International Journal of Industrial Engineering
2023;7(1):1-6.
ISSN: 2456-8449

Sentiment Analysis in Social Media Networks

A. Gayathri[1*], S. Yuvarani[2], P. Srinivasan[3], K. Jayasakthi Velmurugan[4]

[1]*Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, Tamil Nadu, India.*
[2]*Department of Computer Applications, Thanthai Hans Roever College (Autonomous), Perambalur, Tamil Nadu, India.*
[3]*Department of Electronics and Communication Engineering, Sona College of Technology, Salem, Tamil Nadu, India.*
[4]*Department of Computer Science and Engineering, Jeppiaar Engineering College, Chennai, Tamil Nadu, India.*

[*]*Corresponding author: gayathribala.sse@saveetha.com*

Abstract. After the growth of social media platforms like Twitter, there has been a marked increase in the quantity of textual information available online, including news stories and historical records. More individuals are using the internet and different forms of social media to share their thoughts and emotions with the world. Because of this, there has been a rise in the quantity of user-generated phrases that express emotions. It's natural that researchers will investigate new approaches to understanding people's emotions and reactions. In addition to providing novel hybrid systems that combine text mining and neural network for sentiment categorization, this research evaluates the efficacy of many machine learning and deep learning techniques. More than a million tweets from across five different topics were utilized to create this dataset. The datasets were split that 75% was used for training and 25% was used for testing. The findings reveal a maximum accuracy rate of 90%, demonstrating the superiority of the system's hybrid learning technique over conventional supervised methods.

Keywords: Sentiment analysis, social media, Machine learning, big data

INTRODUCTION

The consistent interest in deep learning methods for multilingual sentiment analysis of social media is shown by twenty-four research covering twenty-three various languages and eleven various social media platforms. In contrast to other assessments, ours covers a longer time span (from 2017 to 2020) and places more emphasis on identifying the similarities and underlying principles shared by the various approaches taken to multilingual sentiment analysis. Found (i) that there has been a change in focus toward studying bilingual and code-switching strategies [1]. (ii) the seeming plateau of simpler architectures built off a backbone consisting of embedding layers, feature extractors based on single CNNs or LSTMs, and classifiers. Despite findings suggesting the more challenging tasks demand more detailed architectures, (iii) the paucity of techniques solving multilingual aspects-based sentiments analysis using deep learning, and (iv) the unexpected lack of more complicated designs such as the transformer-based [2].

Social media platforms have been primarily contributing to user-generated content since the advent of Web 2.0 and the freedom to assist the distribution of information, sharing of ideas, and expression of opinions about current global-level events, services, goods, etc. This kind of social media data consists of user attitudes linked with a wide range of topics addressed online. Detecting the themes being discussed on social media platforms and online-mannerly analyzing user feelings towards such topics is vital for keeping up with the pace of streaming data that it creates on social media platforms. This is the impetus for the paper's proposed topic-level sentiment analysis algorithm, which is based on deep learning [3]. The suggested method is unusual because it first uses topic-level attention mechanisms in a long short-term memory network to do sentiment analysis at the sentence levels using online latent semantics indexing with regularization constraints. The suggested model stands out from the others since it can perform topics-level sentiment analysis and scalable topic modeling over real-time short text input. The average recall for the SemEval-2017 Tasks 4 Subtasks B datasets, an example of in-domain topics-level sentiments analysis, is 0.879, while the average recall for the newly developed datasets collected from Twitter using the hashtags #ethereums, #bitcoins, and #facebooks are 0.846, 0.824, and 0.794, respectively [4].

