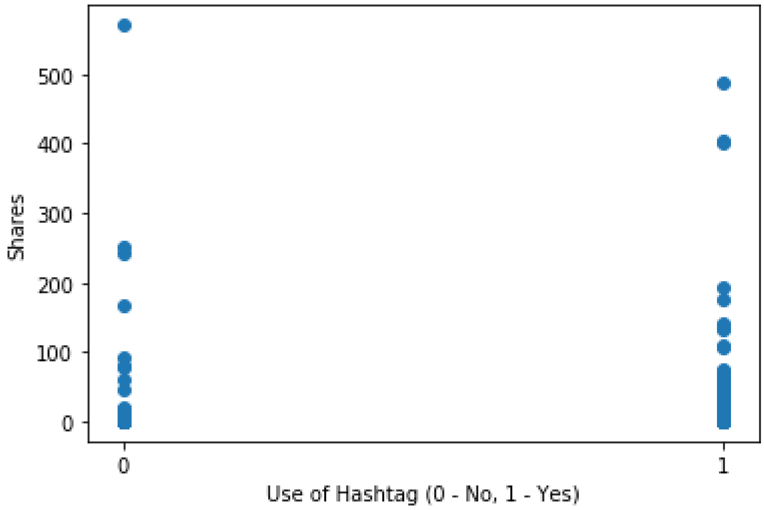


Figure 5.27*: Scatter plot of use of hashtag and shares.*

The Pearson correlation coefficient is highly sensitive to outliers. As there are many data points that lies an abnormal distance from other values, it affects the Pearson correlation analysis. The coefficient is more than 0, which mean there is positive relationship with number of likes, comments and shares. According to ***Table 5.21***, spearman correlation performed slightly better which is 0.208, 0.186, 0.266 for like, comment and share respectively. This result indicates that use of hashtag can get higher number of likes, comments and shares.

Table 5.21*: The correlation between use of hashtag and popularity.*

	Like	Comment	Share
Pearson Correlation	0.042	0.085	0.111
Spearman Correlation	0.208	0.186	0.266

5.8 Summary

A comparison table between six influential factors is shown in ***Table 5.22.*** Through the Pearson correlation analysis, entertaining content has the highest coefficient value for number of likes and comments while the vividness factor has the highest coefficient value for number of shares. In the other way, the factor of entertaining content and use of hashtag both have the highest coefficient value for the number of likes through the spearman correlation analysis. For the number of comments, spearman correlation analysis showed that the factor which has highest coefficient value is entertaining content. Lastly, informational content and use of hashtag both contribute to highest coefficient value for number of shares in the spearman correlation analysis.

Table 5.22: Comparison Table between the 6 Influential Factors.

Factor	Pearson Correlation	Spearman Correlation
Number of Likes		
Vividness	0.122	0.093
Interactivity	0.012	0.016
Informational Content	0.042	0.208
Entertaining Content	0.357	0.199
Positive Sentiment	0.200	0.108
Negative Sentiment	-0.034	-0.187
Use of Hashtag	0.042	0.208
Number of Comments		
Vividness	0.090	0.137
Interactivity	0.019	0.112
Informational Content	0.085	0.186
Entertaining Content	0.104	0.230
Positive Sentiment	0.057	0.104
Negative Sentiment	-0.079	-0.192
Use of Hashtag	0.085	0.186
Number of Shares		
Vividness	0.255	0.252
Interactivity	0.056	0.194
Informational Content	0.111	0.266
Entertaining Content	0.160	0.192
Positive Sentiment	0.011	0.032
Negative Sentiment	0.010	-0.144
Use of Hashtag	0.111	0.266

CHAPTER 6 CONCLUSION

This chapter includes conclusion, validity, limitations and future work.

6.1 Conclusion

This research has achieved its aim which is recognizing the key terms that use to search for Malaysia's tertiary education-related tweets. On the other hand, it helps branding Malaysia's education institution by identifying how people react emotionally to universities in Malaysia. It was found that people highly relate UTAR to the famous singer which performed in UTAR one year ago. Generally, people perception to Malaysia tertiary education is relatively positive and high in trustworthiness.

