Anomaly Detection in Social Media Using Text-Mining and Emotion Classification with Emotion Detection

V. S. Bakkialakshmi(✉) and T. Sudalaimuthu

Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India
bakkyam30@gmail.com

Abstract. Anomaly detection in online social networks identifies abnormal activity the most as illegal behavior. Anomalous behavior identifies malicious activities, including spammers, and online fraudsters. Social networks like Twitter, Facebook users share opinions to the world primarily in text form, such as Microblogs. Anomaly detection in social media depends on text mining analytics. Text mining analytics derives quality information from the text corpus. Text mining is a prominent mining technique that is still under research for further development. Many researchers have proposed anomaly detection techniques using various text mining processes. Each comment or tweetis updated in informal human writings, using NLPT Natural Language Processing techniques unstructured texts are normalized into a standard format to apply ML algorithms. Opinions shared in social media are classified by emotions, emotion classification classified with the different emotions like happy, sad, fear, disgust, anger, surprise, trust. This paper aimsto analyze anomaly detection in social media with micro-blogs. The study deepens with text mining and emotion classification techniques from different authors.

Keywords: Anomaly detection · Text mining · Emotion classification · Social media networks · Emotion detection

1 Introduction

As information technology progresses at such a quick pace, social media has emerged as a new phenomenon in Online Social networks (OSN). Every day, people use social media to share their thoughts on a variety of topics, products, and services with their friends and followers, making it a valuable resource for text mining and sentiment analysis. Facebook, Twitter, and a slew of other sites use social media to communicate. Twitter is a popular social networking service with a large user base [1]. Figure 1 depicts a global model of the number of tweets sent each second. There is not a standardized way to mine and analyze social media business data in the literature. Text mining and sentiment analysis using a collection of R packages [2–7] for mining Twitter data and sentiment analysis are described here as open-source approaches that can be used on other social media sites. To demonstrate the value of studying user-generated online views via

D. S. Guru et al. (Eds.): ICCR 2021, CCIS 1697, pp. 67–78, 2022.
https://doi.org/10.1007/978-3-031-22405-8_5

Microblogs, a case study of two UK retailers is presented. By doing this, businesses may analyze their performance from the perspective of their customers without having to conduct costly and time-consuming client surveys.

It is the automated process of identifying and revealing previously unknown information, as well as linkages and patterns in large collections of unstructured textual data. In massive quantities of text, text mining seeks out previously unknown information. Information retrieval systems return documents that are connected to the query entered into the system, rather than random results [8]. Data mining methods, such as classification, clustering, association rules, and many more, are used in this field of study to sift through textual sources in search of new information and connections. [44] Information retrieval, data mining, machine learning, statistics, and computational linguistics are all used in the process of text mining [9]. To begin, a collection of unstructured text documents is amassed for analysis. The documents are then pre-processed to eliminate common terms, stop words, and stemming. The pre-processing creates a Term document matrix, a structured representation of the documents, in which each column represents a document and each row represents an occurrence of a term in the document as a whole. It is the last phase that uses advanced data mining techniques such as term clouds and tag clouds to find patterns and relationships in the text. Then, it visualizes these patterns with tools like these [10].

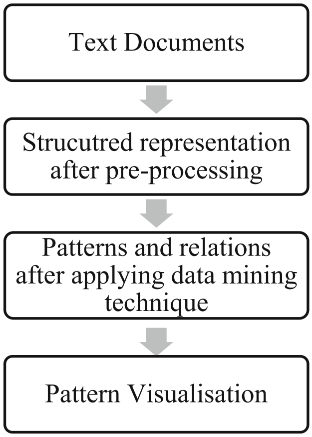

Fig. 1. Process of text mining for emotional classification

To determine the attitude or polarity of human opinions or reviews published to rate items or services, it is also known as opinion mining or subjectivity analysis [11]. Blogs, reviews, and Microblogs can all be subjected to sentiment analysis because of the textual nature of those forms of opinion expression. Microblogs are short text communications like tweets with limited characters and the sentiment analysis can be performed more easily on these microblogs than on other types of thought [12]. Document sentiment analysis or sentence sentiment analysis are two methods of conducting sentiment analysis. In the first scenario, the entire document is analyzed to establish the strength of an opinion, with the elements defining the product or service being extracted first. The second text, on the other hand, is broken down into sentences, and each one is assessed individually to establish the degree of agreement or disagreement [13].