Received: 15.10.2022 Revised: 25.11.2022 Accepted: 19.12.2022

Looked at the model's throughput (how many subjects it can identify per second), average responses time (in seconds), and average response time (in milliseconds) for handling sentiment analysis queries to evaluate its scalability. Large-scale topics modeling over streaming data and topics-level sentiment analysis are both possible, thanks to the experimental findings. According to the Worlds Health Organization, vaccine skepticism was one of the top ten dangers to the world's health in 2019 [5]. Information, ignorance, and deception concerning vaccinations are widely disseminated on social media platforms nowadays. It may be possible to learn what elements contribute to vaccination confidence across time and place by monitoring social media talks about vaccines. Analyzed 1,499,227 vaccine-related tweets posted on Twitter between June 1, 2011, and April 30, 2019, using a hybrid method for opinion mining. Our analysis found that 69.36% of the tweets were deemed neutral, 21.78% were deemed positive, and 8.86% were deemed negative. Over time, the share of good and negative tweets rose while the share of neutral tweets fell [6]. The polarity of reviews may be gauged with the use of sentiment analysis (SA). This study presents a hybrid strategy, GWOPS that combines hybrid algorithms [GWO and (PSO), respectively, since the optimal solution to address all optimization issues has not yet occurred. Features selection filters help narrow the search area. GWOPS is used during NN classifier training for feature selection. Our experiment relied on data gathered from Twitter on a variety of subjects. The quality and efficacy of the proposed algorithm, GWOPS, are accessed via a series of tests and comparisons with three established optimization techniques [7].

LITERATURE SURVEY

In tandem with the development of the digital platform, sentiment analysis is becoming increasingly important in fields such as cyber-vulnerability review, group discussion for cyber-danger study, malicious movements-based overviews web diary, more narrowly focused site, and casuals' association related to the study of cyber-criminals' exercise. The choice and methods used to pique interest in the present are often adapted to fit the worldviews and surveys of others in terms of emotion and sentiment. For this reason, the conventional looking-through approach is used to finalize the choice, conduct research, and evaluate the performance of others [8]. This is verifiable on many levels, from the personal to the institutional to the societal. To determine whether a piece of content contains theory information and what kind of enthusiastic data it imparts via cyber-malicious post overviews, i.e., whether the way behind this content is certain positive or negative, this work employs a broad inclinations assessment that involves the tasks of ordinary languages processing (NLP). Since revolutionary developments, cybercriminals have begun to host cyber occasions using online social networking and security platforms, which have been successfully blocked by users.

Consequently, it is of incredible assistance for the company and personal use, among others, to get an understanding of the feelings behind the content and educational material submitted by online users. The work may be arranged in accordance with various depths of content handling, such as asking for the furthest point of words, phrases, or even complete curricular sets. An AI-based technique is used here to examine a widely used system for cyber-weakness overviews [9]. Today, everyone has daily interactions with various forms of social media. Now do extensive research and statistical analysis using data collected from social media platforms. This essay uses Natural Languages Processing and Sentiments Classifications with Recurrent Neural Networks to draw conclusions and analyze the expressions (comments, hashtags, posts, tweets) of the user of the Twitter social media platforms based on the main trend (by keywords, which is primarily the 'covid' and corona virus themes in this articles). This is where we do our analysis, compilation, visualization, and summarization before moving on to the next step. Even though 'modern' tweets are notoriously vague, the trained algorithm is substantially better at identifying emotional polarity. Particularly when using RNN. Our newly developed and trained RNN model is used in these newly scraped data sets (organized by keyword's subject) to deduce what emotional manifestation happened on a particular topic at a given time [10].

Since its introduction, the digital currency has caused quite a stir in the financial markets. Artificial intelligence (AI) technologies and methods are the current focus of academic and corporate research into the question of how organizations may profit from analyzing the massive amounts of data now accessible online. Given that public sentiment drives the market and that social media offers a safe space for people to express their thoughts, businesses, and governments might utilize AI's natural language processing (NLP) technology to assess public opinion. Non-fungible tokens (NFTs) are a novel and rather unorthodox tool that are quickly becoming a promising new industry. Unlike the stock market, where prices can be determined with pinpoint accuracy, NFTs lack such quantitative benchmarks. Rather, NFT markets are pushed and pulled by things like public opinion, expectations, buyers' perceptions, and producers' goodwill. This research analyzes the public's sentiments towards

