Global Business and Management Research: An International Journal
Vol. 13, No. 3 (2021)

Application of Text Mining to Analyze Customer Opinions on Social Media

Dr. Rolou Lyn R. Maata*
*Assistant Professor, Gulf College, Muscat, Sultanate of Oman
E-mail: rolou@gulfcollege.edu.om

Dr. Aaron Paul Pineda
Lecturer, Higher College of Technology, Madinat Al Zayed, United Arab Emirates

Dr. Ferdinand J. Epoc
Assistant Professor, University of Buraimi, Sultanate of Oman

Dr. Ronald Cordova
Assistant Professor, Gulf College, Muscat, Sultanate of Oman
** Corresponding Author*

Abstract

Purpose: This study aims to evaluate and analyze customers' sentiments that would provide relevant support to businesses in decision making and in redesigning companies' digital marketing strategies and campaigns in the future.

Design / Methodology / Approach: The researchers decided to use text mining and sentiment analysis processes and simulation approaches to analyze customers' opinions. The researchers extracted data using Twitter API using Rapid miner version 9.8 with a total tweet of nine hundred forty-seven (947) tweets using keyword search "clothing brand" distinctively in Oman. These data were analyzed using Tableau Desktop professional edition version 2020.4 and rapid miner 9.8.

Findings: The results of this research study indicated that generally positive customer's opinions and feedback have been revealed. Further, the results also revealed that social media data can be transformed into valuable knowledge and be a source of improvements from business particularly in digital marketing. Therefore, text mining and sentiment analysis processes are considered effective technique and powerful tool to determine business value.

Research Limitations / Implications: This research study demonstrates how user interactions and customers' opinions on social media specifically twitter can be used for digital marketing which later on will have a huge impact to the intangible characteristics of the business such as brand, value, reputation, and loyalty.

Originality/Value: This research study was conducted to shed light on how social media comments and feedback can be transformed into knowledge to improve marketing strategies in most businesses in Oman. The results of the study could turn business into making decisions. Further, this study can provide valuable pointers to (a) policy makers as a guide in reviewing a law related to digital marketing, (b) businesses in Oman where it can be considered as basis for re-designing their marketing strategies and campaigns, (c) students and future researchers for them to learn more on this research and apply it in their future research.

Keywords: text mining, sentiment analysis, twitter, digital marketing
Category: Applied Paper

1. Introduction

Due to the advancement of information technology, social media have become popular and commonly used platforms for content sharing and social networking. It is an internet-based application that allows users to interact and share valuable information in real time. It plays a vital role in impacting individual's knowledge, culture, and personality. Social media creates forum where people bring together to exchange and express ideas and opinions freely and connect with people around the globe to seek for possible advice, support and guidance.

In today's generation, social media is used immensely in the business world to attract and build strong customer networks, increase market reach, low-cost advertising, and marketing, and helps develop business brand reputation. Social media platforms such as Instagram, facebook, twitter, etc. help businesses reach their target customers to sell and promote products and services. Twitter for example, is considered as one of the most well - known social media platforms widely used by customers to express their opinions called "tweets" on a specific product, service, or brand. It replaces the traditional method of marketing and sharing customers' comments and feedback that produces vast amount of customers' opinions daily. Marketers use social media data to understand consumers' behavior, personality, and environment to plan their marketing strategies and campaign, and to boost customers' engagement, improve brand awareness and reputation (Howell & Ertugan, 2017). Thus, gathering these consumers' opinions and analyze them appropriately would help businesses create or redesign marketing strategies and campaigns (Tuten & Solomon, 2014). In addition, the study of Bin & Chan (2014) proved that consumers' opinions create valuable insights into the markets, and it helps produce information relevant for decision making to improve business marketing strategy.

This paper demonstrates the text mining and sentiment analysis processes in reviewing customers' opinions towards clothing brand using twitter data and provides valuable results as basis for decision making. It consists of four (4) sections namely: literature review, methodology used, detailed discussions on the simulation results, and conclusion. According to (Moreno & Redondo, 2016) there are numerous literature available about text mining of twitter data, hence, the outcome of the study attempts to contribute to the developing field of text mining and data analytics.

