Multi-Trend Twitter Sentiment Analysis: Collaborative Approach for Improved Results

[1] Keerthika J,[2] Hemapriya N,[3] Abinaya R,[4] Karpagathareni S and [5] Bhavadharani S
[1,2,3,4,5]Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore, India
[1] keerthika.jcse@sece.ac.in,[2] Hemapriya.s2019cse@sece.ac.in, [3] abinaya.r2019cse@sece.ac.in,
[4] karpagathareni.s2019cse@sece.ac.in,[5] bhavadharani.s2019cse@sece.ac.in

Article Info
A. Haldorai et al. (eds.), *2nd International Conference on Materials Science and Sustainable Manufacturing Technology,* Advances in Computational Intelligence in Materials Science.
Doi: https://doi.org/10.53759/acims/978-9914-9946-9-8_12

Abstract— Twitter has a significant number of daily users through which tweets are utilized to communicate their thoughts in this era of growing social media users. This paper offers a way to haul out sentiments from tweets as well as a method for sorting out various tweets as optimistic, adverse, or unbiased. It refers to identifying and classifying the sentiments expressed in the text source. The existing Twitter APIs for data extraction are used to mine public Twitter data. Tweets would be chosen based on a few carefully chosen keywords related to the domain of our concern. In our proposed method, we collected various sentiment data from a variety of tweets to train and produce more precise and reliable sentiment classifiers for each trend. This method automatically extracts the key elements of subjects from online user evaluations. Since tweets are generally unstructured in format, they must first be converted into structured format. And after that, the data is fed into several models for training and used to rank the best sentiment classifier. The intention of this design is to arrive at a model that can classify sentiments of real-world data using Twitter.

Keywords— Recommendation System, Sentiment Analysis, Literature Review, Content-Based Filtering.

I. INTRODUCTION

User-based content, such as product evaluations, blogs, microblogs has been expanding dramatically and dynamically with the growth of Web 2.0 websites. Classification of sentiment information in vast quantities of user-based info may aid in acknowledging the wider public feelings about a handful of subjects, such as themes, labels, calamities, proceedings, personalities, and many more. Experts, for instance, have proposed that examining the emotions expressed in tweets has a significant possibility for forecasting the variance of presidential election results and stock market values. Classifying the feelings of numerous microblog entries is also beneficial as a time- and money-saving alternative to traditional polling.

Sentimental Analysis

Sentimentality examination is a NLP method that determines a document's emotional undertone. Organizations frequently use this kind of analysis to forecast and group customer attitudes on a given good, service, or concept. It is the process of identifying optimistic and adverse attitudes in text. Commerce often uses it to analyze common media facts for sentiment, measure brand standing, and improve relationships with their clienteles. This is a difficult challenge to address due to the high degree of un-structuredness in natural language.

Importance Of Sentimental Analysis

As individuals bring out their opinions and state of mind much more easily than was possible, sentiment analysis is becoming an essential technique for tracking and comprehending sentiment in all forms of information. Many brands are able to infer what makes their clienteles pleased or irritated by routinely examining customer feedback which is in the form of comments in review responses and social platform chats, allowing them to create goods and facilities that meet their needs.

II. RELATED WORKS

It is becoming clear that recommender systems are useful tools for both business and academia. Sentiment analysis seeks to detect and extract beliefs, attitudes, and behaviors from people and groups.

Numerous issues of recent relevance in the areas of statistics and machine learning are put forth[1]. They can be framed within the context of curved made easy methods. Due to the increasing size and complexity of contemporary data, the ability to solve problems using a large number of training instances is becoming increasingly important. As a result, decentralized collection, or storage of these datasets, as well as associated distributed solution techniques, are critical. The alternating direction multiplier approach is especially well suited to large-scale issues emerging in numbers, ML, and linked turfs.