NFT as expressed on Twitter. In addition, this research uses attitude and emotional analysis to deduce why NFTs are becoming so widely used in the secondary market [11]. Using the Pearson Product-Moments Coefficients (PPMCC), categorize tweets into positive and negatives sentiment as well as eight different emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). Tweets were overwhelmingly upbeat (72%), and optimistic feelings like expectation and trust were widespread. To the best of our knowledge, this is the first attempt at a financial and emotional analysis of tweets about NFTs [12]. After Face book and Instagram, Twitter is the third most used OSN in the globe. Its data model and API for accessing that model are both quite simplistic in comparison to those of competing OSNs. Because of this, it is well suited for social network studies that aim to deconstruct and dissect vibrant networks with hundreds of millions of members to examine the patterns of online activity, the structure of social graphs, the attitude towards different entities, and the nature of harmful assaults. Indeed, Twitter has become a prominent research medium, cited in over 100,000 scholarly works in the last decade [13].

Most research that makes use of Twitter has been the subject of good review and comparison studies, but there have been few attempts to map this research landscape in its entirety. This paper describes an attempt to create a map of current research issues on Twitter, with a particular emphasis on three key areas: the structures and features of the social network; sentiments analysis; and dangers, including spam, bots, false news, and hate speeches. Also, outline sampling and data access recommended practices in addition to presenting Twitter's fundamental data model [14]. This overview also provides a foundation for the computational methods like Graph Sampling, NLP, and ML that are employed in these disciplines. In addition to discussing the important discoveries and the state of the art in these methodologies, it also examines the results of current reviews and comparative research. Overall, predicting that this survey will serve as a roadmap for the future expansion of the themes provided and aid scholars in developing a coherent conceptual model of Twitter [15].

PROPOSED SYSTEM

Contents communities (YouTube, Instagrams), social networking (Facebook, LinkedIn), blogs (Reddits, Quoras), and micro-blog (Twitter, Tumblr) are the four main forms of social information service or social media depending on their application use. Micro-blogging sites and Twitter are the most popular social media platforms for gathering data on user opinions, according to the reviewed research. Twitter is used as a source for 85% of the papers under review's sentiment analysis. One of the top ten most frequented websites; Twitter lets its users share and reply to brief messages. Scholars, businesses, and governments may all benefit from the opinions and data shared on Twitter. Twitter is a well-known micro-blogging tool and social media platform where users may share their thoughts and feelings on current events, brands, and more in 140 characters or less. The material and data that everyone can access is what makes Twitter so popular. Using an application programming interface (API), users may search for and copy information on any subject using keywords or a hashtag. The overall layout of the proposed system is seen in Figure 1.

Twitter uses its 500 million daily tweets and its open API to do real-time research and monitor public mood. Eight nations, including those in the West and the East, are searched for and collected through Twitter. Twitter's global user base means the service benefits from the wide range of perspectives, languages, and cultural backgrounds represented there. For the 2016 presidential election, Twitter was utilized to compile user tweets on each candidate, and a community development initiative used it to compile tweets about its activities. Additionally, tweets were obtained from London Heathrow Airport's official Twitter account and examined using sentiment analysis, and messages from customers to the UK energy company were collected and read on Twitter. When it comes to social media platforms, Facebook has the most subscribers worldwide. However, it is not widely used for sentiment analysis because of the issues: the data is unorganized and poorly formatted; users often resort to abbreviations and typos; and so on. This complicates the task of data analysis. Take, for instance, the practice of mining social media for sites, status changes, and comments that hint at user experiences.

The study's data collection included forums, blogs, Expedia, blog spots, traditional media, WordPress, YouTube, Twitter, aggregators, and Facebook. The analysis also reveals that Twitter is the source of 88% of the information. Other social media platforms, such as BlogSpot, YouTube, and WordPress, are not preferred because of the restricted data and perspectives they provide. Any global occurrence, whether it is an event, activity, sport, or calamity, is fair game for sentimental research. An investigation of the differences in opinion on ISIS between Westerners and Easterners is one such example. The outcomes demonstrate that people on opposite sides of the globe agree that ISIS is a terrorist organization. Data security and the risk of security breaches may be better understood via the use of sentiment analysis. It helps shape public opinion by providing a framework for how businesses should react to security incidents. The non-employment rates and the employment sentiment scores on social media were also the subjects of sentiment studies. The study's focus on healthcare provides a clear example of how and when sentiment analysis might be useful. Offer a methodology for providing sentiment analysis as a service, which makes use of the spatiotemporal features to pinpoint the origins of health crises.