2. Literature Review

Growing number of companies have used and adopted social media to interact with their customers as part of their business processes and strategy to enhance company's performance and achieve organization's goals (Smits & Mogos, 2013). Social media platforms such as Twitter produces huge amount of unstructured data from various customers' opinions through product reviews, product comments, and suggestions. These customer-generated data is available freely online. Many researchers all over the world have used these data to transform into knowledge as a tool for businesses to make decisions.

2.1 Text Mining

Text Mining is defined as an automated process of discovering hidden but useful information and revealing new knowledge from unstructured computerized text data sources (Younis, 2015). It targets that unrevealed knowledge from massive amount of data or text (Gupta & Lehal, 2009). Basically, the patterns for text mining are extracted a precise information from natural language

text which is unstructured based on keywords, concepts, entities, phrases, relationships, and even sentences (Agrawal & Batra, 2013).

Many businesses are greatly impacted by social media data as discussed in several research studies conducted on text mining. Text mining is widely used to analyze and interpret social media data in a diverse discipline such as in business, medicine, engineering, education, etc. In a study of Tay et.al. (2020) it revealed that the application of text mining is an effective way to enhance human judgment drawn from public attention. This result was supported by the study of (Yun, et.al, 2020) which considers text mining as an effective approach to analyse residents' sentiments towards the rise of electricity policy. The result of applying this method provides clear understanding of residents' concerns and gives positive impact to the residents, government, and the community. Further, the success of every business relies on how the business is greatly influences by the sentiments of their customers (Manoharan & Sathish, 2020). This research utilized social media data to reveal their research results.

Social media data are publicly available twenty-four seven (24/7) for public consumption. These data can be used by businesses to validate competitors' performance, analyse customers' sentiments and opinions, and redesign digital marketing strategies, etc. by performing text mining and sentiment analysis. The results of the analysis may provide clear direction in achieving business goals, identifying business strengths and weaknesess, and addressing customers' suggestions / feedback to meet their needs (He, et.al., 2013). The study of (Cheng, et.al, 2011) agreed on the effectiveness of text mining in evaluating business performance and customers' feedback through social media reviews. In fact, there are many authors who successfully used text mining techniques in various discipline such as in business (Ingvaldsen & Gulla, 2012) and science (Li, et.al, 2012), and education (Abdous & He, 2011) to analyse huge amount of data from social media platforms.

Further, text mining is performed using software packages. Many software packages and tools are available in the market today for text mining that uses computational algorithms based on natural language processing (NLP). These tools and packages allow users to identify the sample relevant data to extract, synthesize, analyze and interpret to produce relevant knowledge (Feinerer, 2014).

2.2 Sentiment Analysis

In recent years, most of the sentiment analysis conducted by researchers was performed using twitter data in which they concluded that results were accurate and relevant for businesses Ruz, et.al, 2020). Sentiment analysis also known as opinion mining was introduced by (Liu, B., 2010) which uses natural language processes, biometric or even text analysis to identify, mine, perform, analyse and quantify data. This concept analyses data to addresses the real business problem to come up with a real and useful solution.

There are many ways on how to perform sentiment analysis. For instance, a combination of lexicon from deep data collection can be separated based on different levels of emotions and analyse using appropriate tools to avoid bias results (Valravasundaram, et.al, 2020)

On the other hand, web data can also be analysed using sentiment analysis, detection and classification depending on the purpose. Its coverage is from the areas of finance, consumer markets, government, and media. Size of data scrape from twitter, diversity of social web data, and subjectivity are just few of the many characteristics of information involved in performing sentimental analysis.

In a study conducted by (Smits & Mogos, 2013) they emphasized that sentiment analysis contributes to the realization of one's company's strengths and weaknesses allowing them to reinvent their policies, and strategies. In addition, this approach is not only for businesses but also can contribute to other areas of discipline like computing. (Kumar & Jalswal, 2019) tackled about how text mining and sentiment analysis contributes to the body of knowledge and in research domain.
Therefore, all studies integrated in this section proved that this study could contribute to the core value of customers' opinions in creating or redesigning digital marketing strategies and campaigns.

