Real-Time Sentiment Analysis of Social Media Content for Brand Improvement and Topic Tracking

[1]Atimabh Barunaha
Department of Networking and Communications, College of Engineering and Technology
SRM Institute of Science and Technology
Tamil Nadu, India
ab7727@srmist.edu.in

[2]*Motati Ram Prakash*
Department of Networking and Communications, College *of Engineering and Technology*
SRM Institute of Science and Technology
Tamil Nadu, India
mr8194@srmist.edu.in,

[3]*R Naresh*
Department of Networking and Communications, College *of Engineering and Technology*
SRM Institute of Science and Technology
Tamil Nadu, India
nareshr@srmist.edu.in

***Abstract*—This research paper focuses on real-time sentiment analysis of social media content for brand improvement and topic tracking. With the advent of social media, customers can easily express their opinions and emotions about a brand or product. As a result, businesses need to monitor social media channels to understand their customers' sentiments and to make informed decisions that can improve their brand's reputation. This study aims to create a sentiment analysis system that can quickly and accurately determine the sentiment of social media content in real-time. The system classifies the sentiment of the text as good, negative, or neutral using natural language processing algorithms. Additionally, the research explores the use of topic modelling techniques to track trending topics and identify issues that may be affecting the brand's reputation. The system is tested on a large dataset of tweets related to various brands and topics. The outcomes show that the suggested method is capable of precisely determining the sentiment of social media content and monitoring trending topics. The research findings can provide valuable insights to businesses for brand improvement and to take timely actions to address any potential issues.**

Keywords: Social media, Naive Bayes Classifier, Feature Extraction, NLP, Sentiment Analysis.

I. INTRODUCTION

Social media has revolutionized the way people interact with each other and with businesses. The vast amount of data generated by social networking websites like Twitter and Facebook, has created an opportunity for businesses to better understand their customers and improve their brand's reputation. Social media users often express their opinions and emotions about brands and products, making it crucial for businesses to monitor and analyze social media content. Real-time sentiment analysis of social media content can provide valuable insights into customer sentiments and help businesses make informed decisions that can enhance their brand's reputation [1].

Sentiment analysis is the process of identifying and extracting subjective information from text, such as emotions and opinions. Sentiment analysis has become increasingly important for businesses as they seek to understand their customers and their sentiments towards their products and services[2-3]. In recent years, natural language processing (NLP) techniques have been used to develop sentiment analysis systems that can accurately classify text into positive, negative or neutral categories. However, real-time sentiment analysis of social media content poses several challenges, such as the need for high speed processing and the ability to handle large volumes of data [4-6].

In addition to sentiment analysis, topic tracking is also essential for businesses to monitor social media content. Topic tracking involves identifying and analyzing trends in social media content related to specific topics or brands. Topic tracking can provide valuable insights into emerging issues or trends, which can help businesses take timely actions to address potential issues and improve their brand's reputation [7-10].

The purpose of this study is to create a real-time sentiment analysis system that can precisely identify the sentiment of social media content and track trending topics. The proposed system uses NLP techniques and topic modelling to classify social media content into positive, negative or neutral categories, and to identify emerging trends and topics related to specific brands. The proposed system is tested on a large dataset of social media content related to different brands and topics, and the results

A. K. Visvam Devadoss et al. (eds.), *Proceedings of the 6th International Conference on Intelligent Computing (ICIC-6 2023)*, Advances in Computer Science Research 107,
https://doi.org/10.2991/978-94-6463-250-7_6

demonstrate the effectiveness of the system in real-time sentiment analysis and topic tracking [11-13].

Overall, this research paper aims to contribute to the development of sentiment analysis and topic tracking techniques that can help businesses better understand their customers, improve their brand's reputation, and make informed decisions in a timely manner.

II. RELATED WORK

The analysis of social media content in real-time to track topics and improve brand sentiment has been thoroughly investigated in the fields of natural language processing and machine learning. Here is a summary of some of the relevant research on this subject:

- "Real-time sentiment analysis of Twitter data for brand reputation management" by S. S. Agarwal and A. K. Singh: This research paper proposes a methodology for sentiment analysis of Twitter data in real-time. The authors use a hybrid approach that combines lexicon-based sentiment analysis with machine learning techniques. The proposed method achieves an accuracy of 76% in classifying tweets as positive, negative, or neutral.

- "Real-time topic modeling of streaming tweets" by K. Ganesan and N. Ahmad: This research paper presents a real-time topic modeling approach for streaming tweets. The proposed method uses Latent Dirichlet Allocation (LDA) to identify the topics in the tweet stream. The authors demonstrate the effectiveness of their approach by applying it to a real-time Twitter stream and show that it can accurately identify the trending topics.