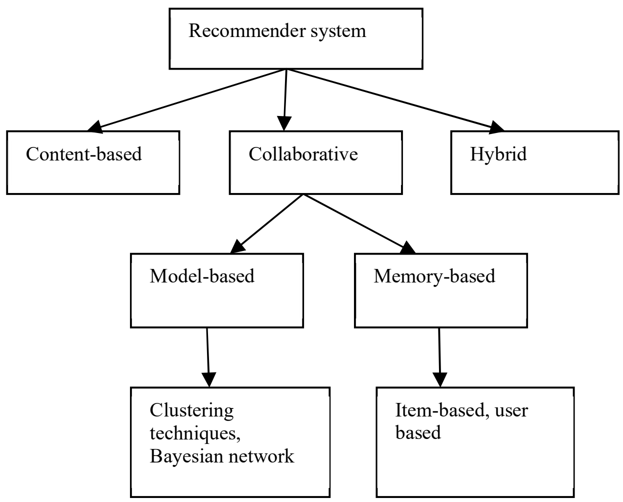

FIGURE 1. System architecture of the proposed system

In addition, it is possible to anticipate rescue efforts by using sentiment analysis to determine what people's emotional requirements are during a crisis. Furthermore, by over-serving and evaluating emotion from texts, sentiment analysis enables determining a person's depression state. The results of sentiment analysis of Twitter's data demonstrate that the platform is more dependable than previously thought; a 94% correlation was established between the two sets of data, suggesting that sentiment analysis has the potential to become as accurate as traditional polling methods. Finally, consumer feedback plays a crucial part in the implementation of sentiment analysis, as it may help businesses and organizations take the necessary steps toward enhancing their product or service and overall company strategy.

A survey of internet users' opinions and experiences with various pharmaceutical and cosmetic products demonstrates this. Using sentiment analysis, airports may pinpoint problem areas with their service and take remedial action, such as listening to customer criticism on social media. Consequently, businesses may benefit from sentiment analysis's ability to examine patterns and traits associated with consumers' eating habits. Businesses can benefit from sentiment examinations in several ways, including finding how well-liked their product and service are by consumers, measuring the efficacy of their brand communications and social media presence, and monitoring the movements of their stock prices via social media. The client's opinion helps us to see where we fall short and improve. Evidence of this may be shown in research that compared customer tweets about the Big Six and the new entrant energy consumer, which is Britain's biggest and oldest gas and electricity provider. The findings show that the Big Six are less optimistic than a newcomer to the energy market. Furthermore, research shows that a high positive sentiment is derived from tweets on community development programs activities, demonstrating the value of sentiment analysis on the success level of a program. The end outcome has the potential to boost the community's level of life. The organization can assess the impact of social media by.

RESULTS AND DISCUSSIONS

The results of the performed systematic gives detail on research into social media sentiment analysis. Three main points may be taken away from this study. Begin by demonstrating how social media sentiment analysis is conducted. Researchers have presented many other methods; however, SentiWordnet and TF-IDF remain the most popular Lexicon-based methods, while Naïve Bayes (NB) and SVMs are the most famous machine-learning methods. Which sentiment analysis technique best relies on the specifics of the available data. The accuracy of both approaches was quite high. Figure 2 illustrates how various machine learning algorithms fare in terms of accuracy.

