Opinion Mining and Sentiment Analysis for Classification of Opinions on Social Networking Sites Using Machine Learning Algorithms: Systematic Literature Review

Rayees Ahmad[1], Dr. Yasmin Shaikh[2]

[1]Research Scholar, International Institute of Professional Studies DAVV, Indore (M.P) India

[2]Assistant Professor (Senior Scale), International Institute of Professional Studies DAVV, Indore (M.P) India

Abstract: This paper is a report of a review on Sentiment Analysis for classification of Opinions on social Networking sites that explored the methods, social media platform used and its application. . In this a generic deep learning framework for predictive analytics utilizing both structured and unstructured data will be presented. The social media, blogs, forums, e-commerce web sites, and so on encourages people to share their opinion, emotions and feelings publicly. Today these Internet Sites are very popular and this resulted a huge amount of raw data that has been uploaded by users in the form of text, photos, audio and videos. People's sentiments and experiences are very valuable information in decision making process of any business but to get benefit from these opinion and experiences, the accumulated content should be extracted and analysed properly. This can be done by using sentiment analysis. A systematic review of studies published between 2015-2020 was undertaken using the following trusted and credible database including IEEE Xplore, ACM, Emerald Insight, Science Direct and Scopus. After the initial and in-depth screening of paper, 24 out of 85 articles have been chosen from the review process. The articles have been reviewed based on the aim and objectives of the study. The result shows most of the articles applied opinion-lexicon method to analyses text sentiment in social media, extracted data on microblogging site mainly Twitter and sentiment analysis application can be seen in world events, as improving quality and strategy in business, Market Research, Decision Making, political forecasting an election result, monitor disease outbreak, create awareness on the importance of data security, perception towards a particular spot, and improve locate and response to the disaster.

Keywords: Opinion, Sentiment, Machine learning (ML), Natural Language Processing (NLP), Social media.

I. INTRODUCTION

With the big evolution of social networking sites such as Facebook, Instagram, Twitter, Blogs, and Forums etc. the web has been emerged into a huge volume of user reviews and opinions about particular aspects of products or services. People like to share and post their experiences, thoughts, opinions, feelings, and preferences according to their understanding and observation about the services. Their point of view or impression may be positive, negative or neutral. This opinion is used for identifying trends, user interest, and prediction of stock markets, political polls, and market researches, enhancing the user experience by presenting the things of their own interest and to influence them towards a particular direction. Such data is very useful for decision makers or policy makers of the country since it is impossible for the policy makers to interact with the mass and collect their opinions effectively. For one particular aspect, one may have a positive opinion while some other may have a negative opinion at the same time. Thus, classifying opinion and sentiment of people is a difficult task. Furthermore, the shared reviews and feelings are not in specifically structured format, thus identifying its positivity or negativity perspective automatically, is also convenient. Therefore, analysis of an unstructured format of text and extract the information for determining the user's sentiments requires special machine learning techniques and semantic algorithms for their classification. In sentiment analysis, major tasks listed are subjectivity and sentiment classification, sentiment lexicon generation, opinion spam detection and quality of reviews.

Sentiment Analysis (also known as *opinion mining* or *emotion AI*) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify and study affective states and subjective information. It is an extension of data mining, which utilizes natural language processing techniques to extract people's opinion from World Wide Web. Sentiment analysis (SA) or opinion mining (OM) is the computational study of people's opinions, sentiments, attitudes and emotions toward an entity. It focuses on developing automatic systems that can analyse natural language texts to determine the sentiment expressed in them. The word Sentiment is often using in wide sense to refer to expressions of subjectivity, opinion, attitude, affect, orientation, emotions, and