In other words, to obtain social network data and clean and filter it fairly while also extracting its features, as well as properly storing and managing it [14]. Semantic social network learning at a deep level [15]. Specifically, to achieve heterogeneous data matching by realizing text feature and visual feature association learning, we need a technique for indexing and sorting social network data based on different search requests to implement social network search in deep learning [45]. It is possible to develop multiple ranking techniques by taking into account the current circumstances and assigning varying weights to various variables, such as text and visual elements as well as social elements and temporal and spatial ones [16–20].

Figure 2, OSNs term used very popularly by many authors. Online Social Network (OSN) such as Twitter, Facebook, LinkedIn attracts people with common interest and activities [1]. Social media platform becomes a communication window for personal as well as business promotions. Most of the companies promote their brand on social media platforms and increase their sales. Meanwhile, the user metadata are gathered by those social media and used for promoting ads for the same individual user [2]. From the user's side view, a home-based individual can also start a business and make money in enhanced Online social media networks according to Ravneet and Sarbjeet [1].

Anomaly Detection is also referred to as Outlier- Detection. Anomaly detection plays a lead role in text mining concepts and techniques, fetching anomalies are of critical importance in social media to prevent medical scan, bank transactions, image processing, Anomaly detection is defined as anomalous behavior or abnormal actions detected in regular patterns, from the given data set. Labeled or unlabeled data can be supervised or unsupervised for anomaly detection.Anomalies in social media networks arise by malicious individuals or Online fraudsters by changing their interaction patterns. Some of the recent updated malicious activities are spamming, cloning profiles, jeopardizing the identity. Users usually make interactions in social media in text form. Text-based sharing of opinions is an easy and fastest way of communication. Human nature always expresses any comment with some emotions, expressed emotions say about the state of mind and situation of that particular user.

Social media is commonly used to register every individual's perspective of views. Each comment that is shared by people are assigned has raw data. For every social event happening, public opinions that are raised make the event viral or non-viral. The collected corpus data is always said to be linguistics data. Data sets collected contains informal, non-grammatical, colloquial kind of sentences or texts. The initial part of the data needs to be preprocessed before implementing some training models. The irregular form of human writing knows to be a natural language, making the natural language understanding to produce desired output is a challenging job [21–24].

Multiple interpretations in the text content of the same sentence make the computer understanding harder. Text preprocessing is a critical part of work, mostly done with Natural language processing (NLTK Tools). Raw data might consist of linguistic diversity, any insertions, links. Data pre-processing steps follow with removing stop words, removing links & URLs, changing uppercase and special characters to lower case, removing

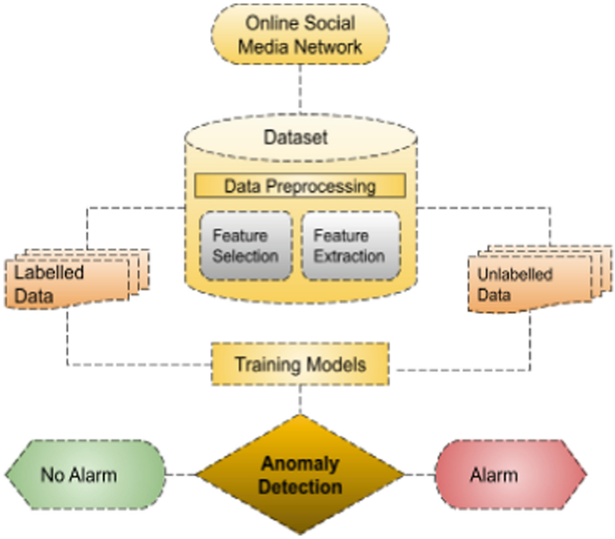

Fig. 2. OSNs anomaly detection.

punctuations after stemming process deriving the final machine-understandable format. Short messages converted to feature space [25–27].The feature dimension process is divided into feature selection& feature extraction. Feature selection selects the required subset from the existing original data. Feature selection simplifies the Data set to make them undemanding in shorter training times. Enhanced rationalization by reducing the overfitting of data. The subset selection starts with an empty data set. Each feature is compared with the remaining data. The next process is to estimate classification/regression error for adding a new feature. Finally, selecting features that give maximum improvement. Same time, the feature selection process stops when there is no significant improvement found [28–34].