Automatic polarity prediction of users uploading sentiment data (positive or negative) (e.g., reviews, blogs),Although manually labeled text data can be used to train traditional classification algorithms, the labeling process is time-consuming and costly[2].In different tweets, users frequently use different terms to express their emotions. The performance of a classifier trained in one trend on other tweets will be minimal due to the variances between these tweets. We develop a universal solution to sentiment classification in this work when we have some labeled data in various trends that are treated as source Trends but no labels in the target trends.

In [3] there is an argument that there is an enormous amount of web-based opinion data that requires automatic technologies to assess and comprehend how people feel about certain issues. The sentiment lexicon is quite important in most applications that use sentiment analysis. Since word polarity is delicate, it is well recognized that there is no sentiment lexicon that is universally best. The same word might signify distinct polarity regarding various features, even in some worse circumstances and in the same trends. For instance, the word "big" while writing a laptop review is not good for battery life but good for screen life. The author of this research concentrated on the challenge of learning a sentiment vocabulary that is not just trend-specific but also context-dependent.

Learning how people think and how they classify good and bad is an essential and crucial phase in the information gathering process. As people's views and thinking can be obtained from online evaluation pages and personal content pages become more widely available and popular, new chances and trials come out, as we can now dynamically make use of IT facilities to pursue and comprehend the ideas of people. Thus, at least in part, the unexpected urge in curiosity in the upcoming systems has resulted in a sudden drop-in engagement in the sector of opinion extracting and text analytics,which is associated with the computer intervention of observing sentiment, and prejudice in messages. This study looks at approaches and plans that claim to directly support systems for discovery view-focused info. Furthermore, it focuses on methods intended to deal with some new tasks brought on by sentiment-aware applications, which are already current in more out-of-date fact-based analysis [4].

In [5], there is a survey that in the wake of political involvement among young people and the general population in general, everyone is not only eager to communicate their political views but is also interested in hearing the opinions of the majority. As is well known, social media platforms provide the ideal entry point for this newfound urge for political activism since they make it simple to gather data on the many facets of popular opinion. These websites are starting to have a significant influence on how individuals think and behave.

The opinions on social media sites that enable global access to what people believe about issues and themes related to daily life. Therefore, using such a source of data to understand public opinion can be highly beneficial in a variety of situations. The mission of the research field known as information extraction is to construct precise automated methods for analyzing opinion data. Although there has been a lot of prior research on opinion mining, most early studies focused on text documents like movie and product reviews.[6]

Sentiment examination looks for views (or polarity) conveyed in text about a particular subject or issue. Individuals are posting their views and opinions on a regular basis as a result of the explosive growth of websites for social networking including blogs services, social media platforms and text message services are all examples of these. Because insights or emotional states gleaned from text could be relevant for subsequent applications such as recommended, analytical research, status monitoring, or political concept prediction, sentiment analysis has gained a lot of attention. In the past, scholars concentrated on determining the valence of text using linguistic cues taken from reviews' textual content. In addition to ratings, many recommendation and review websites provide a variety of information on target entities and opinion holders (hence referred to as users). [10-13]

K-Means Clustering is used as a classification technique in the base paper, indicating that this is an unsupervised Ml process. It also produces outcomes that are similar to ours, but instead of using an unsupervised method, the supervised method is established by using Support Vector Machine technique. This produces more precise results than K-Means, K-means clustering is sensitive to the initial centroids chosen, resulting in different results each time the algorithm is run. Because K-means is intended to work with numerical data, it may underperform when dealing with text data, which necessitates additional pre-processing steps.As a result, Flume is used as a tool to collect data from people in timely manner, and save is employed for pre-process of the metrics before using algorithms to reduce time and increase speed to get the output. The major drawback on using K-means is when there are clusters with different densities and sizes, k-means has problems clustering the data. K-means is unsuccessful because the objective function that it tries to reduce rates the actual clustering solution as being inferior.