3. Methodology

3.1 Context of the Study

The clothing brand industry is just one of the many industries that have come into the world of social media for the betterment of their business and has huge number of users in all social media platforms. In this study, we focused to mine the comments and feedback of customers from one clothing brand business in Oman.
Traditionally, clothing businesses promote products and services through various traditional methods of marketing channels like tv advertising, magazines, leaflets, etc. Due to the rapid advancement of technology and the widespread us of social media like for example Instagram, Facebook, and twitter customers are easy to reach out. In a study on gratification theory, (West & Turner, 2010) customers would look into appropriate media outlets and valuable information for gratification purposes. Thus, social media become progressively famous media outlet for target customers.

3.1.2 Methodology Used for Text Mining and Sentiment Analysis

The methodologies used for this study are as follows:

a. Simulation method was used in this study as the most appropriate method to evaluate and analyze selected customers' opinions extracted from twitter to form a dataset. This method helps the researchers come up with evidence-based decisions that were processed using Tableau desktop professional edition version 2020.4 and rapidminer studio version 9.8 and enable to visualize the results in a well-structured manner.

b. Text Mining and Sentiment Analysis method - adapted from the study of Adhous & He (2012) and was used to simply demonstrate how researchers process the dataset collected from twitter (refer to Figure 1 below)

1. Text Pre-processing involves extraction and preparation, and text collection. There are many processes involved in this step: first, data access and data cleaning. The researchers extracted twitter data messages using twitter API and use keyword search to retrieve messages with pre-defined keyword. Second, we do perform data cleaning using rapid miner to get the exact tweet messages, clean the data which are unnecessary for the simulation process and remove all those spaces and punctuation or with URLs. This process converts the unstructured data into a structured on depending on the needs of the study.

2. Text Analysis involves processes structured representation of data that was produced in the previous step. In this step, applying txt mining is important. This may include lexicon-based

approach in sentiment analysis in which it identifies the tweet as positive or negative tweets and scoring function must assigned score to every tweet.

3. Visualization & Action are processes that take place after the analysis. The results shown in previous step will carry over on to the next step. The wordcloud and the summary of plots showing the frequency of customers' tweets and sentiment score were visualized. Based on these results, the conclusion and future work were drawn.