The feature Extraction process reduces the initial set of raw data and makes them into a manageable groupof features. Feature extraction aims to reduce the features by creating brand new features from the existing features and discarding the original one. Feature Extraction (Feature Projection) usually transforms data from high dimensional space to lower-dimensional space. Same way feature extraction reduces the amount of redundant data for any data analysis, that transforms words and sentences into feature vectors.Supervised anomaly detection represents a smooth method for training models with labeled data and predicts the unseen data. Algorithms like KNN, SVM are statistically built around supervised classification. Unsupervised anomaly detection that provided with unlabeled data fails to meet the required task. Predicting anomalous activities becomes critical with unseen actions. The unlabeled dataset has only one knowledge saying that less than 1% of a dataset has anomalies.

2 Anomaly Detection in Social Media

The literature proposes numerous graph-based intrusion detection methods. A recent focus is on anomaly identification in social networks, with particular attention paid to graphical aspects. Based on structural metrics, graph-based networks may better handle the topological characteristics of the network by using multiple graph metrics. The method in [6] examined a variety of node- and agent-based properties, then combined and calculated their combined value recursively. Many academics have looked at structural qualities that represent topological elements of networks to uncover patterns in the characteristics of the objects. Finding anomalous links between nodes can help uncover unusual behavior, for instance [7], by identifying links between the nodes that aren't normal. Prior scholars have utilized the concept of graph theory to depict networks with ease and precision utilizing graph metrics patterns created by the discipline.

The authors in [8] proposed Random walk, an outrank algorithm that assumes anomalous objects have a low connectivity score because there are fewer similarities between them and the other objects in the group. The "Oddball algorithm" described in [9] employs geonet density power laws to discover anomalies, such as stars or cliques that are close together. New features such as Average Betweenness Centrality (ABC) and Community Cohesiveness are employed by Reza et al. to distinguish between nearby stars and nearby clique abnormalities. Using the same power-law curve fitting approach in [11] found near the star and near clique anomalies in Twitter's dataset.

Many circumstances necessitate the assignment of scores according to the degree of abnormality in the nodes even though anomalies are considered a binary choice problem. Using the local outlier factor in. [12] assigned each object a similarity score based on how similar it is to an outlier. Most studies have employed scores to assign the number that represents the degree of anomalousness to an object based on how closely it resembles its surroundings. It highlights the benefits and drawbacks of various structural and behavioral anomaly detection methods and strategies, including After weighing the pros and cons of several approaches, we concluded that behavioral strategies fell short due to privacy concerns. There is no way to make behavioral data available to the general public.

3 Text Mining

Researchers of [10, 11] explored that text mining majorly reads unorganized forms of data that saymeaningful information patterns in a short period. In all social media sites, the people communicate of text is shared in some words or sentences either short or long, those sentences are not written with proper grammar and spellings. Even thoughpeopleallaroundtheglobesharetheirknowledge, ideas, and interests in social media, they are not interested to follow structured sentences, right spelling, error-less grammar. For acquiring the structured sentences, various vocabulary-based-text mining approach is used. Text mining focuson correct datathatcanbestructuredor unstructured.Some of the text mining approaches are given in Table 1.

Table 1. Text mining in social media networks

S. No	Author	Text mining methodology & approach
1	Eman M.G. Younis	Lexicon-Based Sentiment Analysis Approach
2	Zielinski et al.	Forest of classifiers approach
3	Shilpy et al.	Bayesian Supply Regression Classification Methodology
4	Myneni et al..	KNN classification Methodology
5	Aditya Akundi et al	Refined Hashtags Selection Approach
6	Namugera et al.	Lexicon Based Approach with LDA
7	Kia Jahanbin et al.	Fuzzy Rule-Based Evolutionary Algorithm
8	Said A.Salloum et al.	Automatic Classification Approach

In General, data collected from social media are not obtained for a research purpose, however, the data should be changed from unstructured to structured data. Overall available data from social media are 80% unstructured data and 20% structured data. Text mining techniques need to find the words according to the needs of NLP in an automated way. Text mining is completely involved with Natural languageprocessing.

