

**SENTIMENT ANALYSIS
ON
ONLINE SOCIAL MEDIA**

A PROJECT REPORT

Submitted by

NITISH M (2116210701184)

in partial fulfillment for the

award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING



**RAJALAKSHMI ENGINEERING
COLLEGE ANNA UNIVERSITY,**

CHENNAI

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE,CHENNAI

BONAFIDE CERTIFICATE

Certified that this Thesis titled “**SENTIMENT ANALYSIS ON ONLINE SOCIAL MEDIA**” is the bonafidework of “**NITISH M (2116210701184)**” who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr . S Senthil Pandi M.E.,Ph.D.,

PROJECT COORDINATOR

Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai - 602 105

Submitted to Project Viva-Voce Examination held on_____

Internal Examiner

External Examiner

SENTIMENT ANALYSIS ON ONLINE SOCIAL MEDIA

NITISH M (210701184)

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

210701184@rajalakshmi.edu.in

ABSTRACT:

This article is a social media sentiment analysis report that explored the methods, the social media platform used and its application. Social media contain large amounts of raw data uploaded by users in the form of text, videos, images and audio. Data can be turned into valuable information using sentiment analysis. The surge of social media platforms has led to an exponential increase in user-generated content, offering unique opportunities and challenges for sentiment analysis. This research delves into applying sentiment analysis techniques to social media data to understand public opinion, emotions, and trends. We utilize a combination of natural language processing (NLP) and machine learning algorithms to analyze sentiments expressed in posts from platforms like Twitter, Facebook, and Instagram. Our approach involves data collection, preprocessing, feature extraction, and the implementation of supervised learning models, such as Support Vector Machines (SVM) and neural networks, to determine sentiment polarity (positive, negative, neutral). Furthermore, we investigate the effectiveness of advanced techniques, including deep learning and transformer-based models like BERT, in improving sentiment prediction accuracy. The results demonstrate that context-aware models significantly outperform traditional methods in capturing the subtleties of informal, user-generated text. This study underscores the importance of contextual and linguistic nuances in sentiment analysis and proposes a framework for real-time sentiment monitoring and analysis, offering valuable insights for businesses, policymakers, and researchers aiming to understand and respond to societal trends.

INTRODUCTION:

The advent of Web 2.0 is changing the world of social media. On the Internet, social media is not only used to communicate, share information and personal opinions with others, but also companies can communicate, understand and improve their products and services by connecting on social media. The number of social media users is increasing every day and in 2019 it is estimated that there are up to 2.77 billion social media users worldwide . Different types of information are uploaded and shared on social media in the form of text, videos, photos and audio . Social media is full of raw and unprocessed data, and advances in technology, especially machine learning and artificial intelligence, allow data to be processed and transformed into useful data that can be useful to most business organizations . This article focuses on examining related literature published between 2014 and 2019 to better understand the application of sentiment analysis on social media platforms. Sentiment analysis is an approach that uses natural language processing (NLP) technology. transform and interpret opinions from text and classify them as positive, negative or natural feelings . Most previous studies have applied sentiment analysis to product or movie reviews to better understand their customers and make the necessary decisions to improve the product or service..

Scholars have been studying sentiment analysis since the last decade, and most articles began to appear and grew rapidly after 2004 [6]. Sentiment analysis is

divided into three different levels, which are sentence level, document level and feature level. The goal is to classify the opinion about either sentence, document or features into positive and negative feelings . Two main methods have been identified for sentiment analysis, which are a machine learning approach and a vocabulary based approach. The machine learning approach used algorithms to extract and identify opinions from the data, while the vocabulary-based approach counted positive and negative words associated with the data. Scientists have developed a new efficient and accurate model to analyze emotions. But the challenge comes in developing a model where most of it is for English. However, a recent study shows that opinion analysis models exist in other languages, such as Korean , Thai , Arabic , Malay , Portuguese and the chinese. Regarding the application of sentiment analysis, it has been reported to have been done in the context of business and marketing, politics and public relations. An example of an application is an online store, a voting application and world events [14]. Most of the information collected for the study was collected on social networks. Social media contains a huge amount of data about internet users, and we can get all kinds of information about a product, service, place or event, which makes it suitable for opinion research..

2.LITERATURE REVIEW

The rapid expansion of social media platforms has created an abundance of user-generated content that offers unique opportunities for sentiment analysis. This field involves the application of computational methods to extract subjective information from text with the aim of understanding public opinion, emotional reactions and trends, which are crucial for various applications in business, politics and social research.