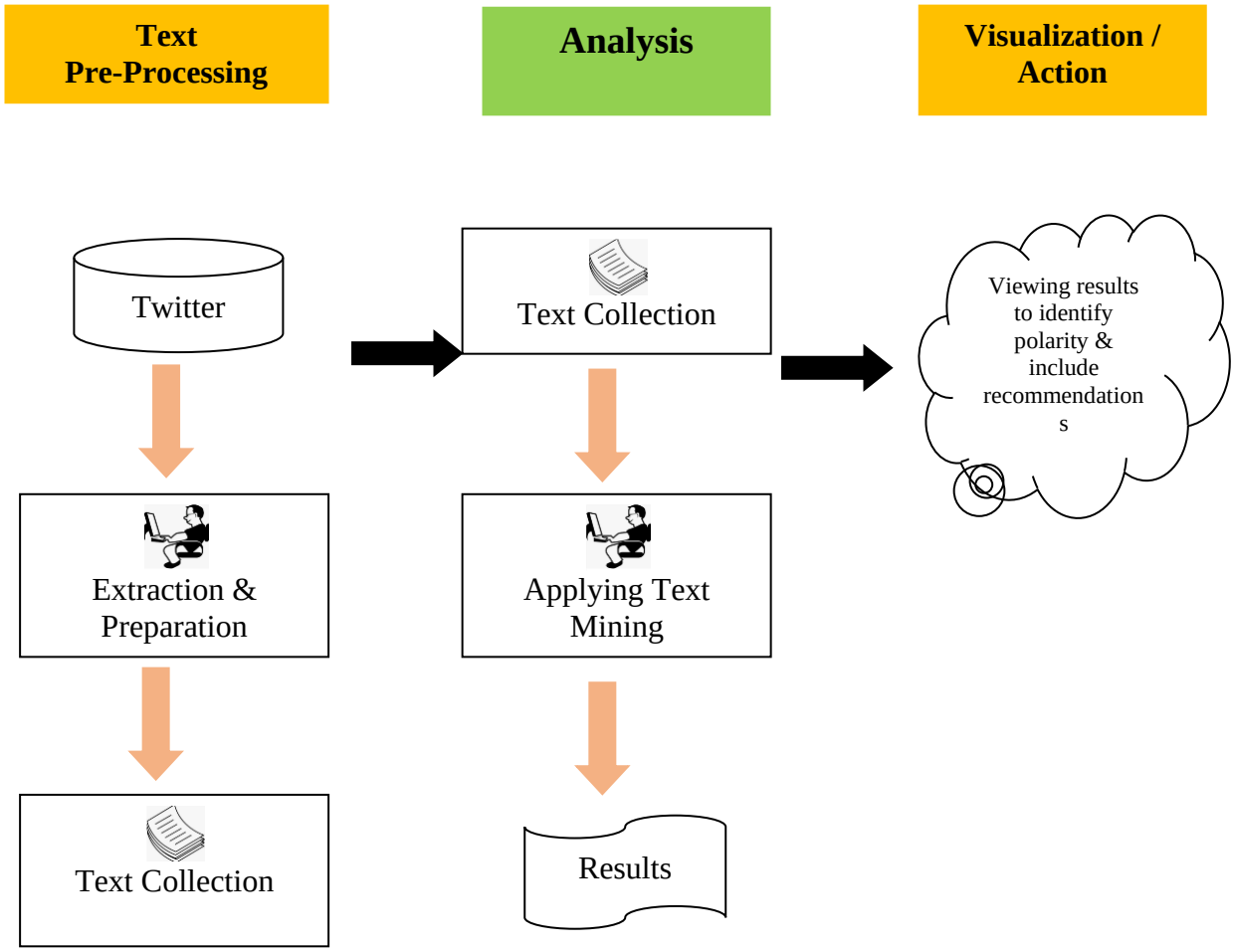

Figure1: Text Mining and Sentiment Analysis Methodology Used

4. Findings and Discussion

The simulation involved collecting a total of nine hundred forty-seven (947) tweets from a clothing brand available in Muscat Oman in a period of one (1) week. These tweets were processed and mined as textual information retrieved from Twitter using the account name of the clothing brand. The rapidminer and twitter API have been utilized to access the data in twitter, extract tweet text, and cleaning the data. Twitter API facilitates twitter authentication and access messages using keyword search (Danneman & Heimann, (2014).

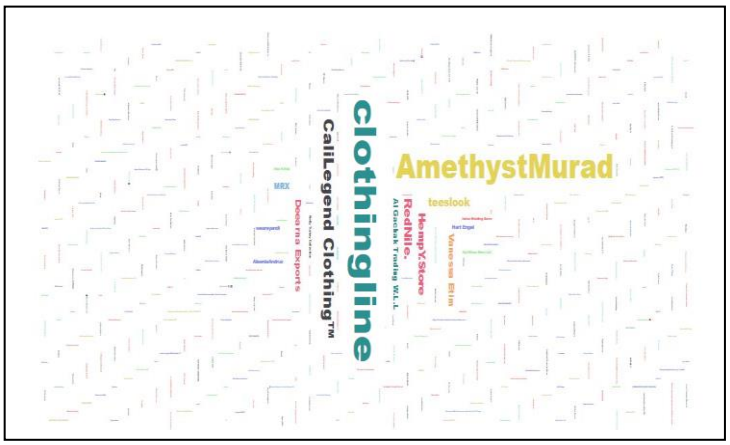

Figure 2: Sample word cloud representation of tweets

Figure 2 illustrates the sample output of the word cloud representation of tweets that was done using tableau software. It clearly demonstrates the importance of each word in the cloud. The size of the word represents the number of mentions in tweets. In this output, it clearly shows that the word "clothing line" and "amthystmurad" were mentioned frequently. This is an indication that the most common word used by users were displayed in a large text in the word cloud.

In relation to the sentiment analysis, the lexicon-based approach was utilized in this study. After applying the text mining approach, the bag of words (BOW) which serves as the representation of tweets is done. To perform a sentiment analysis, we must have an opinion lexicon and the scoring functions. The scoring function assigns score to every tweet following the equation below adapted from the study of (Younis, 2015):

Sentiment Score= Σ positive words – Σ Negative words
… (Equation 1)

Table 1: Sentiment Score Definition

	Sentiment Score
Neutral	Score is 0 if both words either positive or negative does not exist Score is 0 if both words (positive or negative) are equal
Positive Polarity)	Score is positive if total positive words is > total negative words
Negative Polarity	Score is negative if total positive words is < total negative words