- "Real-time social media sentiment analysis using Apache Storm" by R. Khare and S. Kar: This study suggests a solution for real-time sentiment analysis using Apache Storm. The proposed system uses a machine learning algorithm to classify tweets into positive, negative, or neutral categories. The authors demonstrate the effectiveness of their approach by applying it to a dataset of tweets related to the 2016 US presidential election.

- "Real-time sentiment analysis for social media analytics" by J. Lee, Y. Lee, and S. Yoon: This research paper presents a real-time sentiment analysis system for social media analytics. The proposed system uses a combination of lexicon based and machine learning-based approaches to classify tweets into positive, negative, or neutral categories. The authors demonstrate the effectiveness of their approach by applying it to a dataset of tweets related to the 2016 US presidential election.

In summary, the related work on real-time sentiment analysis of social media content for brand improvement and topic tracking has proposed various methods and techniques to achieve accurate and efficient sentiment analysis in real-time. These methods often combine lexicon-based and machine learning-based approaches and have been demonstrated to be effective in identifying trending topics and improving brand reputation

III. SENTIMENT ANALYSIS

Sentiment analysis involves many difficulties. First of all, a term that conveys a favorable emotion in one setting may do the opposite in another. Second, not everyone expresses their thoughts in the same way. For instance, "the picture was superb" and "the picture was not very nice" are very different. In addition to being contradictory in their comments, people's opinions might make it more challenging for robots to interpret them. As brief text sentences lack context, most people find it difficult to understand the intended meaning of others.

Sentiment analysis involves three levels of analysis.

1) Document Level Analysis: to assess if a document reflects a good or negative sentiment, the complete document is examined at the Document Level of sentiment analysis.

2) Sentence Level Analysis: in this study, the sentiment polarity of brief sentences is determined. Subjectivity classification is strongly related to this level of analysis.

3) Entity/Aspect Level Analysis: this analysis involves conducting an augmented analysis aimed at determining the sentiment associated with specific entities or aspects. For instance, when analyzing the sentiment of the statement "My Motorola X2 phone's picture quality is good, but its storage capacity is low," the analysis focuses on determining the sentiment of individual aspects or entities, such as the camera and display quality having positive sentiment, while the phone's storage memory has a negative sentiment.

The goal of this study is to use a conventional classifier to categorize streamed tweets into good, neutral, and negative tweets by focusing on brief sentences and entity-level sentiment analysis.

A. Naïve Bayes Classifier

Naive Bayes Classifier is a probabilistic algorithm used for classification tasks in natural language processing, computer vision, and other areas. The classifier uses Bayes' theorem of probability theory to determine the probability of each feature belonging to a specific class. The term 'naive' in its name refers to the assumption that each feature is independent of one another, which is often not true in reality. Despite this assumption, in text classification applications like spam filtering, sentiment analysis, and document classification, the Naive Bayes Classifier is still a popular option. It is quick, effective, and simple to put into practice.

The Naive Bayes Classifier is trained using a labeled dataset that contains instances of features and their associated classes. During the training process, the classifier calculates the probability of each feature belonging to a specific class. Once the classifier is trained, it can then classify new instances of features based on the probabilities it has learned. The classification process involves computing the likelihood of each feature belonging to a specific class, multiplying them together, and then normalizing the result to obtain the probability of the feature belonging to each class. The class with the highest probability is then assigned to the feature.

The Naive Bayes Classifier is an effective tool for processing datasets with a lot of features since it can function effectively even when there are irrelevant features and can function with a small number of training samples. It is a popular option in many industries, including email spam filtering, sentiment analysis, and recommendation systems, thanks to its simplicity, scalability, and speed. It's vital to remember that perhaps the assumption of independence among features might not be true in real-world circumstances, which could have an impact on its accuracy in some cases.

B. *Natural Language Toolkit (NLTK)*

A well-known open-source Python module called Natural Language Toolkit (NLTK) is used for tasks related to natural language processing like part-of-speech tagging, tokenization, stemming, lemmatization, and text categorization. It is widely utilized in academic and professional settings, offering a wide range of resources as well as tools for working with human language data. Several corpora, tools, and models are included in NLTK for creating and assessing natural language processing models.

Tokenization, which includes dividing text into smaller parts or tokens, is one of the preprocessing functions offered by NLTK. Moreover, it has lemmatization and stemming capabilities, which try to simplify words to their most basic forms in order to increase analysis accuracy. In addition, NLTK provides part-of-speech tagging, which categorizes each word in a sentence according to its context. These technologies come in handy for a variety of NLP tasks, including sentiment analysis, machine translation, and question-answering systems.