1. Development of sentiment analysis techniques

Initial sentiment analysis methods mainly used vocabulary-based approaches based on dictionaries of words loaded with predefined sentiments to measure text polarity (Taboada et al., 2011). Although these methods were simple, they often failed to address the nuances of the informal and contextual language of social media. The introduction of machine learning techniques marked a significant advance, enabling more complex and adaptive models. Pioneers such as Pang et al. (2002) used machine learning algorithms such as Naive Bayes, Maximum Entropy and Support Vector Machines (SVM) to classify emotions.

2. Machine learning methods

Supervised learning techniques have been widely used in sentiment analysis. SVMs excel in their high-performance text classification tasks (Joachims,

1998), while neural networks have shown great promise in detecting complex data patterns (Collobert et al., 2011). Various feature extraction methods including n-grams, part-of-word tagging, and word embedding have been used to improve model accuracy (Mikolov et al., 2013).

3. Advanced models and deep learning

The advent of deep learning has changed sentiment analysis even more. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), especially long-term memory (LSTM) networks, have shown excellent capabilities in sequential data processing and context capture (Kim, 2014; Hochreiter and Schmidhuber, 1997). Recently, transformer-based models such as Bidirectional Encoding Representations of Transformers (BERT) have created new standards for sentiment analysis by effectively understanding context (Devlin et al., 2019).

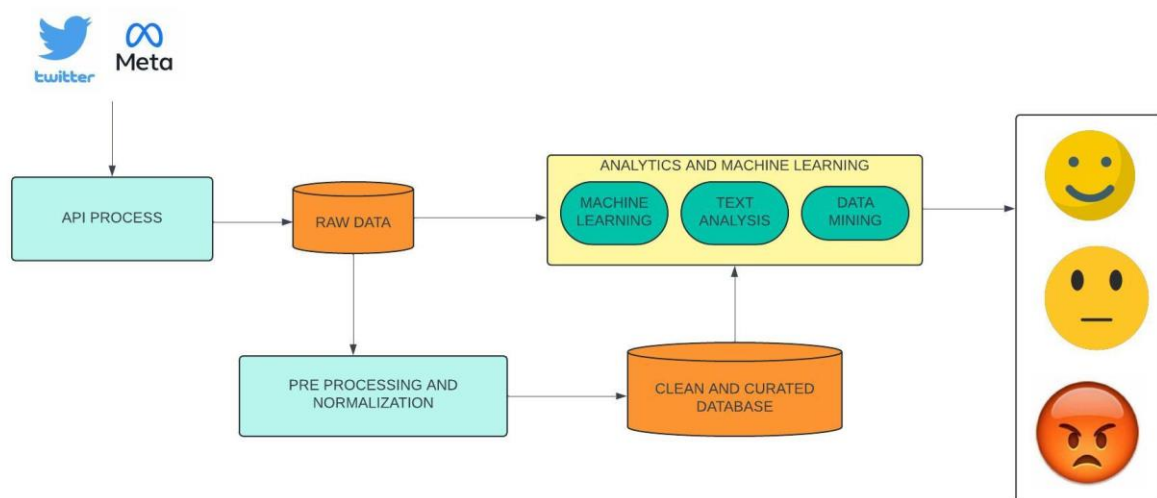
4. Challenges in Social Media Sentiment Analysis

Despite these advances, Social Media Sentiment analysis still faces several challenges. The informal, noisy and often sarcastic nature of social media text makes it difficult to identify emotions (Maynard and Greenwood, 2014). In addition, this task is made more difficult by the rapid development of language and the presence of domain-specific jargon. Meeting these challenges requires constant innovation and adaptation of sentiment analysis methods.

5. Applications and Implications Social media

sentiment analysis has several applications. In business, it facilitates brand monitoring, customer reanalysis and market research (Liu, 2012). Politically, it helps to gauge public opinion during elections and monitor social trends (Tumasjan et al., 2010). Socially, it helps to understand public reactions to events and monitor mental health (De Choudhury et al., 2013). The ability to analyze sentiment in real time offers significant strategic advantages in various sectors.

6. Future Directions Future research on sentiment analysis should focus on improving the accuracy and efficiency of models, especially when dealing with the dynamics and diversity of social media text. Multimodal sentiment analysis combining text, images and videos is an emerging area of interest (Zadeh et al., 2018). In addition, there is a growing effort to develop models that can understand and interpret the cultural and contextual factors that influence emotions..