Argument Mining: According to the study by [12], to identify Arguments on Twitter supervised classification with two tasks say facts recognition and source identification. Argument mining from any variety of textual corpora aims directly at natural language arguments. With several approaches done earlier, two main tasks are identified, they are arguments extraction and relations prediction. To address the tasks of classifying the instances from the given datasets, the study uses supervised classification to separate augmented tweets from non-augmented tweets. By applying a supervised classifier, the factual tweets are identified and resultsproduced.

Sentiment Mining: [9] Corpus data in Sentiment mining deals with Etymological Investigation. The etymological investigation is finding the root of the word's origin, like finding the language & country origin.

3.1 NLP Techniques

Tweets majorly involved short and long microblogs, understanding each person'sthoughts of opinions based on sentiments used in that microblog. For Sentimental Analysis N-Gram's method is used. N-Gram's approach cross three sentiments say" Pos"," neg"," neutral". DependingonthesizeoftextN-Gramsisclassified:

- Size 1 -Uni-gram
- Size 2 -Bi-gram
- Size 3 -Tri-gram

N-Gram means continuous sequence of words or letters of N count from the text corpus or speech corpus.

3.1.1 Micro-blogs for Emotion Recognition

Twitter is one of the well-utilized social media platforms. Twitter tweets are considered data-sets and are mostly used by researchers. Tweets are also called micro Blogs that allow users to share opinions, whereasTwittermicro Blogs havea maximum of 280 characters. Retweets allowthe users to communicate with each other in their style. Microblogging in Twitter spreads opinions that may not be valuable for microblogging space fundamental views, considering fundamentally or dimensional basis with clustering [39–43].

4 Emotion Classification

Twitter found in 2006 with 300 million active users monthly and projected 340 million users in 2024. Reason twitter is popular is easily accessible and a good media for data communication and classification as in Table 2. Reason with numerous counts of features the automatic text classification is hard. Twitter recently extended to 280 characters, with the limited word count the classification becomes even harder. With multiple use cases emotions captured, user happy emotion reflects in his positive comments. Comments can be detected and sent to the developer team; the team gets satisfied automatically that the user enjoys the app [35–38].

The authors of [6] illustrated the emotion classification techniques, methods, and challenges. Emotion recognition can be either text, speech, videos, bio-Sensor. Emotional recognition in daily life needs different data types available for everyone. The action should be fast to uncover and focus lies on Real-Time Emotion Classification. Twitter data obtained from the real-time datasets are processed via hashtag as one of the features i.e. when a person tweets, the emotion of the tweet can be detected. The hashtag Corresponds with his feeling at the same moment. Author [8] used the wang et al. [13] dataset and tried to improve the accuracy in emotion classification. Author [8] used three filters' tweets with a hashtag at the end alone, discarding the contents of the tweet below than 5 words, tweets with URLs or quotations were removed.

The preprocessing technique is done with stemming, stemming reduces the feature space. Feature extraction transforms words and sentences into a numerical representation say feature vectors. The authors used classification algorithms and a combination of N-Gram & TF-IDF. To fetch the accuracy following classification metrics were used, precision, recall & F1-Score. The accuracy of N-Gram & TF-IDF is improved to 5.01% after the estimations of TP, TN, FP, and FN. Authors [8] reduced real-time emotion classification processing time as small as possible. Emotion classification can be distinguished or contrast emotions from one another. Various researchers developed several architectures based on service-oriented architecture. REST says Restful Web Service is a lightweight, maintainable, and scalable service. With REST service no processing is done from the client-side. Emotion classification is done on the server-side; the REST services help the Real-Time aspects smooth even in mobile devices.

4.1 EMOTION: Detection vs Classification

TF-IDF: Features include the representation of a bag of words and the occurrence of words. Classification will be affected if Emotions are out of balance. Normalization is the process of dividing the word occurrences by the total words in a document. Term Frequencies are the new features (TF). The combination of term frequencies and down-scaling weights of dominant words is called Term Frequency Times Inverse Document Frequency (TF-IDF). The approach used - Lexicon Based Approach 9 Emotion Classification Algorithm discussed. Algo's compared with a focus on precision and Timing. Accuracy Enhanced to 5.83%.