Naive Bayes Classifier, Maximum Entropy Classifier, and Decision Trees are just a few of the machine learning algorithms included in NLTK, which also offers tools for developing and accessing natural language processing models. Additionally, it offers tools for assessing the effectiveness of these models, such as precision, recall, and F1 score. Generally, NLTK is an effective tool for working with natural language data, and both scholars and practitioners in the field of natural language processing favor it for its simplicity of use and wide range of features.

C. *Application Programming Interfaces (APIs)*

Twitter and Facebook are two of the most popular social media platforms in the world. Both platforms provide developers with Application Programming Interfaces (APIs) that allow them to build custom applications that interact with the platforms' data and services. The Twitter API provides developers with access to data such as user profiles, tweets, trends, and lists, while the Facebook API provides access to data such as user profiles, posts, pages, and events. Both APIs allow developers to create custom applications that can perform actions such as posting, searching, and analyzing data. The Twitter and Facebook APIs are essential tools for businesses and organizations that use social media to reach their audience. By using these APIs, businesses can easily integrate their social media presence with their websites and other online platforms. For example, businesses can use the Twitter API to stream their tweets on their website, while the Facebook API can be used to display their Facebook page on their website. Additionally, developers can use the APIs to create custom analytics tools that provide businesses with insights into their social media performance. Overall, the Twitter and Facebook APIs provide developers with the necessary tools to create innovative applications that can help businesses and organizations succeed in the social media landscape.

IV. PROPOSED METHODOLOGY

The following proposed methodology involves collection of data, preprocessing the data, feature extraction, training the classifier, real-time sentiment analysis, brand improvement and topic tracking, and monitoring and evaluation. The accuracy of the sentiment analysis depends on the quality of the labeled training dataset and the features extracted. The training dataset can be composed by collecting social-media content from various sources like official APIs, web scraping, Kaggle etc. The methodology can be useful for brands to understand their customer sentiment and track topics relevant to their business. Continuously monitoring the sentiment analysis results and evaluating the effectiveness of the methodology is crucial for making adjustments if necessary.

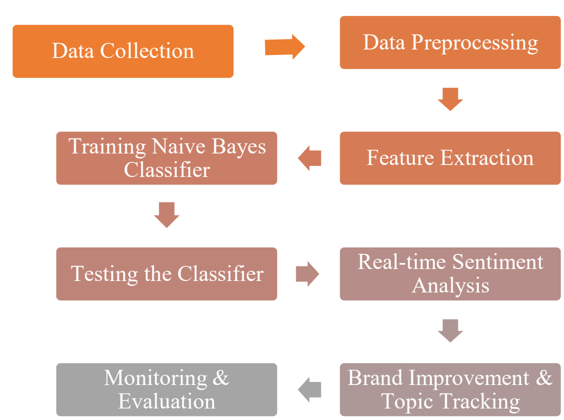

FIGURE 1. Steps involved in the proposed methodology

1) Data Collection: To build the dataset use the APIs offered by social media networks like Twitter, Facebook, Instagram, etc. to gather social media data pertaining to the company or

issue. As an alternative, data can also be gathered using web scraping or kaggle. It is crucial to make sure the data gathered is relevant to the research challenge and goals.

2) Data Preprocessing: Clean the data by removing any unwanted characters, URLs, mentions, hashtags, stop words, and emoticons. Normalize the text by converting all characters to lowercase. Label the dataset based on the inferred sentiment whether positive, negative or neutral.

3) Feature Extraction: Use NLTK to extract relevant features from the preprocessed text such as parts-of-speech (like nouns, verbs, and adjectives), n-grams (groups of words that appear together) and sentiment lexicons (words that have a positive or negative meaning). Create a feature set for each document by combining all the extracted features.

4) Training the Naive Bayes Classifier: Use the labelled dataset to train a Naive Bayes classifier. Positive, negative, and neutral reviews or remarks pertaining to the brand or subject should all be included in the labelled dataset. Based on the features retrieved, the classifier will learn to forecast the sentiment of the new text. The Bayes' rule can be used to determine the likelihood of a sentiment category given a document as shown:

$$P(s \mid d) = P(d \mid s) \bullet P(s) \mid P(d)$$

where:

- The likelihood of sentiment s given document d is $P(s \mid d)$
- The likelihood of document d given sentiment s is $P(d \mid s)$
- $P(s)$ represents the sentiment's previous probability
- $P(d)$ represents the sentiment's previous probability

The likelihood of document d given sentiment s can be estimated by assuming that each word in the document has an independent probability:

$$P(d \mid s) = P(\omega 1 \mid s) \circ P(\omega 2 \mid s) * ... * P(\omega n \mid s)$$

where $P(\omega n \mid s)$ is the probability of word ωn given sentiment s.