3.1. The sentiment analysis method used in social media

Based on the papers reviewed, all of the paper demonstrated the usage of either Lexicon based method, Machine learning method or a mix of both method when implementing sentiment analysis. The results show in conducting sentiment analysis, 7 of the reviewed paper uses the lexicon-based method, 10 papers use machine learning and 7 papers show the combination of both methods.

Lexicon based method is known as an unsupervised learning method. Lexicon method does not require any training data and only depends on the dictionary.

Most of the study adapted Sentiwordnet and TF-IDF method when conducting sentiment analysis. This approach is calculated based on the occurrences of the terms in the text data with other positive or negative words in the predeveloped polarity lexicons like Sentiwordnet [4]. As for TF-IDF method, it works by converting the words into a number and it is calculated using the term frequency-inverse document frequency method [16].

The techniques are based on lexical resources, and the effectiveness of the whole approach strongly depends on the quality of the lexical resources. It is based on the polarity of the text, which can be obtained from the polarity of the words that make it up. Due to the complexity of natural languages, this approach is not designed to cover all aspects of language, especially when dealing with slang, sarcasm, and negation [17].

Using keywords is not enough. Some problems exist, such as some words have different meanings depending on the application, some sentences containing emotion words may not express emotion, and many sentences without emotion words may also indicate opinion [18]. However, the dictionary-based method has its own advantages, such as easy counting of positive and negative words, flexible adaptation to different languages and speed of analysis.

The machine learning method is part of supervised learning and training data is needed to process the method.

In machine learning method, the most used method is SVM and Naive Bayes model. Different machine learning models, but these are the most commonly used. Naïve Bayes is successful when applied to a corpus of well-formatted text[19], while with a support vector machine it gives good performance on a poorly formatted dataset. However, the machine learning method performs poorly on Facebook when people post random lengths and many typos, and requires a huge number of training samples to adapt the method because the size of the dataset affects the size. and output quality [18, 20]. Furthermore, analysis using machine learning is time-consuming, taking hours in a complex machine learning model, especially when training is required [11]. With a smaller training data, the process is faster, but it leads to a worse classification accuracy [21].

Interestingly, the researchers claim that both analysis methods perform very similarly in terms of accuracy [21].

Options to combine these two approaches are possible, mainly lexicon-based sentiment classification that includes a sentiment score function and Naive Bayesian multinomial event models from a machine learning approach to predict sentiment direction. Rather than relying on a single method, studies have shown that combining both methods is more effective [19]. Therefore, to improve the result, it is recommended to combine both methods, because they complement each other and the result is improved compared to using only one approach. A combined approach is valuable in detecting the phenomenon [21]. It can also improve the processing of unstructured data [22].

3.2. Type of social media platform used to collect data for sentiment analysis

Social information services or social media can be classified into four types according to the use of these applications:

Content communities (Youtube, Instagram), Social networks (Facebook, LinkedIn).), Blogs (Reddit, Quora) and Microblogs (Twitter) , Tumblr [23].

According to the revised document, among the four types of social media, microblogging sites, especially twitter, is the most popular social media platform used to collect information about users' opinions.

85% of rated articles use Twitter to collect data for sentiment analysis. Twitter is one of the 10 most visited websites and allows users to send text messages and communicate with each other. Twitter is also used to express opinions and provide

very valuable information to researchers, business organizations and even the government.

Twitter is a famous microblogging tool as a social media platform where people can express their feelings about a particular person, event or product. What makes Twitter popular is content or data that is readily available for public use. The API allows people to access and copy information on any topic they want based on keywords or hashtags.

Twitter does real-time analysis and captures public opinion because Twitter has about 500 million tweets every day and allows the public to access its data through an API [24]. Twitter is used to search and collect tweets from 8 different countries in the West and East.

Twitter has users all over the world, so it is full of opinions and views from different countries, different languages and different understandings [25].

For example, Twitter is used to collect user requests about a particular presidential candidate during an election [26] and to collect tweets written about a community development program [27].

In addition, Twitter collected messages from a customer of a British energy company [28] and analyzed tweets downloaded from the official Twitter account of London Heathrow Airport for sentiment analysis [29].

Facebook has the largest social media users in the world. But it's not very popular for sentiment analysis because the data is messy, poorly structured, and people often use short forms and lots of typos. This makes parsing the data difficult.

For example, using Facebook and Twitter are user experiences to get desirable pages, status updates and comments.