Table 2:Summary of sentiment polarity performed by rapidminer with their Corresponding scores and frequency

Sentiment Score	Neutral	Positive					Negative		
	0	1	2	3	4	5	-3	-2	-1
Frequency	40	189	258	97	72	93	38	96	64

Table 1 illustrates the distribution of sentiment polarity based on their corresponding scores and frequency. A total of seven hundred nine (709) tweets were positive wherein there were one hundred ninety-eight (198) tweets that refers to negative and forty (40) tweets were neutral. Overall, the tweets for this clothing brand in Oman can focus on the negative comments from their customers to improve their services. Therefore, businesses can come up with a clear marketing plan or strategy to address the issues or concerns raised by the customers based on the results of this study.

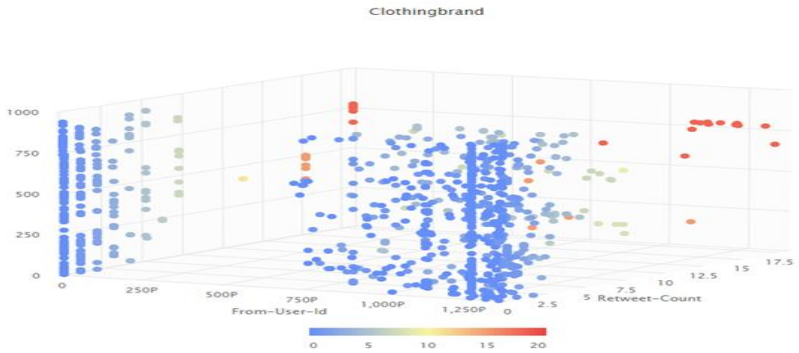

Figure3: Distribution of Number of Tweets per User

Figure 3 demonstrates the distribution of tweets and re-tweets per user ID. It clearly shows that tweets messages have been re-tweeted and shared frequently. The blue dots simply means that the tweets have been re-tweeted from the range of 0 to 10 and the red dots simply means that the tweets have been re-tweeted more than 15 times. Therefore, it is clear evidence that customers tend to retweet messages concerning "clothing brand".

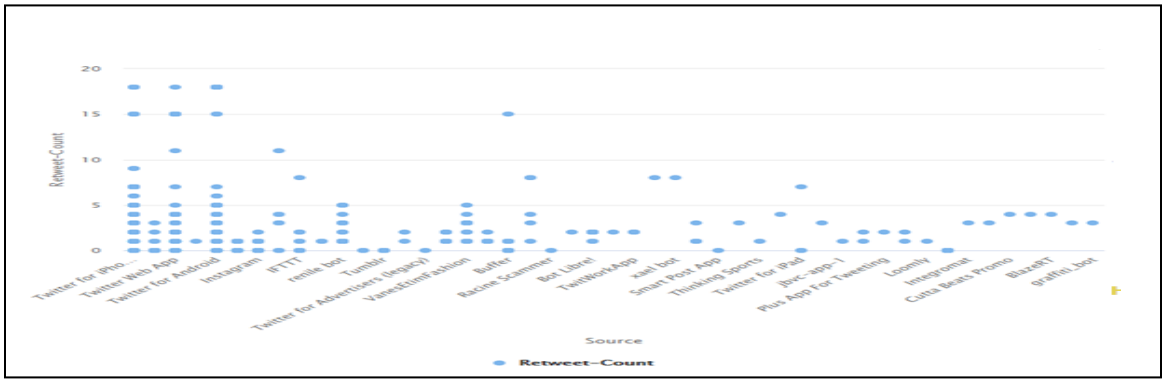

Figure4: Distribution of Sources of Tweets

Figure 4 shows the distribution of sources of tweets based on the tweet messages from the customers. The above graph shows that many of the users uses Iphone device to tweet and re-tweet their messages. This simply means that businesses nowadays must invest a lot in digital marketing as many customers have already access to various devices like Iphone.

5. Conclusion

After thorough analysis of the data used in this study, the researchers concluded that businesses can collect, pre-process text data, analyze and visualize data using Rapidminer and Tableau packages. It is also highly feasible utilize customers' sentiments and opinions using social media to re-design their marketing strategies (digital marketing), marketing plans, and adapting business intelligence in the future. The result of the analysis will provide businesses opportunity to improve customer services and business value to provide better service and enhance customer relationships. The main contribution of this study is a sequential view of how customers' opinions changed the digital marketing strategies and campaign of the business based on sentiment analysis of the twitter data.