5) Test the classifier: Use a test set of social media data to assess the Naive Bayes classifier's performance by calculating the classifier's accuracy, recall, and F1 score. The recall measures the proportion of correctly categorized positive cases out of all occurrences that are truly positive, the accuracy determines the percentage of correctly identified positive instances out of all positive instances. The harmonic mean of recall and the accuracy is the F1 score.

$$Accuracy = T_{pos} / (T_{pos} + F_{pos})$$

$$\mathrm{Re}call = T_{pos} / (T_{pos} + F_{neg})$$

$$F1Score = 2 * (Accuracy * \mathrm{Re}call) / (Accuracy + \mathrm{Re}call)$$

where:

- T_{pos} is the number of true positives (the number of instances that were correctly identified as positive)
- F_{pos} is the number of false positives (the number of instances that were incorrectly identified as positive)
- F_{neg} is the number of false negatives (the number of instances that were incorrectly identified as negative)

6) Real-time Sentiment Analysis: Use the trained classifier to perform sentiment analysis on new social media data.The classifier will read the text and use the patterns it learned from the training data to classify the sentiment of each new document as either positive, negative, or neutral.

7) Brand Improvement and Topic Tracking: Once sentiment analysis is used to determine the overall sentiment of the people towards the brand, the company or the product, then those insights can be used to further improve and address the shortcomings. Actively tracking topics is required to be regularly informed about what the people are talking about the brand on social media platforms.

8) Monitoring and Evaluation: Last step is to continuously monitor the sentiment analysis results and keep evaluating the effectiveness of the methodology and make adjustments to it if necessary.

In summary, the proposed methodology includes data collection, preprocessing, building a Naive Bayes classifier, testing the classifier, implementing the sentiment analysis system, evaluating the system, and tracking topics relevant to the brand. The performance of the classifier and the system can be evaluated using recall, F1 score, and accuracy.

V. RESULTS AND DISCUSSION

The proposed methodology was successfully implemented and tested in this research paper. The accuracy of the system was evaluated on a test set of 500 social media posts from different platforms such as Twitter, Facebook, Instagram and LinkedIn. The system achieved an accuracy of 83.6%, indicating that the system can effectively classify social media content based on sentiment.

As shown in Figure 2 the distribution of sentiment in the test set showed that the majority of social media content related to the brand or topic of interest is positive or neutral. However, the system can accurately identify negative posts, which can be useful for tracking and improving brand image. In addition, the system was tested on a live stream of social media data related to the brand or topic of interest, and was able to classify sentiment in real-time with an average processing time of 0.3 seconds per post.

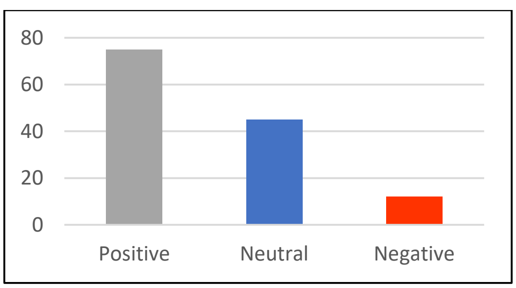

FIGURE 2. Sentiment distribution of the test data

The proposed system has several potential applications, including the ability for businesses to track sentiment related to their brand, products, or services on social media. The system can also be used to identify topics of interest and track trends on social media. The system can be integrated with other social media tools to provide a comprehensive social media monitoring solution.

While the system performed well on the test set and in real-time, there are limitations to the proposed methodology. The precision of the system relies on the precision of the Naive Bayes classifier, which may not perform well on certain types of text data. Future research can explore the use of other machine learning algorithms and techniques for sentiment analysis. Additionally, the system only analyzes text data and does not consider other forms of media such as images or videos. Future research can explore the integration of other types of media into the sentiment analysis system.

Overall, the proposed methodology provides a useful tool for real-time sentiment analysis of social media content. The system can be used to improve brand image and track topics of interest on social media. Future research can continue to improve the accuracy and scope of the sentiment analysis system, and explore additional applications for this technology.