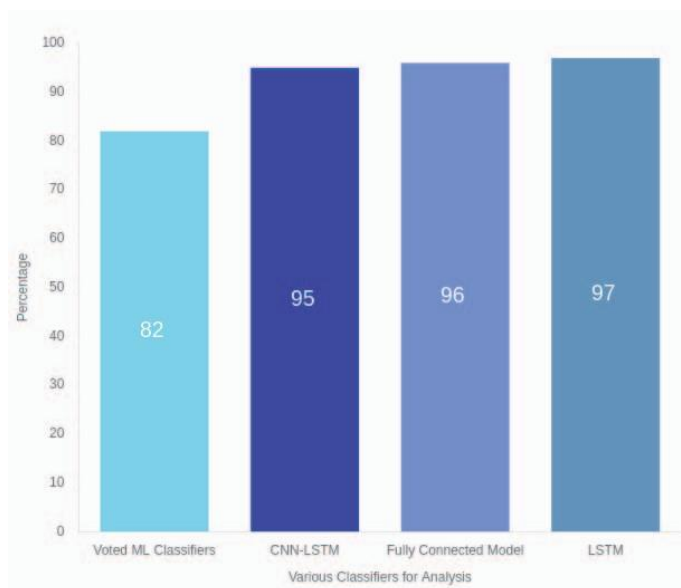
Conducted studies that collect data from various social media sites include forum, blogs, Expedia, blogspot, mainstream media, WordPress, YouTube, Twitter, aggregator, Facebook. And the result shows that 88% of the data is from Twitter [31]. Other social media sources are not recommended due to the limited amount of information or opinions it collects, such as Blogspot, YouTube and WordPress.

3.3. Application context of sentiment analysis

The application of sentiment analysis varies from business and marketing, politics, health to public activities. Sentiment analysis is not limited to one application, but offers a wide range of applications in various fields to facilitate decision making. Sentiment analysis can be applied to world events such as events, activities, sports or disasters that occur in the world [20, 32]. Some examples are a study to compare the attitudes of people in the West and the East towards ISIS. The result shows how the two sides of the world view ISIS as a terrorist [25]. Sentiment analysis can also be used to raise awareness of information security and the risk of information security breaches. It also serves as guidance for companies on how to respond to data security breaches in shaping public perception [24]. Sentiment analysis was also performed on

the social media unemployment and employment situation score [31]. We see the application of sentiment analysis in health care and where research uses sentiment analysis as a service framework, proposing and using spatio-temporal functions to identify disease hotspots [23]. In addition, mood analysis can identify people's mood needs during a disaster and prepare appropriate reactions for rescue [33]. Sentiment analysis also makes it possible to find out a person's level of depression by transcribing and analyzing sentiments from the text [19]. Sentiment analysis can be used to predict political elections, because it shows that data analyzed by Twitter is more reliable as a platform with a 94% correlation with polling data and can become a platform that can compete with advanced voting techniques. [26]. Finally, customer feedback plays an extremely important role in the application of sentiment analysis, because it can help a company and organization take the necessary steps to improve their product or service and business strategy. This can be seen in a study that derives the views and experiences of social media users on pharmaceuticals and cosmetics [30]. Sentiment analysis can also be used to identify areas that need improvement in the quality of airport services and implement appropriate corrective measures, such as paying attention to passenger feedback on social media [29]. Sentiment analysis can analyze the trends and characteristics of people's eating habits, which is useful for a business organization plans its product and marketing strategy [18]. Sentiment analysis offers advantages for business owners to determine their popularity among customers and how the customer thinks about their product or service [22] and to evaluate the effectiveness and ability of the company's brand in communication and social media [34]] and evaluate the movement of their share prices in social media by

[35]. The suggestions given by the consumer are important for us to recognize our weakness. This can be seen in a study comparing consumer tweet opinion data from the Big Six (the UK's largest and oldest gas and electricity supplier and a new energy consumer) [28]. The result shows that the opinion of the big six is more negative than that of the new energy consumer. In addition, social media opinion analysis allows an organization to evaluate the success of a program, such as in a study where a tweet about the activities of a community development program received high positive opinion. The result can help to improve the general standard of living of the community [27].