III. PROPOSED SYSTEM

In the proposed work, we have used Algorithms like Naive Bayes and Drimux SVM which suggest tweets by pairing customers who share their passion. It gathers buyers comments in metrics valuations posted by subscribers for unique tweets and presence for correlations in valuation patterns across users to identify groups of users with shared tastes. One of the key aspects of the Tweet main page ,is a set of "trending topics" that are readily visible at all times. These phrases represent the subjects that are now receiving the most attention on the site's rapidly updating flow of tweets.Twitter concentrates on chapters that are being talked considerably more than normal.In this case, a participant's profile indicates the priorities chosen by the user, whether directly or indirectly . Twitter, for instance, uses the GB method, which proposes tweets entrenched on user ratings and buying history.Each operator has a bundle of tweets that have been researched in a certain way, either directly or indirectly. As a result, a user-tweets rating matrix is created which represents individual interests of customers. Different methods are used to identify lost ratings, such as determining a user's "nearest neighbor" when recommending tweets to new users based on the ratings of those users' closest neighbors. **Fig 1** shows the Block diagram.

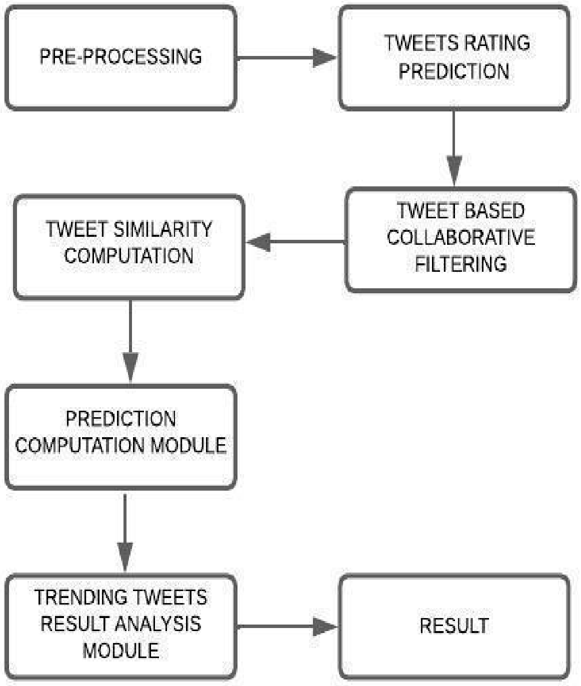


Fig 1. Block Diagram

IV. METHODOLOGY

Pre-Processing

Actual ratings are utilized in the database creation for the Twitter asynchronous system. The dataset's use defines the reliability of the results, so constructing the database is an essential step. Thanks to websites that make available datasets comprising users and tweets with wide - ranging rating histories, it is possible to gather a sufficient quantity of highly expected tweets for suggestions to all subscribers. The data is obtained using Twitter's open framework. The top 10 trending topics on Twitter are updated momentarily. It is unknown how topics are selected to be part of this list, or how often it is updated. However, for a certain hot topic, one can request up to 1500 tweets.

In order to gather this data, two procedures were running. Every 30 seconds, one process asked Twitter for a list of trending topics and kept a separate list. The other procedure used Twitter's search framework to ask for a list of related tweets whenever a new trending subject was discovered. The following three categories were manually created from the popular themes after the data had been gathered:

1.News
2.Memes
3. Trending Topics

The trending topics are annotated using three annotators. To choose an appropriate category, each of them looked at the tweets relating to the popular themes.

Tweets Rating Prediction

There are avaricious and lively obstructive algorithms in this segment. approaches for the Twitter asynchronous system Greedy-algorithm is suggested. It is a live information-based technique that recommends nearly equivalent tweets to those the worker has formerly selected. This method proposes tweets that earlier users who shared your tastes or have liked. It can combine methodologies for collaborative and content-based sifting. The Twitter asynchronous system carries out the two tasks below while making recommendations to each user. First, using a recommendation system, the ratings of unrated tweets are anticipated based on the information currently available. a fresh technique of categorizing Twitter trends that incorporates the rating of the finest hashtags for tweets on specific topics and trends selection. The process of feature selection is facilitated by the use of a diversity of feature position algorithms, including TF IDF and bag_of_words model. As a result, the feature content is shrunk, the organization process becomes more effective, and the key features are brought to the surface. We successfully classify Twitter trends using four Greedy and Dynamic Blocking manuscript classifiers, supported by these advanced feature position and feature assortment algorithms. Our study offers a regular class exactness development above the present approaches of thirty-three percent and twenty-eight percent, accordingly, using TF IDF and bag_of_words rankings.Second, the system detects pertinent tweets and endorses them to the handler depending on the outcome of expected ratings.