feelings and tone in the. Sentiment analysis is classified into positive comment, negative comment or neutral comment. Fig. 1 shows the complete process of sentiment analysis that refers how the input is being classified on the various steps.
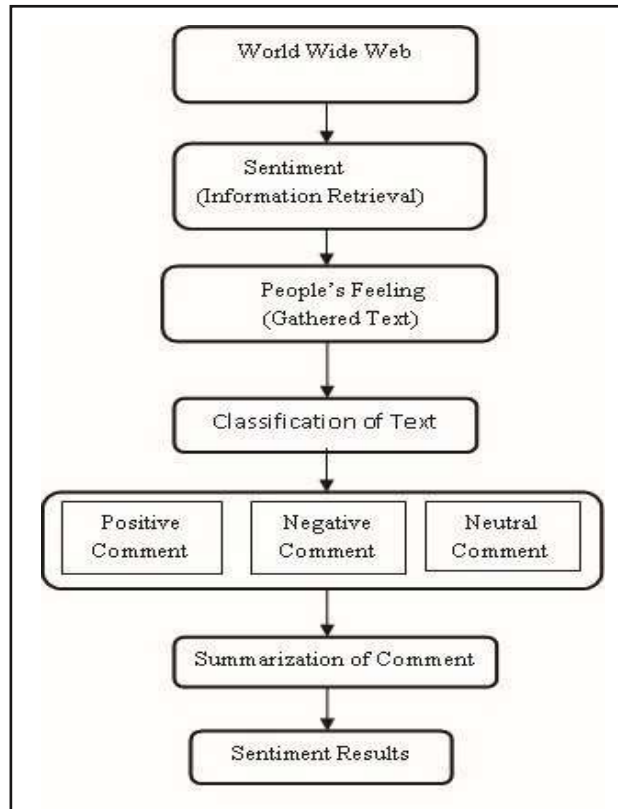

Fig. 1: Process of Sentiment Analysis

The sentiment analysis summarization process contains three main steps,
- Sentiment information retrieval: Review text is retrieved from review websites such as Twitter, Facebook, Amazon and News Sites etc.
- Sentiment classification: classification of Sentiment text (contains subjective information about the topic or issue.) in the blog, reviews, comments, microblogs etc. into positive and negative classes.
- Sentiment summarization: Sentiment results are generated based on features (sentiment sentences) selection about a matter.

II. THE REVIEW

2.1. Design
A systematic review was undertaken using 6 steps guidelines for conducting a systematic literature review in management [15]. First, we start by defining the research question. Then determine the required characteristic for the study. Continue by retrieving potentially relevant literature and selecting pertinent literature. We then synthesize relevant information from the literature and the final step is reporting the result of the review.
2.2. To provide an overview of the review, the following research question was addressed:
RQ1: What is the method used in sentiment analysis of social media?
RQ2: What is the type of social media platform used to applied sentiment analysis?
RQ3: What is the application context of sentiment analysis in social media?
2.3. Retrieving and selecting pertinent literature
The review utilizing five reputable and credible online databases that published literature covering information and computer science area. Search strings keywords used for all five online databases is "Sentiment analysis, social media, Facebook, Twitter". The total articles identified from the database search is 307 articles. 34 articles identified from Emerald Insight, 242 results identified from science direct, 24 results from Association for computing machinery (ACM), 45 articles from Scopus and 45 articles identified from IEEE.

Then the screening of papers is conducted based on the inclusion and exclusion criteria and the screening resulted in 85 articles. Consequently, the screening involved reading the full texts and analysing each article and we obtain 24 finalized articles.

2.4. Synthesizing the literature

The studies were published between 2015 to July 2020. There is a total of 24 articles selected that suits the purpose of this review. The data from the paper is extracted and the primary study findings are analysed and integrated into a Table 1.

Table 1. Summary of reviewed literature.

S.No	Author	Title	Method / Tools	Application / Result	Context
1	Isah, Trundle & Neagu. (2014)	Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis	Lexicon-based and machine learning	Monitor brand in order to act in even of a sudden rise in negative sentiment.	Facebook comment and Twitter
2	Hao & Dai. (2016)	Social media content and sentiment analysis on consumer security breaches	Lexicon-based	Security breaches can be detected in the early stages and prevent further destruction	Twitter
3	Akter, Aziz &Tareq (2016)	Sentiment analysis on Facebook group using the lexicon-based approach	Machine learning	Determine recent trends and characteristics of people food habit.	Facebook Group – Foodbank
4	Yuliyanti, Djatna & Sukoco. (2017)	Sentiment Mining of Community Development Program Evaluation Based on Social Media	Lexicon-based and machine learning	Success level of the community development program	Twitter
5	Saragih & Girsang. (2017)	Sentiment Analysis of Customer Engagement on Social Media in Transport Online	Lexicon-based	Evaluate the business performance of online transport.	Facebook and Twitter comments
6	Hassan, Hussain, Husain, Sadiq, Lee. (2017)	Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression	Machine learning	Find the depression level of a person	Twitter and newsgroup
7	Joyce & Deng. (2017)	Sentiment Analysis of Tweets for the 2016 US Presidential Election	Lexicon-based and machine learning	Calculate sentiment expressed and compare with polling data to see the correlation	Twitter
8	Shayaa, Wai, Chung, Sulaiman, Jaafar & Zakaria. (2017)	Social Media Sentiment Analysis on Employment in Malaysia	Lexicon-based	Negative sentiment score on employment	Multiple channel social media
9	Ali, Dong, Bouguettaya, Erradi & Hadjidj. (2017)	Sentiment Analysis as a Service: A social media-based sentiment analysis framework	Machine Learning	Identify the location of disease outbreaks	Twitter, Reddit, Instagram, news forum.