In the second part of the research, a conceptual framework is proposed to assess the information diffusion that happened in the domain of university. There are 6 influential factors included in the framework as well as 6 hypotheses are studied. Meanwhile, the indicators of information diffusion are number of likes, comments and shares. The result of hypotheses is shown in ***Table 6.1***.

Table 6.1: *Summary of Results.*

	Number of Likes: Coefficient Value	Number of Comments: Coefficient Value	Number of Shares: Coefficient Value
Vividness	Supported: 0.122	Supported: 0.137	Supported: 0.255
Interactivity	Not Supported: 0.016	Supported: 0.112	Supported: 0.194
Informational Content	Supported: 0.208	Supported: 0.186	Supported: 0.266
Entertaining Content	Supported: 0.357	Supported: 0.230	Supported: 0.192
Positive Sentiment	Supported: 0.200	Supported: 0.104	Not Supported: 0.032
Negative Sentiment	Supported: -0.187	Supported: -0.192	Supported: -0.144
Use of Hashtag	Supported: 0.208	Supported: 0.186	Supported: 0.266

According to ***Table 6.1***, administrator of university pages shall include some entertaining content with the use of hashtag in the post to promote certain keywords. In such way, it is believed that the number of likes will increase. The keywords to include in a post can be found in first part of this work. If administrator wish to collect opinions from user invoking comments, he or she shall make the post entertaining. Meanwhile, when the administrator would like the information posted diffuse to more users (sharing), he or she shall make the post informational with the use of hashtag so that user can search the post by certain keyword prefixed by the "#" symbol.

6.2 Validity and Generalizability

According to the Cambridge Dictionary, validity refers to the quality of being based on fact or being able to be acknowledged. It is difficult to avoid any bias in my study since it will involve subjectivity when I interpret and present the data. For instance, this may occur when labelling the samples, and operating the variables. For generalization, it may not enough to cover all university brand as this study only include 6 higher educational institution. However, this is a degree thesis and social media marketing could be extended in the future on university domain.

6.3 Limitations and Future Work

The data period limited to a month from January 19^{th} to February 19^{th}. In future, more input data is expected as there are 320 brand posts in total from 6 university in Facebook. Besides, this study focuses on one data source for information diffusion modelling which is from Facebook. For correlation analysis, it cannot fit a line through the data points (Yanai and Takane, 1992). In future, linear regression analysis may be an appropriate additional measurement to further analyse the relationship between the influential factors and information diffusion. However, some of the variables such as vividness, interactivity, informational content and entertaining content are very subjective, so it is expected to add other variables that is measurable scientifically. A deep learning approach maybe apply to select which influential factor is contributing the most to information diffusion.

BIBLIOGRAPHY

Batrinca, B. & Treleaven, P. C., 2014. Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY,* 30(1), pp. 89 - 116.

boyd, d. m. & Ellison., N. B., 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication.*

Chapleo, C., 2004. Interpretation and implementation of reputation/brand management by UK university leaders. *International Journal of Educational Advancement,* 5(1), pp. 7-23.

Chinnov, A. et al., 2015. *An Overview of Topic Discovery in Twitter Communication through Social Media Analytics.* Puerto Rico, s.n.

de Vries, L., Gensler, S. & Leeflang, P. S., 2012. Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing,* Volume 26, pp. 83 -91.

Gray, B. J., Shyan Fam, K. & Llanes, V. A., 2003. Branding universities in Asian markets. *Journal of Product and Brand Management,* 12(2), pp. 108-120.

Hemsley-Brown, J. & Goonawardana, S., 2007. Brand harmonization on the international higher education market. *Journal of Business research,* Volume 60, pp. 942-948.

Hughes, M., 2018. *Twitter to place new restrictions on its API to stop abuse.* [Online] Available at: https://thenextweb.com/twitter/2018/07/24/twitter-to-place-new-restrictions-on-its-api-to-stop-abuse/
[Accessed 07 August 2019].

Jouini, N. & Pozza, I. D., 2014. Introducing a Relationship Marketing perspective in the measurement of Online Community success. *IPAG Business School Working Paper Series*.

Kaplan, A. M. & Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons,* Volume 53, pp. 59-68.

L. Bunzel, D., 2007. Universities sell their brands. *Journal of Product and Brand Management,* 16(2), pp. 152-153.

Lukic, M., Tuba, E. & Tuba, M., 2017. *Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns.* s.l., 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI).

Mathioudakis, M. & Koudas, N., 2010. *TwitterMonitor: Trend Detection over the Twitter Stream.* Indianapolis, Indiana, USA, ACM.

Mei, L., Xiang, W., Kai, G. & Shanshan, Z., 2017. A Survey on Information Diffusion in Online Social Networks: Models and Methods. *Information,* 8(4), p. 118.

Newman, M. E. J., 2003. The Structure and Function of Complex Networks. *SIAM Review*, 45(2), pp. 167-256.

Parau, P. et al., 2013. *Using Community Detection for Sentiment Analysis.* Cluj-Napoca, Romania, IEEE.

Piatetsky, G., 2019. *Kdnuggets.* [Online]
Available at: https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html
[Accessed 05 August 2019].

Pringle, J. & Fritz, S., 2018. The university brand and social media: using data analytics to assess brand authenticity. *Journal of Marketing for Higher Education*, pp. 19-44.

Pringle, J. & Huisman, J., 2011. Understanding Universities in Ontario, Canada: An Industry Analysis Using Porter's Five Forces Framework. *Canadian Journal of Higher Education* , 41(3), pp. 36 - 58.

Stieglitz, S. & Dang-Xuan, L., 2013. Emotions and Information Diffusion in Social Media - Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, Volume 29, pp. 217-247.

Stieglitz, S. & Dang-Xuan, L., 2013. Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4), pp. 1277-1291.

Stieglitz, S., Dang-Xuan, L., Bruns, A. & Neuberger, C., 2014. Social Media Analytics: An Interdisciplinary Approach and Its Implications for Information Systems. *Business & Information Systems Engineering*, 6(2), pp. 89-96.

Stieglitz, S., Mirbabaie, M., Ross, B. & Neuberger, C., 2018. Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, p. 156–168.

Temple, P., 2006. Branding higher education: illusion or reality?. *Perspectives*, 10(1), pp. 15 - 19.

Twitter, 2019. *Rate Limiting.* [Online]
Available at: https://developer.twitter.com/en/docs/basics/rate-limiting.html
[Accessed 07 August 2019].

Waeraas, A. & N.Solbakk, M., 2009. Defining the essence of a university: lessons from higher education branding. *Higher Education*, Volume 57, pp. 449-462.

Wee Eng Kim, V. & Periyayya, T., 2013. Student Expectations and Branding Strategies among Private Institutions of Higher Education in Malaysia. *Malaysian Journal of Chinese Studies*, Volume 2, pp. 69-81.

WeiXian, W., 2016. *Popularity of Brand Posts on Sina Weibo: A Correlation Analysis of the Influential Factors on Tuborg's Brand Community*, s.l.: Uppsala University.

Wirth, R. & Hipp, J., n.d. CRISP-DM: Towards a Standard Process Model for Data.

Zadeh, A. H. & Sharda, R., 2014. Modeling brand post popularity dynamics in online social networks. *Decision Support System*, 65(C), pp. 59 - 68.

Zeng, D., Chen, H., Lusch, R. & Li, S.-H., 2010. Social media analytics and intelligence. *IEEE Intelligent Systems* , 25(6), pp. 13-16.

APPENDIXES

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.:** 2
Student Name & ID: Tan Sze Mei 1603383	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University	

1. WORK DONE [Please write the details of the work done in the last fortnight.] • Note down on feedback on final year project I report.
2. WORK TO BE DONE • Search for Facebook data extraction method. • More reading to propose a framework / model.
3. PROBLEMS ENCOUNTERED • Facebook data privacy concern.
4. SELF EVALUATION OF THE PROGRESS • More determination and perseverance in research are needed.

Pradeep

Supervisor's signature

SzeMei

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.:** 4
Student Name & ID: Tan Sze Mei 1603383	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- Studied few information diffusion model to fit into university branding case study.

2. WORK TO BE DONE

- Explore more model.

3. PROBLEMS ENCOUNTERED

- Some models are difficult to understand.

4. SELF EVALUATION OF THE PROGRESS

- Lack of knowledge.

Pradeep

Supervisor's signature

SzeMei

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.:** 6
Student Name & ID: Tan Sze Mei 1603383	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- Decided which conceptual framework to refer to.

2. WORK TO BE DONE

- Determine which influential factor that will affect information diffusion to study.

3. PROBLEMS ENCOUNTERED

- The measurement of influential factor is subjective.

4. SELF EVALUATION OF THE PROGRESS

- Slow progress.

Pradeep

Supervisor's signature

SzeMei

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.:** 8
Student Name & ID: Tan Sze Mei 1603383	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University	

1. WORK DONE [Please write the details of the work done in the last fortnight.] • Determined which influential factor to include in the conceptual framework
2. WORK TO BE DONE • Collect data from Facebook
3. PROBLEMS ENCOUNTERED • A lot of data needs manually label for the parameter included.
4. SELF EVALUATION OF THE PROGRESS • Manual progress

Pradeep

Supervisor's signature

SzeMei

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.:** 10
Student Name & ID: Tan Sze Mei 1603383	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University	

1. WORK DONE [Please write the details of the work done in the last fortnight.] • Model done tested with Pearson Correlation test.
2. WORK TO BE DONE • Increase data size. • Update report.
3. PROBLEMS ENCOUNTERED • Low correlation found between influential factors and information diffusion.
4. SELF EVALUATION OF THE PROGRESS • Lack of knowledge on data analytics.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.:** 12
Student Name & ID: Tan Sze Mei 1603383	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- Prepared presentation slides.
- Updated report.

2. WORK TO BE DONE

- Amendment on report.

3. PROBLEMS ENCOUNTERED

- Cannot provide firm explanation about what experiment have proven.

4. SELF EVALUATION OF THE PROGRESS

- Slow progress.

Pradeep

Supervisor's signature

SzeMei

Student's signature

Poster

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY, UNIVERSITI TUNKU ABDUL RAHMAN
BACHELOR OF COMPUTER SCIENCE
FINAL YEAR PROJECT

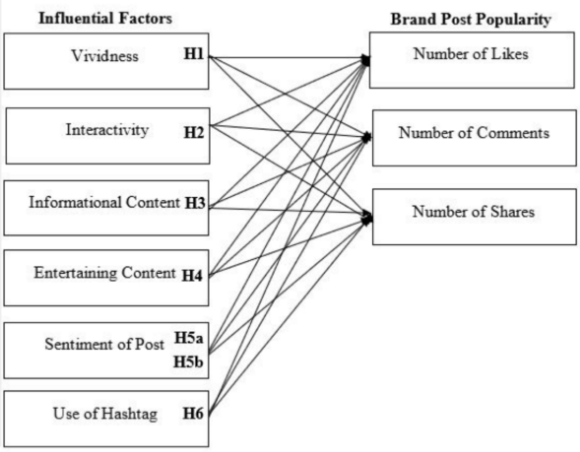
Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University

INTRODUCTION

As for now, social media is being part of an individual life. Even academic institution tends to shift its marketing channel to social media. In this research, how information in social media based on sentiment diffuse among the users is investigated.

METHODS

A conceptual framework of brand post popularity inspired by the study of de Vries et al. (2012) is proposed and tested using Pearson Correlation and spearman correlation analysis.



RESULTS

According to table shown below, five hypotheses are fully supported while negative sentiment and interactivity are partially supported.

	Number of Likes: Coefficient Value	Number of Comments: Coefficient Value	Number of Shares: Coefficient Value
Vividness	Supported: 0.122	Supported: 0.137	Supported: 0.255
Interactivity	Not Supported: 0.016	Supported: 0.112	Supported: 0.194
Informational Content	Supported: 0.208	Supported: 0.186	Supported: 0.266
Entertaining Content	Supported: 0.357	Supported: 0.230	Supported: 0.192
Positive Sentiment	Supported: 0.200	Supported: 0.104	Not Supported: 0.032
Negative Sentiment	Supported: -0.187	Supported: -0.192	Supported: -0.144
Use of Hashtag	Supported: 0.208	Supported: 0.186	Supported: 0.266

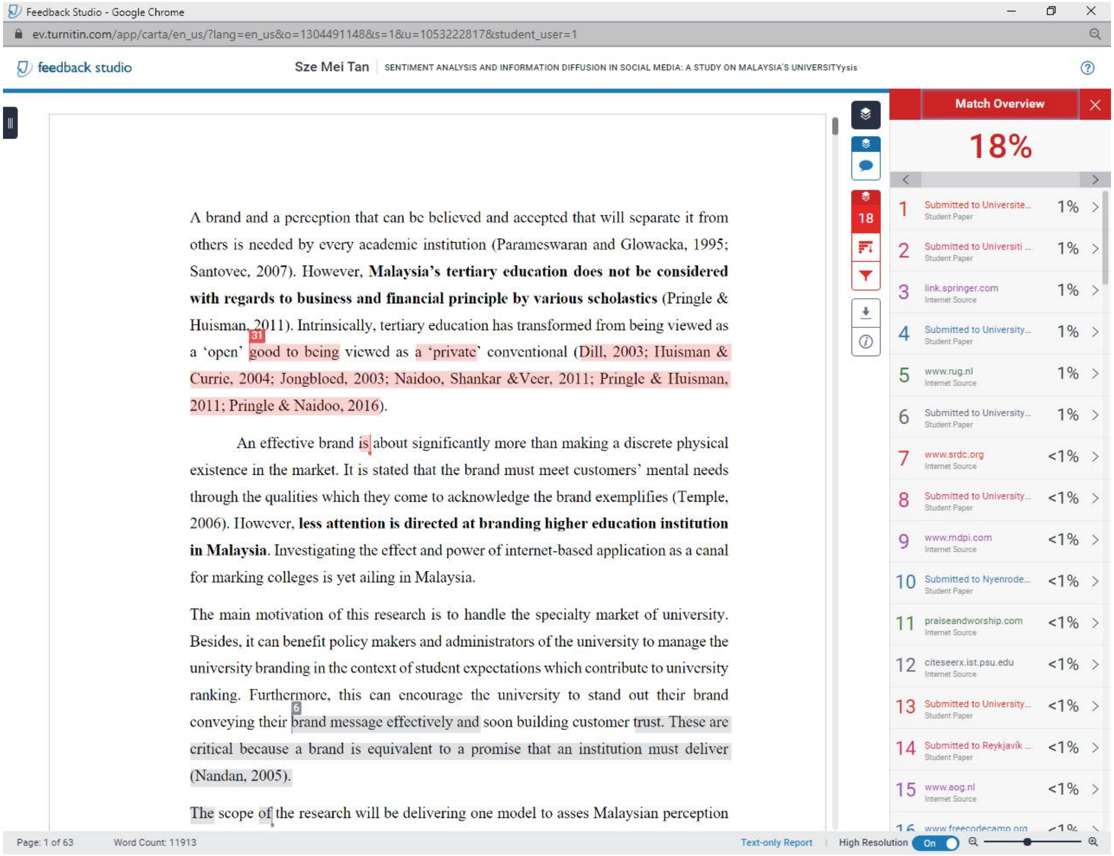
CONCLUSION

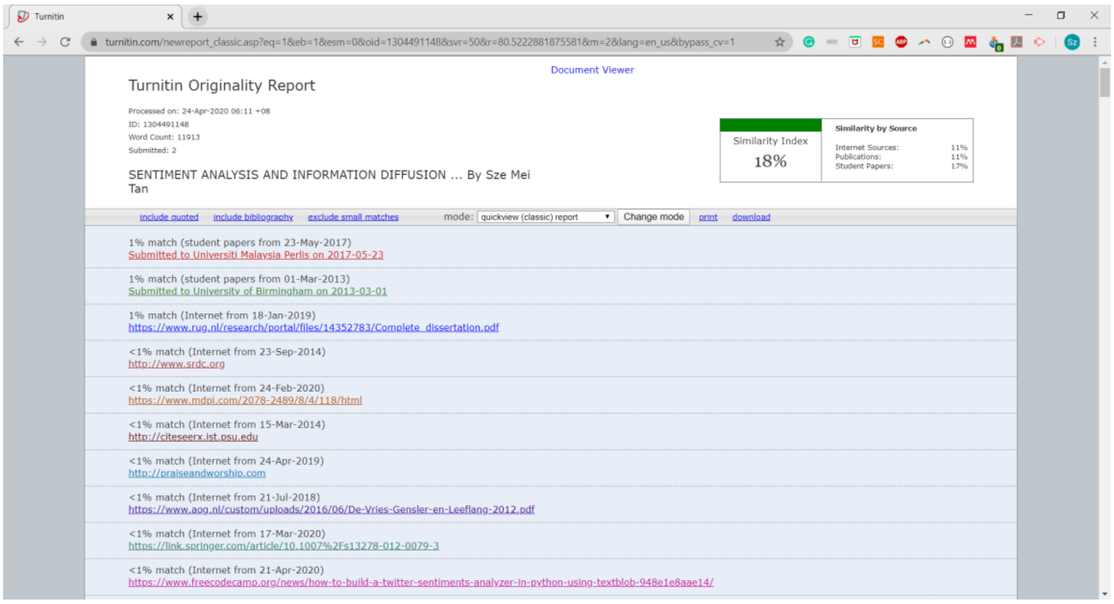
⌘ University page administrator shall include the entertaining element in the post to boost the number of likes, comments.

⌘ Besides, to diffuse the information to more people, the post shall be informational with the use of hashtag for invoking sharing purpose.

by: TAN SZE MEI

Plagiarism check summary





Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Tan Sze Mei
ID Number(s)	16ACB03383
Programme / Course	BCS (Hons) Computer Science
Title of Final Year Project	Sentiment Analysis and Information Diffusion in Social Media: A Study on Malaysia's University

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: ___18_____ % **Similarity by source** Internet Sources : ______ 11________% Publications : _______11________% Student Papers : ______ 17________%	
Number of individual sources listed of more than 3% similarity: 0________	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Pradeep

______________________________ Signature of Supervisor

Name: Dr. Pradeep a/l Isawasan

Date: 24/4/2020

______________________________ Signature of Co-Supervisor

Name: ____________________________

Date: _____________________________

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	16ACB03383
Student Name	Tan Sze Mei
Supervisor Name	Dr. Pradeep a/l Isawasan

TICK (√)	**DOCUMENT ITEMS** Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
✓	Front Cover
✓	Signed Report Status Declaration Form
✓	Title Page
✓	Signed form of the Declaration of Originality
✓	Acknowledgement
✓	Abstract
✓	Table of Contents
✓	List of Figures (if applicable)
✓	List of Tables (if applicable)
	List of Symbols (if applicable)
✓	List of Abbreviations (if applicable)
✓	Chapters / Content
✓	Bibliography (or References)
✓	All references in bibliography are cited in the thesis, especially in the chapter of literature review
	Appendices (if applicable)
✓	Poster
✓	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report. SzeMei ______________________ (Signature of Student) Date: 24/4/2020	Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction. Pradeep ______________________ (Signature of Supervisor) Date: 24/4/2020