Tweet Based Collaborative Filtering

The following segment compares a collection of comments/text that the active user has rated to the target tweets, determining the degree of similarity, and after that it selects N maximum identical content. The corresponding commonalities between Tweets are also calculated. The prediction is calculated using the tweets that are the most comparable. The actual retrieval and selection of movies from the movie database is handled by the information filtering module. Information is filtered based on the knowledge gained from the learning module. the user's standardized ratings are entered into the rating database after passing the user knowledge exam. The steps listed below are used to recommend a movie to the user UI based on information from the rating database. M is supposed to be a variable representing all the users. N is accountable for the movies. Total amount of movies that users have not reviewed.

1) Determine the correlation between each of the other (N-1) films and each film F n that has not been rated by user ui.
2) Choosing S movies, which are typically highly connected with F, is based on the correlation coefficient values. This will combine with F to make a group of S related films.

Tweet Similarity Computation

Finding the people who have rated both of the tweets in this module's similarity computation between a and b is the first step in the process. There are numerous methods for calculating similarity. The adjusted cosine similarity method used by the suggested system is more advantageous since it subtracts the relevant couple of co-rated items and seems to have a user average.There is a contrast between tweets a and b.

Prediction Computation Module

In this component, the weight vector technique is employed to derive the predictions. By calculating the sum of the user's ratings on tweets that are similar to the target tweets, weighted sum determines the forecast of the target tweets for a user u. A prediction is made regarding a tweet sent by user u. Technique based on content.The versatility of posts on twitter for a given customer is measured using utilities that user has assigned to gather twitter posts that are similar to tweets.Only tweets with a high degree of user preference similarity are recommended.

Trending Tweets Result Analysis Module

Information about users, movies, and ratings is kept in several tables in the movie database development module. As a result, the system can accurately collect data from the database and obtain user-explicit movie ratings. Tweets similarity computation and prediction computation modules have been applied in tweets-based collaborative filtering techniques. On the basis of a user's non-purchased movies, recommended lists are developed. Therefore, we computed system projected ratings for all of the user's logged-in non-purchased movies.To determine the system's anticipated rating, we first collected the five twitter posts that were the most like the target movie, and used the weighted sum technique to calculate the expected rating. Predicted value falls between 1 and 5 on a scale of 1 to 5. In order to assess the accuracy of the projected ratings made by this module and displayed in the graph, the accuracy measure Mean Absolute Error has been used (MAE).

With the collaborative technique, the dataset is divided into various subsets, and SVMs are trained on each subset separately. The subsets can be made by randomly splitting the dataset or by employing techniques like cross-validation. Each SVM is trained to identify if text documents have a positive or negative sentiment.

Following training, the SVMs' predictions are pooled to produce a final prediction. A voting mechanism is one method of combining the predictions, in which each SVM "votes" for a certain sentiment, and the sentiment with the most votes is selected as the final prediction. Using a weighted average, where each SVM's prediction is weighted based on its accuracy on the validation set, is another technique to combine the predictions.

The suggested method for sentiment analysis uses Twitter data to extract feelings and categorize them as upbeat, negative, or impartial. The strategy involves gathering sentiment data from a variety of tweets and training classifiers for each trend. The data is first transformed into a structured format, and then it is fed into several training models. In order to classify sentiments of real-world data using Twitter, a model must be developed. The study also emphasizes the significance of sentiment analysis in recognising the opinions of the general people on numerous issues as well as the possible advantages of sentiment analysis in predicting stock market values and presidential election outcomes.