10	Chedia Cynthia & Tan (2017)	Social media sentiment analysis: lexicon versus machine learning	Lexicon-based and Machine learning	Sentiment analysis on consumer generated content	Facebook brand pages
11	Fatyanosa & Bachtiar (2017)	Classification method comparison on Indonesian social media sentiment analysis	Lexicon-based and Machine learning	Sentiment on Jakarta governor election	Twitter
12	Shahare (2017)	Sentiment analysis for the new data based on social media	Machine learning	Process and identify emotion level from news data	News from blogs
13	Suman, Gupta & Sharma (2017)	Analysis of Stock Price Flow Based on Social Media Sentiments	Machine learning	Relate the flow of stock price	Stock Twists
14	Mansour. (2018)	Social Media Analysis of User's Responses to terrorism using sentiment analysis and text mining	Lexicon-based	Most user view ISIS as a threat and fear	Twitter
15	Ikoro, Harmina, Malik, & Batista-Navarro. (2018)	Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers	Lexicon-based	Analyze energy provider company and the sentiment that users show	Twitter
16	Mahtab, Islam & Rahaman (2018)	Sentiment Analysis on Bangladesh Cricket with Support Vector Machine	Lexicon-based and machine learning	Analyze people sentiment expressed towards cricket	Facebook Group – Bangladesh Cricket
17	Abd El-Jawad, Hodhod & Omar (2018)	Sentiment Analysis of Social Media Networks Using Machine Learning	Machine Learning	System to provide insight on how people perception	Twitter
18	Karamollaoğlu, Doğru, Dörterler, Utku & Yıldız (2018)	Sentiment Analysis of Turkish Social Media Shares through Lexicon Based Approach	Lexicon-based	Measure the perception or influences of the phenomena	Twitter
19	Poecze, Ebster, Strauss & Christine (2018)	Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts	Machine learning	Optimize brand communication and understanding consumer feedback	Facebook page of YouTube Gamers
20	Ragini, Anand & Bhaskar (2018)	Big data analytics for disaster response and recovery through sentiment analysis	Lexicon-based and machine learning	sentiment towards the needs of affected people during disaster	Twitter
21	Vishal & Uma (2018)	An Extensive study of Sentiment Analysis tools and Binary Classification of tweets using Rapid Miner	Machine learning	Identify an efficient tool which can help an enterprise	Twitter

22	ElRahman, AlOtaibi & AlShehri (2019)	Sentiment Analysis of Twitter Data	Machine learning	Popularity between two restaurant – KFC and Mcdonalds	Twitter
23	Ramanathan & Meyyappan (2019)	Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism	Lexicon-based	Feedback on Oman tourism	Twitter
24	Olga Artemenko, Volodymr, Nataliia, Khrystyna (2020)	Using sentiment text analysis of user reviews in social media for E-Tourism Mobile Recommender systems	Lexicon-based and machine learning	Process and identify user generated big data in mobile E-tourism recommender systems	Twitter

III. RESEARCH METHODOLOGY

The existing work on sentiment analysis can be classified as the problem of text classification in different forms such as Document-level sentiment analysis, Sentence-level sentiment analysis, and Aspect-based sentiment analysis. The purpose is to classify the opinion either from sentence, document or features into positive and negative sentiment [7].

Classification of OM at various levels	Assumptions made at different levels	Tasks associated with various levels
1. OM at Sentence level.	1. Sentences comprise solely one opinion expressed by one opinion holder. This might not be true in a lot of situations. 2. Sentence boundaries are delineated in the specified document.	**Task 1:** Identification of specified sentence as subjective or opinionated. **Classes**: objective as well as subjective. **Task 2:** opinion classification of the specified sentence. **Classes:** positive, negative as well as neutral.
2. Opinion Document level	1. All documents focus on a particular object as well as comprise opinions conveyed by a particular opinion holder. Not applicable for blogs as well as forum posts because there might be several opinions on various objects in such sources.	**Task 1:** opinion classification reviews **Classes:** positive, negative, as well as neutral.
3. Opinion Feature level.	1. The data sources focus on attributes of one object conveyed by one opinion holder. Not applicable for blogs as well as forum posts because there might be several opinions on various objects in those sources.	**Task 1:** Identification as well as extraction of object attributes, which have been commented on by opinion holders. **Task 2:** Determination of whether the opinions on the attribute are positive, negative or neutral. **Task 3:** Grouping of features synonyms. Present features-based summaries of several reviews.

Study of the evolution of sentiment analysis research can be done by the analytical tokens, or building blocks and the implicit information associated with those tokens. The existing approaches can be grouped into four main types they are keyword spotting approach, lexical affinity method, statistical methods, and concept-based techniques.

A.   *Keyword spotting*

This approach makes use of unambiguous words such as happy, sad, afraid, and bored to classify the text into different categories. "Today was a great day" as being affectively positive, it is likely to assign the same classification taking an example of sentence like "Today wasn't a great day at all." Keyword spotting also relies on the presence of actual effect of words and not for a sentence. Sometimes, a sentence conveys through their meaning rather than affect adjective words. For example, the following sentence "My husband just wants to take custody of my children away from me" states strong emotions, but uses no affect keywords, and therefore is ineffective. Lexical affinity is slightly more sophisticated than keyword spotting.

B. *Lexical affinity*

This approach not just detects obvious affect words, but also assigns arbitrary words a probable "affinity" to particular emotions. For example, lexical affinity might assign the word "accident" a 75-percent probability of indicating a negative effect, as in "car accident" or "hurt by accident." This approach usually trains probability from linguistic corpora. Although it often outperforms pure keyword spotting, there are two main problems with this approach. First, negated sentences (I avoided an accident) and sentences with other meanings (I met my friend by accident) trick lexical affinity, because they operate solely on the word level. Second, lexical affinity is often biased towards a category, directed by the linguistic corpora's source. Due to which it becomes difficult to build and reuse the model for different systems.

C. *Statistical methods*

This approach is popular for affect text classification. Researchers use statistical methods on projects, by using a machine-learning algorithm for a large training corpus of affectively annotated texts, the system will learn combination of things as in keyword spotting and lexical affinity. Generally, statistical methods are semantically weak, which means that individually with the exception of obvious affect keywords a statistical model's other lexical or co-occurrence elements have little predictive value. Due to which statistical text classifiers give accurate results only when they are given sufficiently large amount of data as input. So, these methods will affectively classify the reviews or opinion text on large for example page or paragraph level, they don't perform well when it comes to smaller texts such as clauses or sentences.

D. *Concept-based approaches*

To accomplish semantic text analysis methods such as semantic networks and web ontology are used. This helps the approach to identify the affective and conceptual data related to natural language opinions. In this approach instead of blindly using the matches of keywords and counts of word co-occurrence by using the large semantic knowledge dictionaries, and rather depend on the actual meaning associated with natural language contexts. Concept- based approaches can identify minute expressed sentiments when compared to purely syntactical techniques. Concept-based approaches can analyze multi-word expressions that don't explicitly convey emotion. The concept-based approach relies heavily on the depth and breadth of the knowledge bases it uses. Without a comprehensive resource that encompasses human knowledge, an opinion mining system will have difficulty grasping the semantics of natural language text.

IV. CONCLUSION

In this paper, Sentiment analysis and opinion mining using intelligent techniques was proposed to deal with social media data. It has been observed that various techniques can be used to achieve a sentiment analysis on social media data and others. There are various methods introduces by researchers, still, the most common method used in lexicon method is Senti Wordnet and TF-IDF While for Machine learning is Naïve Bayes and SWM.

 Now a days, sentiment analysis is one of the hot research areas for the researchers. The information gathered from the online data sources like blogs, microblogs, forums, reviews sites, e-commerce websites, etc. has been playing an important role in expressing people's feelings, thoughts, emotions, and opinions for the particular topic, event or issue. Here we identify that the most of the reviewed papers use twitter as their social media context to extract information. This is due to the availability, accessibility and richness of twitter content.

In this paper a cognitive study or sentiment analysis techniques and tools has been presented. The proposed methodology provides important phases the sentiment of text whether it is positive or negative. This paper will be helpful to the researchers of the sentiment analysis domain.

REFERENCES

[1] Statista. (2019) Number of social media users worldwide 2010-2021. Available from: https://www.statista.com/statistics/278414/number-ofworldwide-social-network-users.

[2] 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]   Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuutila. (2018) "The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers." Computer Science Review 27: 16-32.

[7]   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.

[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., & 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. 9598): 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]   Poecze, 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).

[36]   Morinaga, S., Yamanishi K., Tateishi, K. And Fukushima, T., ―Mining Product Reputations on the Web. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2002.