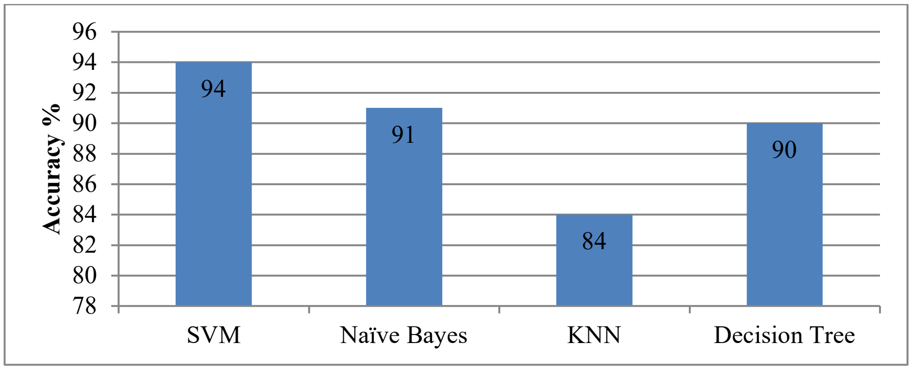

FIGURE 2. Accuracy performance of different ML algorithms

Things to think about include the textual and temporal organization of the data. It is advised to use a lexicon-based approach if the data structures are unorganized, there is a lack of data, and there is not much time to do a study. Machine learning-based approaches work well with larger datasets since they need more time and data for training. Combining a lexicon with a machine learning approach is recommended to boost both the quality and accuracy of the final product. Second, determine the most popular social media platform for gathering data for sentiment analysis. Twitter is often used as a data mining platform. Twitter is used as a social media setting in most of the examined papers. This is because of the wide distribution, ease of use, and depth of Twitter's content. Millions of tweets may be found on any given subject every day. This suggests that social media is becoming a reliable information resource. However, other social media platforms like blogs, Word Press, YouTube, and others get comparatively less coverage. It's possible that the information shown on various social media platforms varies; therefore, looking elsewhere for answers might be fruitful. The suggested system's sentiment analysis findings are shown in Figure 3.

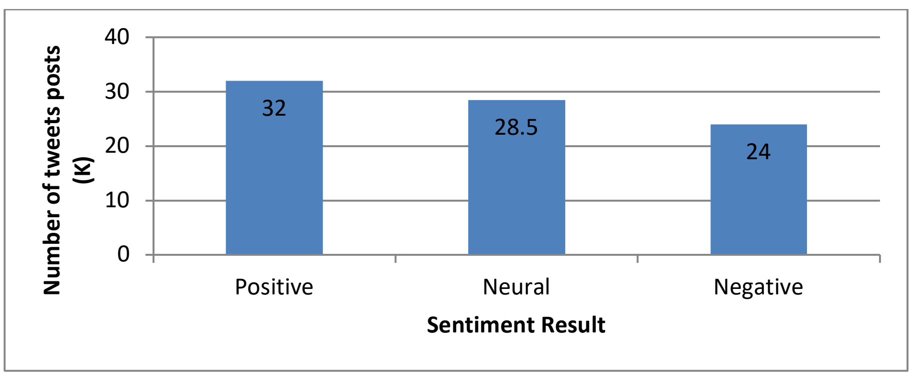

FIGURE 3. Sentimental results of the proposed system

Finally, showcase the value of sentiment analysis in social media. Sentiment analysis has many potential uses, including but not limited to business quality and strategy improvement; political election outcome prediction; disease outbreak monitoring; data security awareness raising; public opinion of a particular sport; and disaster location and response enhancement. This demonstrates the importance of sentiment analysis in gaining insight into people's perspectives and guiding their decision-making processes. Additional research is required to design a universal model of sentiment analysis that can be applied to various kinds of data, examine other possible social networking sites to get users' opinions and broaden the context of sentiment analysis applications to make future recommendations.