V. SYSTEM DESIGN

Fig 2, 3 shows use case diagram and Activity diagram

Use Case Diagram

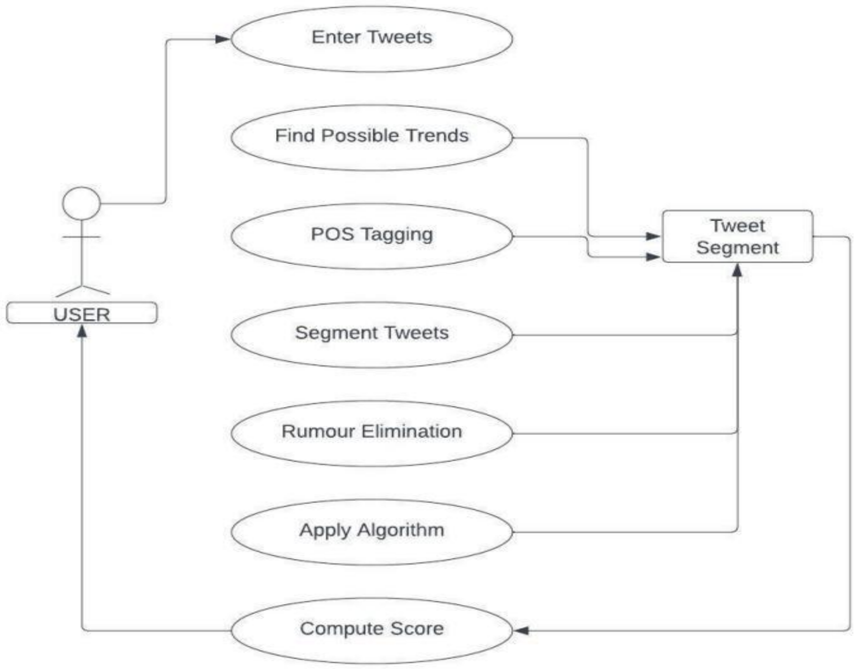


Fig 2. Use Case Diagram

Activity Diagram

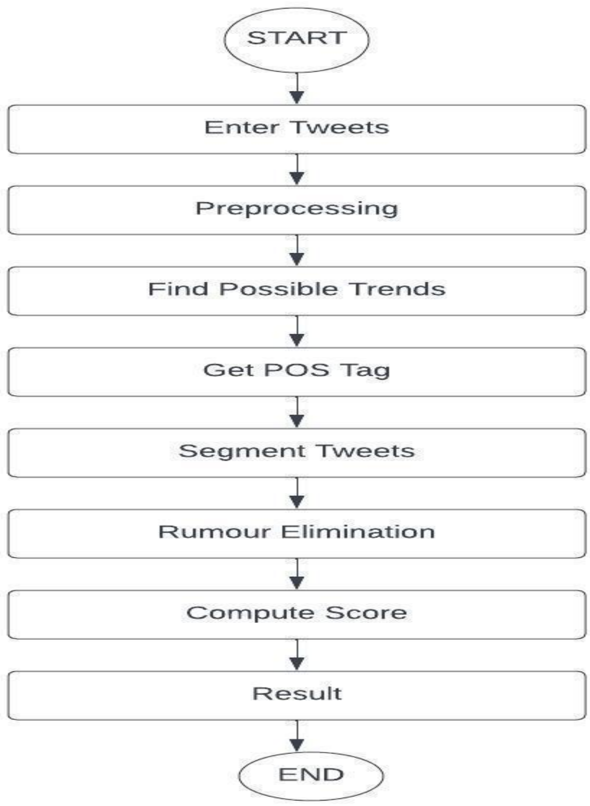
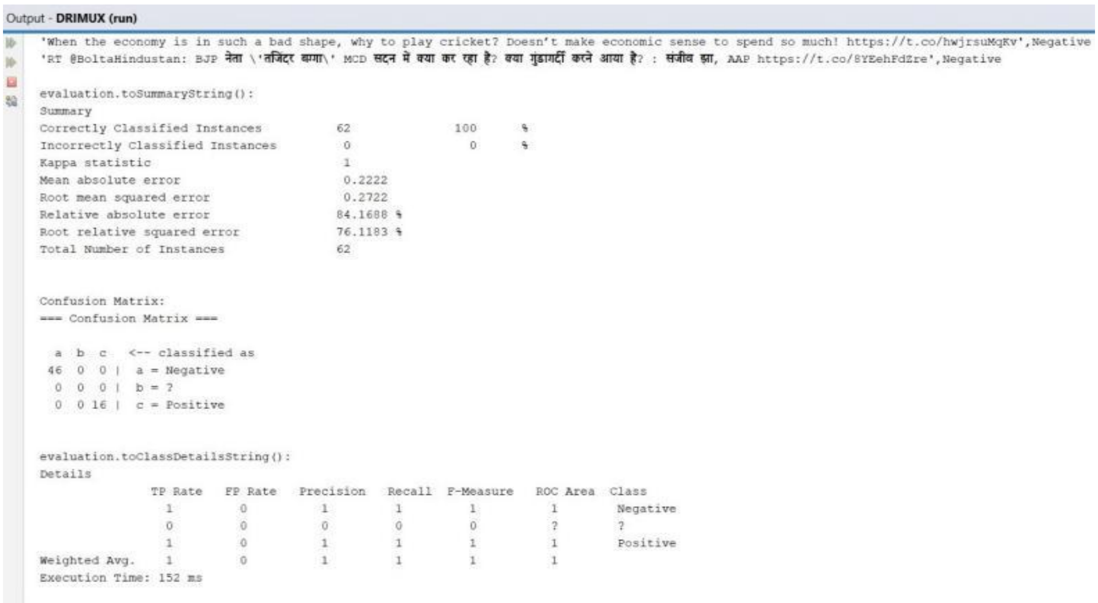