Detect electrical potential by muscle cells, Skin Conductance. This is measured with the moisture level & body Posture Measurements System. Expensive Should only be operated by a trained person Lab Facility needed with controlled environment. Lexicon-based technique: Use dictionaries of words with emotional value. Advantages: Data with ground truth not required. no features can be extracted from the relevant dataset.

Table 2. Recent detection and classification models

S. No	Type	Method
1	Detection	Detection Method with SVM KNN Random Forest classifier
2	Detection	ROI based Expression Detection with ANN classifier
3	Detection	ROI based Expression Detection with Optical Flow-Based Analysis
4	Detection	Expression Detection with 3-NN and MLP
5	Classification	CNN Based Approach with one input and output layer, and three hidden layers
6	Detection	Active Appearance Model
7	Detection	Sobel edge detection method with neural network classifier
8	Detection	Detect contours using MLP and RBF with ANN classifier
9	Detection	Active Template Method detects the keypoint locations
10	Detection	Sobel edge detection method with PCA
11	Classification	Bilinear Pooling with CNN
12	A unified model for emotion Detection using CNN	Hashtags are used to create labeled items with emotions
13	Six classifiers are used utilizing SVM for basic emotion classification	Emotion-word hashtags are used for tweet labeling

(*continued*)

Table 2. (*continued*)

S. No	Type	Method
14	Logistic regression detection	Uses personality predictive features with Myers-Briggs personality type
15	multi-task DNN Classifier	This model can map the arbitrary text into semantic vectors
16	RNN detection	sentence-level extraction of opinion expression
17	SVM classification	This model analyses electoral tweets for emotion, sentiment, intent, or purpose in a tweet
18	Random forest detection	Generalized model to detect bots
19	Random forest classification	Ensemble of classifiers to identify sentiment
20	Rule-based learning	The study identifies informative keywords
21	CNN	Word embedding layer with the emotion classification for detection of labels

5 Challenges

Exploring suspicious behavior is an important undertaking since it can reveal malicious activity and necessitate a thorough understanding of what is considered to be normal and abnormal user behavior. Graph theory is a powerful tool for simulating social networks.

- By extracting useful information about users' behavior and separating anomalies from typical users, a user can obtain structural graph metrics.
- To detect anomalies, existing structural indicators, such as degree, brokerage, edge count, and centrality of betweenness, are inaccurate and generate false positives.
- As a result, the goal of this study is to identify anomalies in social networks and limit the number of false positives and negatives in the detection process.
- Because of its powerful representation of interdependent items, structural features of networks are significant for characterizing nodes.
- Few people in most social networks break from the majority's patterns, which can be identified by the network's structural qualities. Research question: "Detection of anomalies in social networks by use of structural graph metrics and comparisons with other graph metrics".

6 Conclusion

In this paper, we surveyed various detection models that enable the system to research social media technologies. Social media anomalies alert people like natural disasters and disease outbreaks early on, allowing them to get prepared and informed. The utilization of various detection models for text mining in social media may be used for identifying bullying, terrorist attacks, and disseminating false or misleading information.

To avoid catastrophes and attacks, the detection models must detect malicious activity early and precisely. It is getting difficult to find patterns in the social media data as it gets more widely available, which is helping with the detection process. Traditional anomaly detection scenarios lack the social data that social media platforms like Facebook and Twitter. Anomalies can be categorized into two categories: point abnormalities and widespread anomalies. Graphs, unstructured texts, and sequential data are the most common input types for the anomaly.