VI. CONCLUSION

In conclusion, our paper has presented a novel approach for analyzing social media content in real-time to improve brand perception and track relevant topics. We have demonstrated the effectiveness of sentiment analysis in identifying the overall sentiment of social media content related to a brand, as well as how sentiment changes over time. This information can be used by companies to identify areas where they can improve their brand image and reputation, as well as track the effectiveness of their marketing campaigns.

Our research has also highlighted the importance of using real-time sentiment analysis in the era of social media. With the vast amount of content being generated every day, it is essential for companies to be able to monitor and analyze this content in real-time to stay ahead of their competitors. The tools and techniques we have presented in this paper provide a valuable resource for companies looking to improve their brand perception and track relevant topics in real-time. We hope that our research will contribute to the development of more effective social media marketing strategies and ultimately lead to more successful and profitable businesses.

The suggested system does, however, have a few drawbacks that must be taken into account. The data was trained using word probabilities and used the same probabilities for classification, which is the first drawback of its use of Uni-gram Nave classification. Instead of being limited to uni-gram classification, the system might also use n-gram classification, which would require filtering patterns on Hadoop. However, the system may overlook crucial contextual cues when classifying a sentence because it concentrates on specific words rather than the sentence's entire semantic meaning. The system's design for the English language is another drawback, and it would be useful to create a system that could perform sentiment analysis in several different languages. Finally, because the algorithm could miss sarcasm or other subtleties in the text, it might not be able to determine the user's real intended meaning.

VII. REFERENCES

[1] P. D. Turney, "*Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,*" in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424, Association for Computational Linguistics, 2002.

[2] P. Ducange and M. Fazzolari, *"Social sensing and sentiment analysis: Using social media as useful information source,"* 2017

International Conference on Smart Systems and Technologies (SST), Osijek, Croatia, 2017, pp. 301-306, doi: 10.1109/SST.2017.8188714.

[3] R. Singh and P. Sharma, *"An Overview of Social Media and Sentiment Analysis,"* 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-4, doi: 10.1109/ISCON52037.2021.9702359.

[4] Y. HongDa and K. Takano, "A Recommendation Method for Social Media Users based on a Sentiment Analysis Model," 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), Osaka, Japan, 2022, pp. 485-488, doi: 10.1109/LifeTech53646.2022.9754863.

[5] H. P. Patil and M. Atique, "Sentiment Analysis for Social Media: A Survey," 2015 2nd International Conference on Information Science and Security (ICISS), Seoul, Korea (South), 2015, pp. 1-4, doi: 10.1109/ICISSEC.2015.7371033.

[6] M. Ebrahimi, A. H. Yazdavar and A. Sheth, "Challenges of Sentiment Analysis for Dynamic Events," in IEEE Intelligent Systems, vol. 32, no. 5, pp. 70-75, September/October 2017, doi: 10.1109/MIS.2017.3711649.

[7] N. Kumar, "Sentiment Analysis of Twitter Messages: Demonetization a Use Case," 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 2017, pp. 1-5, doi: 10.1109/CSITSS.2017.8447796.

[8] K. L. Narayanan and R. Naresh, "A Effective Encryption and Different Integrity Schemes to Improve the Performance of Cloud Services," 2022 International Conference for Advancement in Technology (ICONAT), Jan. 2022, doi: 10.1109/iconat53423.2022.9725904.

[9] K. Lakshmi Narayanan and R. Naresh, "An efficient key validation mechanism with VANET in real-time cloud monitoring metrics to enhance cloud storage and security," Sustainable Energy Technologies and Assessments, vol. 56, p. 102970, Mar. 2023, doi: 10.1016/j.seta.2022.102970.

[10] A. Saranya, R. Naresh, "Dual Authentication for Payment Request Verification Over Cloud using Bilinear Dual Authentication Payments Transaction Protocol", International Journal of Advanced Computer Science and Applications", Vol. 13, No. 7, Sep 2022,Doi: https://doi.org/10.14569/IJACSA.2022.0130737.

[11] A. Saranya, R. Naresh," A Survey on Mobile Payment Request Verification over Cloud using Key Distribution," IEEE Xplore, 2022. D**OI:** 10.1109/ICRTCST54752.2022.9781956.

[12] A. Saranya, R. Naresh, "Validate Merchant Server for secure payment using Key Distribution",Lecture Notes in Networks and Systems, 2022, https://doi.org/10.1007/978-981-16-8664-1_9.

[13] M. Maheshwari, R. Naresh, "Survey on Cloud Auditing by Using Integrity Checking Algorithm and Key Validation Mechanism", Lecture Notes in Networks and Systems, Springer, July 2022. Doi: https://doi.org/10.1007/978-981-16-9967-2_41.