Fig 3. Activity Diagram

Our evaluation which was done on twitter text/content analysis model proposed above is for common users, where the user shares their thoughts on twitter in the form of a comment or a tweet, these are very useful for measuring the tweets classified on the feelings. Among the many solutions that have been developed over the past few decades, twitter asynchronous systems have been employed to help users by suggesting related and pertinent tweets. Numerous developments have been made in this area to get a top-notch, well-tuned Twitter asynchronous system.. When compared to k-means, SVM provide greater accuracy.[7-9] When computing, the k-means algorithm seems to take much time resulting in the have disadvantages in comparison to other organization approaches, so in comparison, k-means is a lethargic method and does not achieve much result in the period of developing a solution.

VI. EXPERIMENTS AND RESULTS

WEKA and SPSS modelers, two widely used tools, were utilized in our trials. Feature selection, clustering, classification, regression, and other modeling algorithms are supported by the commonly used machine learning tool WEKA. Popular data mining software SPSS Modeler has a distinctive graphical user interface and high forecast accuracy. It is extensively used in national security, law enforcement, medical research, resource planning, and business marketing. In each trial, the classification accuracy was assessed using 10-fold cross-validation. The Zero R classifier, which only forecasts the majority class, was used to obtain a baseline accuracy. After putting our model to the test using various K values, we discovered the K value at which the system performs with the greatest precision. On the test set, this model had a classification rate of about 79%. The obtained result using precision, recall and confusion matrix is shown in **Fig 4.**



```
Output - DRIMUX (run)
'When the economy is in such a bad shape, why to play cricket? Doesn't make economic sense to spend so much! https://t.co/hwjrsuMqKv',Negative
'RT @BoltaHindustan: BJP नेता \'तजिंदर बग्गा\' MCD सदन मे क्या कर रहा है? क्या गुंडागर्दी करने आया है? : संजीव झा, AAP https://t.co/8YEehFdZre',Negative

evaluation.toSummaryString():
Summary
Correctly Classified Instances        62        100    %
Incorrectly Classified Instances       0          0    %
Kappa statistic                        1
Mean absolute error                    0.2222
Root mean squared error                0.2722
Relative absolute error               84.1688 %
Root relative squared error           76.1183 %
Total Number of Instances             62


Confusion Matrix:
=== Confusion Matrix ===

  a  b  c   <-- classified as
 46  0  0 |  a = Negative
  0  0  0 |  b = ?
  0  0 16 |  c = Positive


evaluation.toClassDetailsString():
Details
              TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                1        0         1         1        1          1       Negative
                0        0         0         0        0          ?       ?
                1        0         1         1        1          1       Positive
Weighted Avg.   1        0         1         1        1          1
Execution Time: 152 ms
```