4. Conclusion

A systematic literature review informs social media sentiment analysis research. The paper contains the following three contributions. First, we show which method is used to analyze sentiment social media. Various methods have been presented in studies, but the most common lexical-based methods are SentiWordnet and TF-IDF, while in machine learning Naive Bayes and SVM. The choice of the appropriate seed analysis method depends on the data itself. Both methods showed similar accuracy. Things to consider are text structure, time, and amount of information. If the data structure is messy, data is sparse, and time for analysis is limited, a dictionary-based method is recommended. Bigger data is suitable for machine learning based method because it requires more time and data to train. To improve the quality and accuracy of the result, it is proposed to combine both vocabularies and machine learning methods. Second, we identify which is the most common social media site used to obtain data for opinion analysis. The most popular social media site for gathering information is Twitter. Most of the reviewed articles use Twitter as a social media context. This is due to the availability, accessibility and richness of content on Twitter. Every day millions of tweets on almost any topic. This shows that social media is becoming a valuable source of information. However, less attention is paid to other social media such as blogs, WordPress, YouTube and mu. The content of each social media can be different and it is worth exploring other sources that can open up new information and findings. Third, we introduce the application of sentiment analysis in social media. Sentiment analysis has a wide application and can be used in various areas such as improving business quality and strategy, predicting political election results, disease

monitoring, increasing the importance of information security, creation in certain fields. sports and disaster site improvement and response. This shows that sentiment analysis plays a huge role in understanding people's perception and helps in decision making. For future recommendations, further research is needed to develop a general opinion analysis model that can be applied to different types of data, to explore other possible social networks to obtain user opinions, and to expand the context of opinion analysis. application.

5. References

- [1] [2] Statista. (2019) Number of social media users worldwide 2010-2021. Available from: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users>. Giri, Kaiser J, and Towseef A Lone. (2014). "Big Data-Overview and Challenges." *International Journal of Advanced Research in Computer Science and Software Engineering* 4 (6). [3] Sivarajah, Uthayasankar, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. (2017) "Critical Analysis of Big Data Challenges and Analytical Methods." *Journal of Business Research* 70: 263-286. [4] Agarwal, Basant, Namita Mittal, Pooja Bansal, and Sonal Garg. (2015) "Sentiment Analysis Using Common-Sense and Context Information." *Journal of Computational Intelligence and Neuroscience* 9 (2015). [5] U. T. Gursoy, D. Bulut, and C. Yigit. (2017) "Social Media Mining and Sentiment Analysis for Brand Management." *Global Journal of Emerging Trends in e-Business, Marketing and Consumer Psychology* 3 (1): 497-551. [6] [7] Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuuttila. (2018) "The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers." *Computer Science Review* 27: 16-32. N, Mishra, and C. K. Jha. (2012) "Classification of Opinion Mining Techniques." *International Journal of Computer Applications* 56 (13). [8] Song, Minchae, Hyunjung Park, and Kyung-shik Shin. (2019) "Attention-Based Long Short-Term Memory Network Using Sentiment Lexicon Embedding for Aspect-Level Sentiment Analysis in Korean." *Information Processing & Management* 56 (3): 637-653. [9] P. Sanguansat. (2016, 3-6 Feb. 2016) "Paragraph2Vec-Based Sentiment Analysis on Social Media for Business in Thailand", in the 2016 8th International Conference on Knowledge and Smart Technology (KST). [10] Itani, Maher, Chris Roast, and Samir Al-Khayatt. (2017) "Developing Resources for Sentiment Analysis of Informal Arabic Text in Social Media." *Procedia Computer Science* 117: 129-136. [11] Chekima, Khalifa, and Rayner Alfred. (2018) Sentiment Analysis of Malay Social Media Text. pp. 205-219. 714 Zulfadzli Drus et al. / *Procedia Computer Science* 161 (2019) 707–714 Author name / *Procedia Computer Science* 00 (2019) 000–000 [12] D. Cirqueira, M. Fontes Pinheiro, A. Jacob, F. Lobato, and Á. Santana. (2018, 3-6 Dec. 2018). "A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media" in the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). [13] Peng, Haiyun, Erik Cambria, and Amir Hussain. (2017) "A Review of Sentiment Analysis Research in Chinese Language." *Cognitive Computation* 9 (4): 423-435. [14] Ebrahimi, M.m Yazdavar, A., and A. Sheth. (2017) "On the Challenges of Sentiment Analysis for Dynamic Events." *Intelligent Systems, IEEE* 32 (5). [15] Durach, Christian F., Joakim Kembro, and Andreas. (2017) "A New Paradigm for Systematic Literature Reviews in Supply Chain Management." *Journal of Supply Chain Management* Wieland 53 (4): 67-85. [16] Das, Bijoyan, and Sarit Chakraborty. (2018) An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation. [17] Khan, Muhammad Taimoor, Mehr Durrani, Armughan Ali, Irum Inayat, Shehzad Khalid, and Kamran Habib Khan. (2016) "Sentiment Analysis and The Complex Natural Language." *Complex Adaptive Systems Modeling* 4 (1): 2. [18] Akter, Sanjida, and Muhammad Tareq Aziz. (2016) "Sentiment Analysis on Facebook Group Using Lexicon Based Approach", in the 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). [19] Hassan, Anees Ul, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, and Sungyoung Lee. (2017) "Sentiment Analysis of Social Networking Sites (SNS) Data Using Machine Learning Approach for the Measurement of Depression", in International Conference on Information and Communication Technology Convergence (ICTC), Jeju, South Korea: IEEE. [20] Mahtab, S. Arafin, N. Islam, and M. Mahfuzur Rahaman. (2018, 21-22 Sept. 2018). "Sentiment Analysis on Bangladesh Cricket with Support Vector Machine", in the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP). [21] Dhaoui, Chedia, Cynthia M. Webster, and Lay Peng Tan. (2017) "Social Media Sentiment Analysis: Lexicon Versus Machine Learning." *Journal of Consumer Marketing* 34 (6): 480-488. [22] Rahman, S. A. El, F. A. AlOtaibi, and W. A. AlShehri. (2019, 3-4 April 2019). "Sentiment Analysis of Twitter Data", in the 2019 International Conference on Computer and Information Sciences (ICCIS). [23] Ali, Kashif, Hai Dong, Athman Bouguettaya, Abdelkarim Erradi, and Rachid Hadjidj. (2017) "Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework", in IEEE International Conference on Web Services (ICWS), Honolulu, HI, USA: IEEE. [24] Hao, Jianqiang, and Hongying Dai. (2016) "Social Media Content and Sentiment Analysis on Consumer Security Breaches." *Journal of Financial Crime* 23 (4): 855-869. [25] Mansour, Samah. (2018) "Social Media Analysis of User's Responses to terrorism using sentiment analysis and text mining." *Procedia Computer Science* 140: 95–103. [26] Joyce, Brandon, and Jing Deng. (2017) "Sentiment Analysis of Tweets for the 2016 US Presidential Election", in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE. [27] Yuliyanti, Siti, Djatna, Sukoco Taufik, and Heru. (2017) "Sentiment Mining of Community Development Program Evaluation Based on Social Media." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 15 (4): 1858-1864. [28] Ikoro, Victoria, Maria Sharmina, Khaleel Malik, and Riza Batista-Navarro. (2018) "Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers", in

Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 95-98): IEEE. [29] Martin-Domingo, Luis, Juan Carlos Martin, and Glen Mandsberg. (2019) "Social Media as a Resource for Sentiment Analysis of Airport Service Quality (ASQ)." *Journal of Air Transport Management*. [30] Isah, Haruna, Paul Trundle, and Daniel Neagu. (2014) "Social Media Analysis for Product Safety Using Text Mining and Sentiment Analysis", in 14th UK Workshop on Computational Intelligence (UKCI): IEEE. [31] Shayaa, Shahid, Phoong Seuk Wai, Yeong Wai Chung, Ainin Sulaiman, Noor Ismawati Jaafar, and Shamshul Bahri Zakaria. (2017) "Social Media Sentiment Analysis on Employment in Malaysia", in the Proceedings of 8th Global Business and Finance Research Conference, Taipei, Taiwan. [32] Karamollaoğlu, H., İ. A. Doğru, M. Dörterler, A. Utku, and O. Yıldız. (2018, 20-23 Sept. 2018). "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach", in the 2018 3rd International Conference on Computer Science and Engineering. [33] Ragini, J. Rexiline, P. M. Rubesh Anand, and Vidhyacharan Bhaskar. (2018) "Big Data Analytics for Disaster Response and Recovery Through Sentiment Analysis." *International Journal of Information Management* 42: 13-24. [34] Poetze, Flora, Claus Ebster, and Christine Strauss. (2018) "Social Media Metrics and Sentiment Analysis to Evaluate the Effectiveness of Social Media Posts." *Procedia Computer Science* 130: 660-666. [35] Suman, N., P. K. Gupta, and P. Sharma. (2017, 11-12 Dec. 2017). "Analysis of Stock Price Flow Based on Social Media Sentiments", in the 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS).