Journal of the Society of Automotive Engineers Malaysia
Volume 5, Issue 4 (Special Issue), pp 54-66, October 2021
e-ISSN 2550-2239 & ISSN 2600-8092

Public Awareness of the Road Safety on Twitter

H. A. A. Kamarul Aryffin[1], S. S. Syed Ahmad[*1], M. Muhammad[2], Z. Mohd Jawi[3] and K. A. Abu Kassim[3]

[1]Center for Advanced Computing Technology (C-ACT), Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia
[2]Aibots Sdn. Bhd., Menara Zurich, Taman Abad, 80300 Johor Bahru, Johor, Malaysia
[3]Malaysian Institute of Road Safety Research (MIROS), 43000 Kajang, Malaysia

[*]*Corresponding author: sakinah@utem.edu.my*

ORIGINAL ARTICLE *Open Access*

Article History:

Received
27 Oct 2020

Accepted
2 Aug 2021

Available online
1 Oct 2021

Abstract – *With the swift advancement of the Internet and information technology, there has been a rapid growth of social media. A huge volume of social awareness data that signifies the public's identification of and response to road safety is comprised in social media portals like Twitter. Evaluating and collecting data concerning social awareness to identify public cognizance concerning roads and safety initiatives is critical for the government. Strategic development, research, social funding, and enterprise functioning are some vital areas. Hence, this study proposes a technique to assess road safety awareness based on data mined from Twitter. This work indicates the aspects that reduce road safety cognizance and explores communication similarities and dissimilarities between the audiences regarding their communications on road safety. The outcomes indicate positive public results concerning road safety information disseminated by the community. Our assessment also indicates that the public interacts actively using social media, which could gradually improve road safety objectives at the community level. The study assesses the potential impact of such outcomes on community cognizance concerning road safety and provides recommendations concerning additional research agenda for the prevailing novel media scenario.*

Keywords: Road safety, text analytics, Twitter data, public awareness

1.0 INTRODUCTION

Social awareness is a sociological aspect that points to spatiotemporally affixed data offering a perspective concerning the sociability and practices followed by humans. Such practices include daily activity, social response, movement, and perception (Liu et al., 2015). Individual cognizance of social settings comprises two aspects: sensing the practical world, cognition, and response (Pentland, 2005). The proliferation of the Internet and developments in information technology have led to a social structure comprising an extensive pool of social awareness information, which is used through social media, cell phones, and network news information that express human behavior, group dynamics, and individual perception. Such social awareness information offers unprecedented insight concerning advanced quantitative perspectives about human behavior and activity trends. Such data is helpful because it could offer intelligent assistance to augment human decisions and guide group conversations (Weng et al., 2018).

Of late, road safety has been termed by the World Health Organization (WHO) as one of the fastest-growing risks to the lives of individuals worldwide. Social consciousness using disseminating info related to road safety is one of the key aspects of all road safety approaches. Road safety awareness initiatives must stay updated with the newest technologies and the latest communication tools are utilized to convey road safety messages in the best possible manner (Sujon & Dai, 2021; Ahmad et al., 2021). The media industry has witnessed an extensive transformation in recent times, and now individuals apart from professional journalists have a significant voice. The novel social media paradigm has ushered into a new communication domain where the world can create transformations in driver behaviors and make roads safer. A media initiative was devised to increase road safety awareness concerning critical safety communication in Malaysia (Ghani & Musa, 2011). Conventional media (newspapers, radio, television, magazines, and webpages) and social media (Facebook, WhatsApp, and short-duration videos) were exchanged. Media consumption varies with demographics. Public attention is strengthened when a targeted message is reinforced using several media sources. Initially, awareness campaigns were deployed using public and private media because these are regarded as legitimate. Social media reaches the masses quickly and extensively and is also cost-effective. Hence, obtaining and evaluating road safety-related social awareness information present on social media is critical for comprehending public reaction concerning safety-related messages circulating on social media. Therefore, this research is set to identify social media usage concerning road safety cognizance.

There is little information concerning the role of road safety data analysis on social media. Therefore, this study aims to find and describe messaging trends and public reaction to social-media-based road safety communication. This work builds value for the discourse by assessing if social media is a potent tool for creating and disseminating information to build public cognizance about road safety. In this endeavor, we used Twitter to gather information concerning the role of social media in creating community awareness about road safety.

With the swift advancement of the Internet and cellular networks, social media has been expanding rapidly. A huge volume of social awareness info that signifies the public's understanding and response to road safety is comprised in social media, like Twitter, Facebook, and blog posts. Because social media offers a tremendous volume of constant real real-time (Adedoyin-Olowe et al., 2014), businesses and research groups have devoted significant attention to it (Jani & Zakaria, 2021; Oyewobi et al., 2021; Maryani et al., 2020). Social media info can be utilized for stock price forecasts, avoidance of epidemics, preliminary event monitoring, election forecasts, humanitarian reliefs, crisis management, and brand management by organizations, and research groups (Wibowo et al., 2021; Aribowo et al., 2021; Mansoor & Ahmad, 2020; Saroj & Pal, 2020). Twitter is a typical example of the swift growth of social media which draws the focus of researchers (Chinnov et al., 2015). Twitter is an Internet-based social network that permits users to send and receive private updates from other contacts (texts of up to 140 characters, called "tweets"). Twitter is a social media platform that provides content distribution, mobility, ease of use, and real-time communication, allowing individuals to express themselves and present their perspectives using comments in real-time (Gu et al., 2016). Perspectives and comments are potential information sources that build public cognizance; the government and other sectors can use these to comprehend public perception (Ahmad et al., 2019). Hence, tweets hold powerful potential as crucial tools to assess social awareness. Several scholars are now interested in Twitter-based research to extract academic and business value (Shaeeali et al., 2020).

Text mining refers to gathering valuable data and processing text to get valuable information that provides potential data trends, internal associations, and correlations based on extensive unstructured and noise text sets (Kasmuri & Basiron, 2017). Keywords form the basic practical unit of the information contained in the text. Text mining has been conducted extensively using keyword occurrence patterns; co-occurrence is then studied to study new products demanded by customers (Xun & Guo, 2017), and to identify and predict promising technology for social advertisement and campaigns (Xiong et al., 2019; Hussain et al., 2021). Twitter is a platform where the media, enterprises, authorities, and others can express their view concerning a specific technology (Ding et al., 2021; Essien et al., 2021; Agarwal et al., 2018). Tweets comprise technological inventions, academic findings, breakthrough developments, innovation by businesses, and public response. The domain and emotional trends contained in such data are critical for understanding public cognizance of road safety. Grammatical building blocks like adjectives, nouns, verbs, and adverbs contained in tweets highlight the essential content and emotional response (Naskar et al., 2020). Tweets typically present the subject using nouns, while adjectives primarily express emotions. Tweets spread rapidly and extensively because they are straightforward, fragmented, efficient, and real-time (Kraaijeveld & De Smedt, 2020). The specifics, comments, and perspectives concerning road safety information and awareness-building keywords strongly influence community behavior. Hence, this study correlates nouns and adjectives with awareness. Concurrently, the mining technique uses awareness keywords to be used together; keyword frequency is evaluated to assess social cognizance concerning road safety data gathered from Twitter.

This research work indicates that the word cloud might be an effective “infographics”, which are “visual representations of information, data or knowledge that present complex information quickly and clearly” (Smiciklas, 2012). Also, “dashboards” are “a graphical user interface that organizes and presents information in an easy-to-read format”. This work provides baseline outcomes concerning social media evaluation using word cloud assessment. This study is based on real tweets to gather perspectives concerning (1) the nature of social response elicited by road safety plans; (2) using a word cloud to process the text posted by the public, and (3) understanding region-specific information and patterns emerging from individual discussions concerning road safety. This study offers a real-world perspective concerning social media use, vital for additional research to obtain an understanding of the potency of social media to augment road safety cognizant and drive impact on locals. The subsequent aim is to determine if social media is productive for building and disseminating a public plan.

We start by offering an overview concerning the works on social media analytics. We then specify the word cloud assessment technique; lastly, we present the outcomes of the social media use trends determined using this work. We conclude with a discussion on study outcomes, drawbacks, and future directions.

2.0 METHODOLOGY

To analyze the social awareness about road safety, this research paper puts forward a social awareness analysis method by considering Twitter data mining. The key idea here is that tweet data regarding road safety are regarded as a data resource, in which we can apply text mining and social network analysis for efficiently extracting and analyzing road safety awareness towards the public, with an aim to educate regarding 'visualization' and 'knowledge discovery' focusing on road safety and social awareness. Data collection is the first step in this study. With the help of Twitter API, retrieving road safety information using keywords can be done.

There were three data preparation steps in this study: data collection, sampling, and raw data pre-processing. To initiate data processing, a series of functions were performed on the tweet texts for removing emojis, URLs, retweets, special characters, hyperlinks pointing to Websites, and hash symbols. This process helped to mention related diseases that could have contaminated the results. We also eliminated stop words in English (e.g., the, for, is) and words like 'Google drive' or *'jalan cerita'* that could have referred to other topics. Also, we changed the tweet texts to lower case, so the AI algorithm classifies "Car" and "car" as the same thing. Then, the tweets were transformed into a corpus (text mining structure). A document-term matrix was created and performed calculation for the Term Frequency – Inverse Document Frequency (TF-IDF), which refers to numerical statistics employed in reflecting the significance of a word in a corpus. To get the scenario's output, an analysis of the tweet data was carried out, while the Twitter API was employed for extracting the tweet data.

Word clouds are helpful to carry out analysis of any kind of text data, such as short answers, essays, or written responses to a survey or opinion questions. In this section, various examples are provided for word clouds to understand the level of public awareness about road safety by evaluating the use of keywords with regards to road user short answer responses.

Word cloud can be defined as a text mining technique for focusing on the most frequently employed word in any text of interest. Word cloud can be created based on various methods of text processing and can be performed as the followings: standardization, tokenization, cleansing, stemming, removing the common word, drawing word cloud, and indexing.

Tokenization involves the splitting of each sentence from the entire text into several words. White-space can be employed as a word delimiter for performing this step. The following general structure can be found in the manuscript:

1. Normally, the process of standardization involves changing the entire characters used in all the documents into lower case.
2. Cleansing helps to eliminate certain special characters and symbols like a question mark, dollar signs, and punctuation.
3. Indexing allows to create of a list of all unique words, which helps in counting the frequency of each word that appears in all texts.
4. In the drawing process, Python Word Cloud is used to illustrate word clouds. The favorable words are presented randomly in a different position and distributed via representations as different shapes, such as rectangular, sphere, or ellipse, and are colored randomly.

5. The size of the font represents the frequency of occurrence of the word. The user can determine the font size for a word with a smaller frequency. For accustoming the font size to larger frequency, linear normalization is employed as follows.

$$S = Max\ S\ \times \frac{(C - Max\ C)}{(Max\ C - Min\ C)} \qquad (1)$$

Where S is the size of the font, $MaxS$ is the maximum size of the font, C is word count frequency, $MaxC$ is maximum word count frequency and $MinC$ is minimum word count frequency.

In general, words that occur together in the same document must be clustered together as a single group to ease the process of understanding. To perform word clustering, we need to conduct the term-document matrix. As elements of data, this matrix would include word frequency. From all the documents, the list of unique words can be outlined as a row of the matrix, while the list of documents as a column. This process was found to be remarkably different from documents clustering which requires building the matrix in the opposite definition of column and row. The word clustering was done using K-Means clustering and an optimal number of word clusters for positive and negative text was decided using the elbow method.

3.0 RESULTS AND DISCUSSION

Post data gathering, text analysis was carried out to evaluate the most common phrases employed in data about awareness about road and traffic safety. From this collection of sentences, a list of texts was gathered. To minimize biases in the process, various words were excluded, including stop words, as the words are commonly used in a language.

To segment the data into the respected cluster, it is required that the negative text data and positive text data pass through the K-Means clustering algorithm to get an optimal number of clusters by determining the sum of squared distance between those clusters. After the run is performed, the result is seen below where both negative and positive data include an optimal cluster number of 9. After identifying the best number of clusters, both negative and positive text data were run through K-Means clustering to get word clusters. In the next step, processing of the word clusters was done to build Word Cloud.

Figure 1 displays a word cloud about all positive texts not displaying any obvious pattern. Post segmentation of the positive text data into 9 clusters, a certain obvious pattern could be observed. As seen in Figure 2, the first- and third-word clouds signify noise or mistakes arising from the AI. The first cluster refers to the accident that occurred while the third cluster points to mentions speeding or driving carelessly.

Figure 3(a) displays the second cluster referring to traffic lights, while Figure 3(b) refers to the fourth cluster mentioning the talks regarding driving safely or wishing the driver for a safe drive. Meanwhile, in Figure 4, the fifth, seventh, and ninth clusters point out mentions about smooth and unblocked traffic. The sixth cluster demonstrates the word cloud referring to under control *(lancar terkawal)*. The eighth cluster refers to the highway and does not include any other obvious details.

Figure 1: Word Cloud for all positive text

(a) (b)

Figure 2: Word Cloud for positive text (a) cluster 1 and (b) cluster 3

(a) (b)

Figure 3: Word Cloud for positive text (a) cluster 2 and (b) cluster 4

(a) (b) (c)

Figure 4: Word Cloud for positive text (a) cluster 5 (b) cluster 7 and (c) cluster 9

Figure 5: Word Cloud for all negative text

(a) (b) (c)

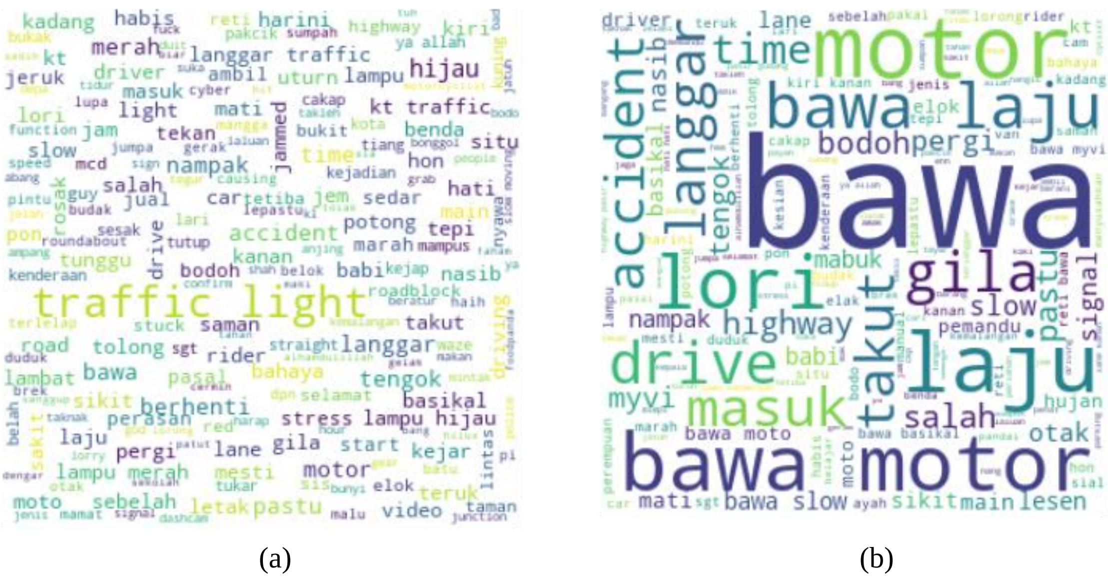
Figure 6: Word Cloud for negative text (a) cluster 1 (b) cluster 3 and (c) cluster 8

(a) (b)

Figure 7: Word Cloud for negative text (a) cluster 2 and (b) cluster 4

With regards to the negative text, Figure 5 shows a word cloud about all negative text that primarily refers to accidents and driving. As shown in Figure 6, the first, third and eighth negative text clusters refer to mentions regarding collisions and accidents. Figure 7(a) shows the second negative text cluster referring to the discussion on slowed traffic. Meanwhile, Figure 7(b) shows the fourth cluster mentioning floods, buses, and difficulty in driving.