Fig 4. Obtained result using Precision, Recall and Confusion Matrix

VII. CONCLUSION AND FUTURE WORK

The current proposal above is for media users, where the number of people messaging can be determined by reviewing the tweets predicated on their emotions.It updates web pages automatically to know where the most latest familiar conversation enclosed by users is.The suggested scheme above displays all twitter posts on Twitter and synchronizes them just on web pages; it's indeed limited and instant to implement.The paper's primary goal would include Twitter data analysis and instantly modifying it on online sites.Back end, we will evaluate the results and automatically update the information in our website for collecting the results in twitter and emotional analysis for every single twitter message to examine the emotions supplied by the twitter user. Our model is distinctive in that it instantly updates hot topics and details on our web page without the requirement for human input.

References

[1]. B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends® in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/1500000011.

[2]. J. Bollen, H. Mao, and A. Pepe, “Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena,” Proceedings of the International AAAI Conference on Web and Social Media, vol. 5, no. 1, pp. 450–453, Aug. 2021, doi: 10.1609/icwsm.v5i1.14171.

[3]. B. O’Connor, R. Balasubramanyan, B. Routledge, and N. Smith, “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series,” Proceedings of the International AAAI Conference on Web and Social Media, vol. 4, no. 1, pp. 122–129, May 2010, doi: 10.1609/icwsm.v4i1.14031.

[4]. M. Hu and B. Liu, “Mining and summarizing customer reviews,” Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Aug. 2004, doi: 10.1145/1014052.1014073.

[5]. S. Kannangara, “Mining Twitter for Fine-Grained Political Opinion Polarity Classification, Ideology Detection and Sarcasm Detection,” Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Feb. 2018, doi: 10.1145/3159652.3170461.

[6]. Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu, “OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media,” IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1763–1772, Dec. 2014, doi: 10.1109/tvcg.2014.2346920.

[7]. N. N. Reddy, E. Naresh, and V. Kumar B.P., “Predicting Stock Price Using Sentimental Analysis Through Twitter Data,” 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Jul. 2020, doi: 10.1109/conecct50063.2020.9198494.

[8]. J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, “Election result prediction using Twitter sentiment analysis,” 2016 International Conference on Inventive Computation Technologies (ICICT), Aug. 2016, doi: 10.1109/inventive.2016.7823280.

[9]. M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, “Sentiment Analysis of Tweets Using Machine Learning Approach,” 2018 Eleventh International Conference on Contemporary Computing (IC3), Aug. 2018, doi: 10.1109/ic3.2018.8530517.

[10]. M. F. Celiktug, “Twitter Sentiment Analysis, 3-Way Classification: Positive, Negative or Neutral?,” 2018 IEEE International Conference on Big Data (Big Data), Dec. 2018, doi: 10.1109/bigdata.2018.8621970.

[11]. S. Kumar, P. Singh, and S. Rani, “Sentimental analysis of social media using R language and Hadoop: Rhadoop,” 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Sep. 2016, doi: 10.1109/icrito.2016.7784953.

[12]. S. Raheja and A. Asthana, “Sentimental Analysis of Twitter Comments on Covid-19,” 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Jan. 2021, doi: 10.1109/confluence51648.2021.9377048.

[13]. T. Chen, R. Xu, Y. He, Y. Xia, and X. Wang, “Learning User and Product Distributed Representations Using a Sequence Model for Sentiment Analysis,” IEEE Computational Intelligence Magazine, vol. 11, no. 3, pp. 34–44, Aug. 2016, doi: 10.1109/mci.2016.2572539.