References

1. Kaur, R., Singh, S.: A survey of data mining and social network analysis-based anomaly detection techniques. Egypt. Inf. J. **17**(2), 199–216 (2016)
2. Savage, D., et al.: Anomaly detection in online social networks. Soc. Netw. **39**, 62–70 (2014)
3. Bindu, P.V., Santhi Thilagam, P.: Mining social networks for anomalies: Methods and challenges. J. Netw. Comput. Appl. **68**, 213–229 (2016)
4. Jahanbin, K., Rahmanian, V.: Using Twitter and web news mining to predict COVID-19 outbreak. Asian Pac. J. Trop. Med. **13**(8), 378 (2020)
5. Li, X., et al.: Twitter data mining for the social awareness of emerging technologies. In: 2017 Portland İnternational Conference on Management of Engineering and Technology (PICMET), IEEE (2017)
6. Sharma, P., et al.: Challenges and techniques in preprocessing for Twitter data. Int. J. Eng. Sci. Comput. **7**(4), 6611–6613 (2017)
7. Mishael, Q., Ayesh, A.: Investigating classification techniques with feature selection for intention mining from the Twitter feed. (2020)
8. Janssens, O., Van de Walle, R., Van Hoecke, S.: A learning-based approach for real-time emotion classification of tweets. Appl. Soc. Media Soc. Netw. Anal. 125–142 (2015)
9. Mishra, G., Varshney, S.: Location-Based Opinion Mining of Real-Time Twitter Data. New Delhi, India (2016)
10. Salloum, Said A., et al.: A survey of text mining in social media: Facebook and Twitter perspectives. Adv. Sci. Technol. Eng. Syst. J. **2**(1), 127–133 (2017)
11. Bala, M.M., Navya, K., Shruthilaya, P.: Text mining on real-time Twitter data for disaster response. Int. J. Civ. Eng. Technol **8**(8), 20–29 (2017)
12. Dusmanu, M, Cabrio, E., Villata, S.: Argument mining on Twitter: Arguments, facts, and sources. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (2017)
13. Wang, W., et al.: Harnessing Twitter big data for automatic emotion identification. In: 2012 International Conference on Privacy, Security, Risk, and Trust and 2012 International Conference on Social Computing, IEEE (2012)
14. Hassani, H., Beneki, C., Unger, S., Mazinani, M.T., Yeganegi, M.R.: Text mining in big data analytics. Big Data Cogn. Comput. **4**(1), 1 (2020)
15. Pathak, A. R., Pandey, M., Rautaray, S.: An adaptive model for sentiment analysis of social media data using deep learning. In: International Conference on Intelligent Computing and Communication Technologies, pp. 416–423. Springer, Singapore (2019)
16. Tiwari, D., Kumar, M.: Social media data mining techniques: A survey. In: Information and Communication Technology for Sustainable Development, pp. 183–194. Springer, Singapore (2020)