Figure 8(a) shows the fifth negative text cluster that refers to the traffic light. Next, as seen in Figure 8(b), the sixth negative text cluster talks about speeding motorbikes. As displayed in Figure 9(a), the seventh negative text cluster refers to talking on driving through a jammed road. Finally, Figure 9(b) displays the ninth negative text cluster referring to traffic blocks.



(a) (b)

Figure 8: Word cloud for negative text (a) cluster 5 and (b) cluster 6

(a) (b)

Figure 9: Word cloud for negative text (a) cluster 7 and (b) cluster 9

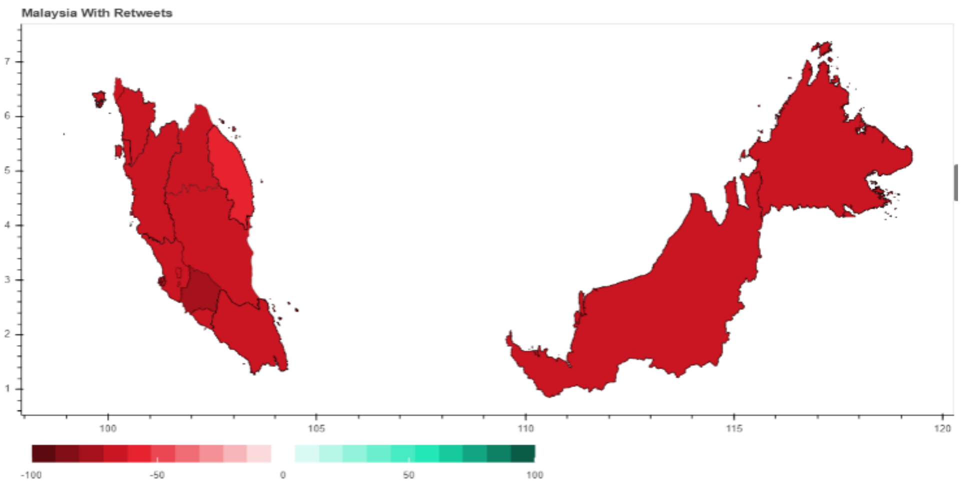


Figure 10: Location that produces tweet

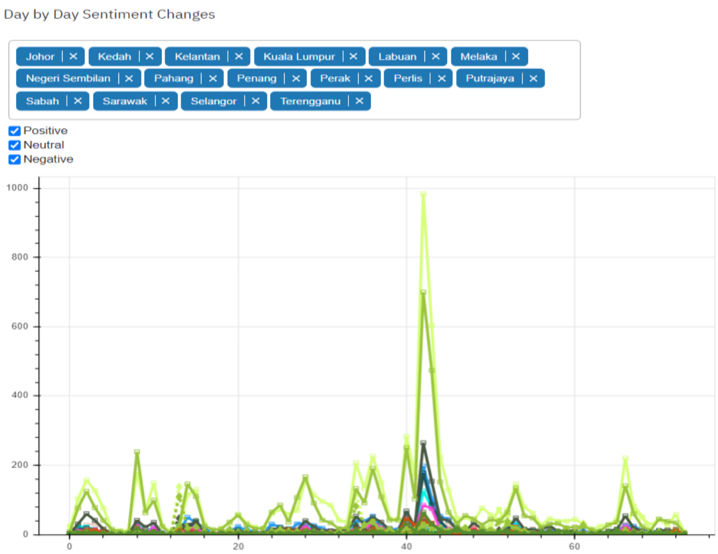


Figure 11: Graph for sentiment vs. day

The tweets are also processed to get the location of the person that produces the tweet. This can be done by utilizing Twitter API. The tweets that are divided into positive and negative tweets provide a (+1) or (-1) score. Based on the score, the percentage was calculated to classify whether the tweets from State A have 100% negativity or 100% positivity. Based on Figure 10, most states have higher negative tweets than positive tweets with Negeri Sembilan as the state with the most negative tweets and Terengganu as the state with the most positive tweets. Figure 11 shows the sentiment changes by the day. This information is good to understand the public awareness about road safety by day. Moreover, further investigation can be done to find the important parameter or event that has a significant influence on the positive or negative tweets from the public.

4.0 CONCLUSION

In the past decades, the evolution of public awareness has impacted the way of designing and implementing road and traffic safety initiatives. Few of the most significant elements of all road safety strategies include communication and education via the dissemination of road safety information. Thus, road safety communicators need to keep up with the latest technologies and employ the latest communication tools for effectively putting across road safety messages. In recent years, the media industry has transformed significantly and provided a voice to other persons and not just professional journalists. Platforms, like social media, have created an entirely new world of communication that would help in changing the behaviors of the driver and improving road safety.

Tweets published on Twitter are mostly the author's comments and opinions, and these tweet texts possess large quantities of the public's sensing and awareness data regarding road safety. These awareness data help to stimulate as well as drive the participation behavior of road users, which can impact or change the formation as well as awareness with regards to road safety. These social awareness data can be analyzed and mined to identify the trends in road safety awareness. Due to these reasons, this paper proposes a method for analyzing social awareness about road safety by considering Twitter data mining. First, we have employed the Natural Language Processing (NLP) module for extracting adjectives and nouns that could indicate social sensing as well as response towards road safety in the text of the tweet. Second, for analyzing and extracting information about social awareness, we have used text mining as well as word cloud analysis. Finally, as an empirical study, we applied the sentiment analysis for confirming the effectiveness and feasibility of the proposed method.

Our successful employment of text analytics confirms its ability to efficiently produce and codify rich and quality knowledge that can be used by researchers to understand the impact of social media more effectively. However, this also has certain limitations and could be used as a key starting point to conduct future research. Thus, future studies can make use of text analytics by covering a longer period to assess trends as well as to include more resources for building robustness in their findings. For example, they can make use of various sources of unstructured data, like blogs and Facebook, which can also be employed to analyze social awareness. We can also integrate these data with Twitter data to get a better understanding of the social awareness about road safety. With regards to the methodology, a semantic theme model analysis, as well as emotional analysis, could be utilized for exploring the public's attitudes, emotions, and responses towards emerging technologies, which can be found in such multi-source heterogeneous data. It is key to mining and analyzing social awareness regarding road safety for gaining a deeper understanding regarding road safety awareness through public participation. Social media also offers a chance for the general public to get involved in discussions about road safety and allow them to share their suggestions and concerns, which could be vital for government R&D strategic planning, best practices, and social investment.

ACKNOWLEDGEMENTS

The authors express their deepest gratitude to ASEAN NCAP Collaborative Holistic Research (ANCHOR) III (UTeM: ANTARABANGSA-ANCHOR/2020/FTMK-CACT/A00032 & ANCHOR: C3X-2) for supporting this research.

REFERENCES

Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F. (2014). A survey of data mining techniques for social media analysis. Journal of Data Mining & Digital Humanities, 2014 (March 2015).

Agarwal, S., Mittal, N., & Sureka, A. (2018). Potholes and bad road conditions: Mining Twitter to extract information on killer roads. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (pp. 67-77).

Ahmad, S., Anis Naseerah, S. O., & Halizah, B. (2019). The impact of social media on human interaction in an organisation based on real-time social media data. International Journal of Data Science, 4(3).

Ahmad, S., Muhammad, M., & Jawi, Z. M. (2021). Public awareness of traffic safety based on data and text analytics. Journal of the Society of Automotive Engineers Malaysia, 5(1), 103-116.

Aribowo, A. S., Basiron, H., Yusof, N. F. A., & Khomsah, S. (2021). Cross-domain sentiment analysis model on Indonesian YouTube comment. International Journal of Advances in Intelligent Informatics, 7(1), 12-25.

Chinnov, A., Kerschke, P., Meske, C., Stieglitz, S., & Trautmann, H. (2015). An overview of topic discovery in Twitter communication through social media analytics. 2015 Americas Conference on Information Systems, AMCIS 2015, 1-10.

Ding, Y., Korolov, R., Wallace, W., & Wang, X. (2021). How are sentiments on autonomous vehicles influenced? An analysis using Twitter feeds. Transportation Research Part C: Emerging Technologies, 131, 103356.

Essien, A., Petrounias, I., Sampaio, P., & Sampaio, S. (2021). A deep-learning model for urban traffic flow prediction with traffic events mined from Twitter. World Wide Web, 24(4), 1345-1368.

Ghani, Y., & Musa, M. (2011). Media exposure for road safety communication campaign: A case study in Malaysia. 2011 International Conference on Social Sciences and Society (ICSSS 2011), Vol 1, October, 417–422.

Gu, Y., Qian, Z., & Chen, F. (2016). From Twitter to detector: Real-time traffic incident detection using social media data. Transportation Research Part C: Emerging Technologies, 67, 321-342.

Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., Ali, A., & Sheikh, A. (2021). Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: Observational study. Journal of Medical Internet Research, 23(4), 1-10.

Jani, N., & Zakaria, M. H. (2021). Early identification of customer engagement issues in relation to social media measurement. Journal of Contemporary Social Science and Educational Studies, 1(1), 1–10.

Kasmuri, E., & Basiron, H. (2017). Subjectivity analysis in opinion mining - A systematic literature review. International Journal of Advances in Soft Computing and Its Applications, 9(3), 132–159.

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions and Money, 65, 101188.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., ... & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. Annals of the Association of American Geographers, 105(3), 512-530.

Mansoor, A. S., & Ahmad, M. (2020). The role of social media use in social coordination among relief local organizations during response to humanitarian crisis in Yemen. Jurnal Pengajian Media Malaysia, 22(1), 51-68.

Maryani, E., Rahmawan, D., & Karlinah, S. (2020). The implications of social media on local media business: Case studies in Palembang, Manado and Bandung. Jurnal Komunikasi: Malaysian Journal of Communication, 36(1), 317-333.

Naskar, D., Singh, S. R., Kumar, D., Nandi, S., & De La Rivaherrera, E. O. (2020). Emotion dynamics of public opinions on Twitter. ACM Transactions on Information Systems, 38(2).

Oyewobi, L. O., Adedayo, O. F., Olorunyomi, S. O., & Jimoh, R. (2021). Social media adoption and business performance: The mediating role of Organizational Learning Capability (OLC). Journal of Facilities Management, 19(4), 413-436.

Pentland, A. (2005). Socially aware, computation and communication. Computer, 38(3), 33-40.

Saroj, A., & Pal, S. (2020). Use of social media in crisis management: A survey. International Journal of Disaster Risk Reduction, 48(March), 101584.

Shaeeali, N. S., Mohamed, A., & Mutalib, S. (2020). Customer reviews analytics on food delivery services in social media: A review. IAES International Journal of Artificial Intelligence, 9(4), 691-699.

Smiciklas, M. (2012). The power of infographics: Using pictures to communicate and connect with your audiences. Que Publishing.

Sujon, M., & Dai, F. (2021). Social media mining for understanding traffic safety culture in Washington state using Twitter data. Journal of Computing in Civil Engineering, 35(1), 04020059.

Weng, S. S., Yang, M. H., & Hsiao, P. I. (2018). A factor-identifying study of the user-perceived value of collective intelligence based on online social networks. Internet Research, 28(3), 696-715.

Wibowo, A., Chen, S. C., Wiangin, U., Ma, Y., & Ruangkanjanases, A. (2021). Customer behavior as an outcome of social media marketing: The role of social media marketing activity and customer experience. Sustainability, 13(1), 1-18.

Xiong, Y., Cho, M., & Boatwright, B. (2019). Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. Public Relations Review, 45(1), 10-23.

Xun, J., & Guo, B. (2017). Twitter as customer's eWOM: An empirical study on their impact on firm financial performance. Internet Research, 27(5), 1014-1038.