6. References

Abdous, M., He, W. & Yean, C.J. (2012). Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade. *Educational Technology & Society, 15(3). Pp. 77-88.*

Agrawal, R. & Batra, M. (2013). A Detailed Study on Text Mining Techniques. *International Journal of Soft Computing and Engineering.* ISSN: 2231-2307, Vol. 2, Issue 6 p. 118 – 121.

Bing, L. & Chan, K.C.C. 2014. A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data, *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, London*, pp. 652-657.

Cheng, L., Ke, z. & Shiue, B. (2011). Detecting Changes of Opinion from Customer Reviews. In proceedings of *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1798 – 1802).*

Danneman, N., & Heimann, R. (2014). Social Media Mining with R. *Packt Publishing Ltd.*

Feinerer, I. (2014). Introduction to the tm package text mining in R. nd): n.page.web.

Gupta, V. & Lehal, G.S. (2009). A Survey of Test Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence,* Vol. 1 (1). p. 60-76.

He, W., Zha, S., & Li, L. (2013). Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry. *International Journal of Information Technology Vol. 33, p. 464 – 472.*

Howell, K., Ertugan, A. (2017). Applying Fuzzy Logic for Sentiment Analysis of Social Media Network Data in Marketing. 9th *International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, Vol. 120, p. 664-670.*

Ingvaldsen, J.E., & Gulla, J.A. (2012). Industrial Application of Semantic Process Mining. *Enterprise Information Systems, 6(2), pp. 139-163.*

Kumar, A., & Jalswal, A. (2019). Systematic literature review of sentiment analysis on Twitter using Soft Computing Techiques, *Special Issue on Concurrency and Computation : Practice and experience.*

Liu, B. (2010) .Sentiment Analysis and Subjectivity, Handbook of Natural Language Procssing pp. 67-666.

Li, L., Ge, R., Zhoud, S., & Valerdi, R. (2012). Guest Editorial Integrated Healthcare Information Systems. *IEEE Transactions on Information Technology in BioMedicine, 16(4), pp. 515-517.*

Manoharan, S. & Sathish (2020). Geospatial and Social Media Analytics for Emotion Analysis of Theme Park Visitors using Text Mining and GIS. *Journal of Information Technology and Digital World Vol. 2 No. 2 pp. 100-107*

Moreno, A. & Redondo, T. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. *International Journal Interact Multimedia Artificial Intelligence,* Vol. 3 No. 6 p. 57

Ruz, G., Henriquez, P., & Mascareno, A. (2020). Sentiment Analysis of Twitter data during Critical Events through Bayesian Networks Classifiers. *Journal for Future Generation Computer Systems Vol. 106, p. 92-104.*

Smits, M. & Mogos, S. (2013). The Impact of Social Media on Business Performance *Association for Information System Library (AISeL) ECIS 2013 Proceedings of the 21st European Conference on Information Systems.*

Sun, Y., Wang. Z., Zhang, B., Zhao, W., Xu, F., Liu, J. & Wang, B. (2020). Residents' sentiments towards electricity price policy: Evidence from Text Mining in Social Media. *Resources, Conservation and Recyling Journal of Elsevier, Vol. 160, pp.*

Tay, L., Woo, S.E., Hickman, L. & Saef, R. (2020). Psychometric and Validity Issues in Machine Learning Approaches to Personality Assessment: A Focus on Social Media Text Mining. *European Journal of Personality. Vol. 34, pp. 826-844*

Tuten, T., & Solomon, M., 2014. Social Media Marketing, 1st ed., Pearson, Harlow, Essex, UK

Valravasundaram, S. , Ravi, L., Abejith, M., Umansankar, S. & Umamakeswari, A. (2020). Sentiment Analysis of Tweets for estimating Critically and Security of events. *Journal of Organizational and end user Computing, pp.9*

West, R. L., & Turner, L. H. (2010). "Uses and gratifications theory". Introducing communication theory: Analysis and application. *Boston: McGraw-Hill*., pp. 392–409.

Younis, E. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. *International Journal of Computer Applications,* Vol. 112 No. 5, p. 44