17. Soong, H.C., Jalil, N.B.A., Ayyasamy, R.K., Akbar, R.: The essential of sentiment analysis and opinion mining in social media: Introduction and survey of the recent approaches and techniques. In: 2019 IEEE 9th Symposium on Computer Applications and İndustrial Electronics (ISCAIE), pp. 272–277. IEEE (2019)
18. Subramani, S., Michalska, S., Wang, H., Whittaker, F., Heyward, B.: Text mining and real-time analytics of Twitter data: A case study of Australian hay fever prediction. In: International Conference on Health Information Science, pp. 134–145. Springer, Cham (2018)
19. Yang, T., et al.: Traffic ımpact area detection and spatiotemporal ınfluence assessment for disaster reduction based on social media: A case study of the 2018 Beijing rainstorm. ISPRS Int. J. Geo-Inf. **9**(2), 136 (2020)
20. Zad, S., Heidari, M., Jones, J.H., Uzuner, O.: A survey on concept-level sentiment analysis techniques of textual data. In: 2021 IEEE World AI IoT Congress (AIIoT), pp. 285–291. IEEE (2021)
21. Lwowski, B., Rad, P., Choo, K.K.R.: Geospatial event detection by grouping emotion contagion in social media. IEEE Trans. Big Data **6**(1), 159–170 (2018)
22. Elagamy, M. N., Stanier, C., Sharp, B.: Stock market random forest-text mining system mining critical indicators of stock market movements. In: 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP), pp. 1–8. IEEE (2018)
23. Tyagi, N., Ahmad, S., Khan, A., Afzal, M. M.: Sentiment analysis evaluating the brand popularity of mobile phone by using revised data dictionary. **7**, 53–61
24. Jia, K.: Chinese sentiment classification based on Word2vec and vector arithmetic in human-robot conversation. Comput. Electr. Eng. **95**, 107423 (2021)
25. Rani, M.S., Sumathy, S.: Online social networking services and spam detection approaches in opinion mining: A review. Int. J. Web-Based Commun. **14**(4), 353–378 (2018)
26. Punel, A., Ermagun, A.: Using the Twitter network to detect market segments in the airline industry. J. Air Transp. Manag. **73**, 67–76 (2018)
27. Nguyen, T., Venkatesh, S., Phung, D.: Academia versus social media: A psycho-linguistic analysis. J. Comput. Sci. **25**, 228–237 (2018)
28. Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., Mora, H.: A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. Indus. Mark. Manag. **90**, 523–537 (2020)
29. Pimpalkar, A., Raj, R. J. R.: Social network opinion mining and sentiment analysis: classification approaches, trends, applications, and ıssues. In: Congress on Intelligent Systems, pp. 755–773. Springer, Singapore (2020)
30. Balakrishna, K., Rao, M.: Tomato plant leaves disease classification using KNN and PNN. Int. J. Comput. Vis. Image Process. **9**(1), 51–63 (2019). https://doi.org/10.4018/IJCVIP.2019010104
31. Iyyanar, P., Kaviya, G.: Employee reputation information system using sentiment analysis and opinion mining. Turk. J. Physiother. Rehab. **32**, 3
32. Hasan, R.A., Alhayali, R.A.I., Zaki, N.D., Ali, A.H.: An adaptive clustering and classification algorithm for Twitter data streaming in Apache Spark. Telkomnika **17**(6), 3086–3099 (2019)
33. Swamiraj, S., Kannan, R.: Stock recommendations using bio-inspired computations on social media. Adv. Nat. Appl. Sci. **11**(9), 306–314 (2017)
34. Zhang, N., Guo, X., Zhang, L., He, L.: How to repair public trust effectively: Research on enterprise online public opinion crisis response. Electron. Comm. Res. Appl. **49**, 101077 (2021)
35. Edo-Osagie, O., De La Iglesia, B., Lake, I., Edeghere, O.: A scoping review of the use of Twitter for public health research. Comput. Biol. Med. **122**, 103770 (2020)
36. Kaliyar, R.K., Goswami, A., Narang, P.: FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications **80**(8), 11765–11788 (2021). https://doi.org/10.1007/s11042-020-10183-2

37. Joshi, A., Karimi, S., Sparks, R., Paris, C., MacIntyre, C.R.: Survey of text-based epidemic intelligence: A computational linguistics perspective. ACM Comput. Surv. **52**(6), 1–19 (2019)
38. Gamal, D., Alfonse, M., El-Horbaty, M., Salem, A.B.: Analysis of machine learning algorithms for opinion mining in different domains. Mach. Learn. Knowl. Extract. **1**(1), 224–234 (2019)
39. Liu, S., et al.: Bridging text visualization and mining: A task-driven survey. IEEE Trans. Visual. Comput. Graph. **25**(7), 2482–2504 (2018)
40. Sunitha, P.B., Joseph, S., Akhil, P.V.: A study on the performance of supervised algorithms for classification in sentiment analysis. In: TENCON 2019–2019 IEEE Region 10 Conference (TENCON), pp. 1351–1356. IEEE (2019)
41. Khan, H.U., Peacock, D.: Possible effects of emoticon and emoji on sentiment analysis web services of work organizations. Int. J. Work Org. Emot. **10**(2), 130–161 (2019)
42. Sathiyanarayanan, M., Junejo, A. K., Fadahunsi, O.: Visual Auxiliary Solutions to Analyse Social Media Data for Improving Marketing & Business. In: 2019 International Conference on Contemporary Computing and Informatics (IC3I), pp. 169–174. IEEE (2019)
43. Li, Q., Chen, Y., Wang, J., Chen, Y., Chen, H.: Web media and stock markets: A survey and future directions from a big data perspective. IEEE Trans. Knowl. Data Eng. **30**(2), 381–399 (2017)
44. Visalaxi, S., Punnoose, D., Sudalai Muthu, T.: An analogy of endometriosis recognition using machine learning techniques. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE (2021)
45. Visalaxi, S., Sudalai Muthu, T.: Automated prediction of endometriosis using deep learning. Int. J. Nonlin. Anal. Appl. **12**(2), 2403–2416 (2021)