ISSN: 2456-8449

CONCLUSION

By analyzing the sentiment of over a million Tweets written in English, this article can compare many machine learning and deep learning techniques. Neural networks, decision trees, Naive Bayes, and recurrent neural networks were all explored, as were hybrid models. With a 90% accuracy rate, sensitivity, and specificity, the hybrid model excels in this study. The purpose of the forthcoming research is to include emotions in the text survey process. Furthermore, considering the staggering volume of tweets sent each minute, a lot of which are written, there are plans to test out the efficacy of the hybrid categorization method with Tweets soon. The supplier and the client both benefit from the widespread distribution of customer evaluations and comments on social media. The suppliers may improve their brand identity by listening to customer comments, and the new user can learn more about the product as a result. In this work, we categorize and analyze data from customer reviews on the shopping website. In this study, use the SVM, NB, KNN, and Decision Tree algorithms to the customer-provided star-rating of the product to categorize it. The accuracy of a combined SVM, NB, KNN, and Decision Tree algorithm is compared.

REFERENCES

[1]. M.M. Agüero-Torales, J.I. Salas, and A.G. López-Herrera, 2021, "Deep learning and multilingual sentiment analysis on social media data," *An overview. Applied Soft Computing.* **107**, pp. 1-12.

[2]. A.R. Pathak, M. Pandey, and S. Rautaray, 2021, "Topic-level sentiment analysis of social media data using deep learning," *Applied Soft Computing*, **108**, pp. 1-12.

[3]. H. Piedrahita-Valdés, D. Piedrahita-Castillo, J. Bermejo-Higuera, P. Guillem-Saiz, J.R. Bermejo-Higuera, J. Guillem-Saiz, JA. Sicilia-Montalvo, and F. Machío-Regidor, 2021, "Vaccine hesitancy on social media: Sentiment analysis," *Vaccines*, **9(1)**, pp. 1-12.

[4]. M.S. Alharbi, and E.S. El-Kenawy, 2021, "Optimize machine learning programming algorithms for sentiment analysis in social media," *International Journal of Computer Applications*, **174(25)**, pp. 38-43.

[5]. R. Rawat, V. Mahor, S. Chirgaiya, RN. Shaw, and A. Ghosh, 2021, "Sentiment analysis at online social network for cyber-malicious post reviews using machine learning techniques," *Computationally intelligent systems and their applications*, pp. 113-130.

[6]. L. Nemes, and A. Kiss, 2021, "Social media sentiment analysis based on COVID-19," *Journal of Information and Telecommunication*, **5(1)**, pp. 1-5.

[7]. C. Qian, N. Mathur, NH. Zakaria, R. Arora, V. Gupta, and M. Ali, 2022, "Understanding public opinions on social media for financial sentiment analysis using AI-based techniques," *Information Processing & Management*, **59(6)**, pp. 1-15.

[8]. D. Antonakaki, P. Fragopoulou, and S. Ioannidis, 2021, "A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks," *Expert Systems with Applications*, **164**, pp. 1-14.

[9]. J.P. Gujjar, and H.P. Kumar, 2021, "Sentiment analysis: Textblob for decision making," *International Journal Scientific Research and Engineering*, *Trends*, **7(2)**, pp. 1097-1099.

[10]. P. Chauhan, N. Sharma, and G. Sikka, 2021, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, **12**, pp. 2601-2627.

[11]. MA. Kausar, A. Soosaimanickam, and M. Nasar, 2021, "Public sentiment analysis on Twitter data during COVID-19 outbreak," *International Journal of Advanced Computer Science and Applications*, **12(2)**, pp. 1-6.

[12]. P. Mehta, S. Pandya, and K. Kotecha, 2021, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," *PeerJ Computer Science*, **7**, pp. 1-15.

[13]. S. Dutta, D. Sarkar, S. Roy, D.K. Kole, and P. Jana, 2021, "A study on herd behavior using sentiment analysis in online social network," *International Conference on Communication, Control and Information Sciences*, **1**, pp. 1-6.

[14]. G. Alorini, D.B. Rawat, and D. Alorini, 2021, "LSTM-RNN based sentiment analysis to monitor COVID-19 opinions using social media data," *IEEE International Conference on Communications*, pp. 1-6.

[15]. P. Nandwani, and R. Verma, 2021, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*. **11(1)**, pp. 1-19.