

Sentiment and Social Media Trend Analysis Using Big Data

Bhavana S

MASTER OF COMPUTER APPLICATIONS

BAPUJI INSTITUTE OF ENGINEERING & TECHNOLOGY, DAVANAGERE

Dr. GEETHA M

MASTER OF COMPUTER APPLICATIONS

BAPUJI INSTITUTE OF ENGINEERING & TECHNOLOGY, DAVANAGERE

ABSTRACT

Social media platforms generate massive volumes of user-generated content in the form of text posts, comments, hashtags, images, and interactions. Analysing this data provides valuable insights into public opinion, brand perception, political sentiment, and emerging trends. The high velocity, volume, and variety of social media data require scalable big data processing and advanced machine learning techniques for effective sentiment classification and trend detection. This paper presents a comprehensive framework for sentiment and social media trend analysis using big data technologies. The proposed system integrates data collection from social platforms, pre-processing pipelines, natural language processing techniques, machine learning and deep learning models, and trend detection mechanisms. Experimental results demonstrate that transformer-based models outperform traditional machine learning approaches in sentiment classification accuracy, while time-series analytics effectively identify trending topics. The system enables organizations to monitor public opinion in real time and make data-driven decisions.

Keywords: *Sentiment Analysis, Social Media Analytics, Big Data, Natural Language Processing, Trend Detection, Deep Learning*

I.INTRODUCTION

The art of communication and the exchange of thoughts and opinions between people have seen a radical transformation with the emergence of technology and the development of various social media. There are hundreds of millions of users on social media like X, LinkedIn, and Reedit offering their opinions and thoughts, comments, and discussion forums. Every second, these social media causes tremendous textual data, which is based on collective public reactions, opinions, and views of people. Instead of viewing all such data as random online messages, I explored it as collective public behaviour and opinion. This amount of data can never be processed by the usual processing mechanisms in a database that has the capability to process the velocity and variety of data emanating from social platforms at a very high pace in case of millions of entries at a time.

This demand led to the introduction of a massive data-based totally dispensed computing device. This makes it viable to cut out needless characters from textual content, transform sentences to tokens, collect fundamental functions, and examine characteristic polarity to decide sentiment in social statistics. Different applications can use evaluation outcomes. Opinion evaluation is prime to brand reputation, client opinion, and rival marketplace analysis. Opinion evaluation can assist government provider vendor's reply to citizen critiques after social/herbal failures. However, social media studies includes technical demanding situations. Social media uses slang, emojis, abbreviations, and sarcasm. The text is not understood due to those. Moreover, social media

systems' perceptions trade quickly. This means that the analysis outcomes might be old quick.

This paper proposes a scalable big data framework for sentiment classification and trend detection that supports real-time monitoring and predictive insights.

II.LITERATURE REVIEW

In the early stages, sentiment analysis mainly depended on lexicon-based methods, where predefined sentiment dictionaries were used to classify text. These methods were easy to understand and implement, but they often failed to handle complex language features such as sarcasm, context, and domain-specific meanings. Previous methods have been incompatible with the speed, quantity, and variety of internet records.

Sentiment analysis solutions fall into the subsequent categories: Lexicon-based and classical machine mastering techniques SentiWordNet and VADER had been key lexicon-primarily based algorithms in early sentiment analysis structures. Word-based totally lexicons estimated a textual content's sentiment polarity the use of a sentiment score, which was once smooth to assemble but lacked a context-aware method to provide correct effects, mainly in complicated or casual terms. Conventional system learning techniques, consisting of decision trees, SVM, and Naïve Bayes Classifiers, were to start with used to improve accuracy. These techniques extracted traits using TF-IDF and Bag of words. Those functions forget order and meaning, making social media sarcasm, slang, and sentence shape difficult to understand.

With advancements in technology, there was an increase in the complexity of deep learning architectures such as CNN, LSTM, and GRU, making it simpler to comprehend the context and meaning behind texts, thus resulting in better outcomes. Recently, transformer architectures like BERT and RoBERTa have made further improvements in comprehending the underlying context in texts.

Other techniques which were not developed in the context of sentiment analysis but researched by other researchers include trend detection techniques that use pattern recognition and topic models such as Latent Dirichlet Allocation, and hashtag detection. In light of the fact that

there is enormous amount of data generated through social media, it is common practice to analyse data using software packages such as Hadoop and Apache Spark.

III.DESIGN AND IMPLEMENTATION

The Sentiment and Social Media Trend Analysis system works in a step-by-step process whereby the information extracted from social media sites undergoes a series of transformations to make it usable. The first step involves the collection of live tweets via the Twitter API and passing the information to Apache Kafka for efficient management of data streams. In the second step, text data analysis is performed through NLP techniques, whereby all the irrelevant parts such as URLs, emoticons, and stop words are removed from the texts.

Once this process is completed, the processed output will then be passed to the sentiment analysis model, where machine learning or deep learning techniques will be utilized for categorizing the text into either a positive, negative, or neutral category. Thereafter, the findings will be presented in graphs or word clouds for ease of understanding. This will allow the users to interact with the dashboard, helping them identify any emerging trends. The overall process works well right from the data collection stage to data presentation.

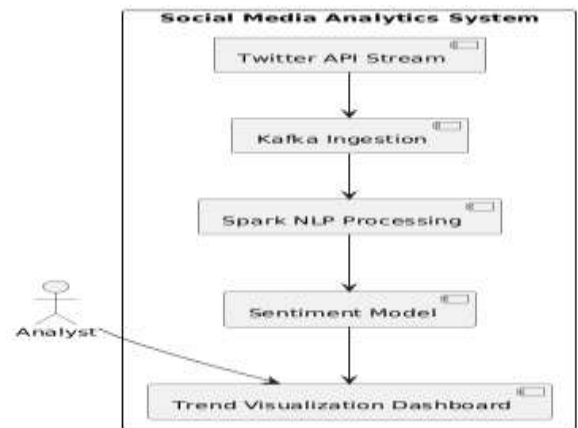


Figure: Architecture design & implementation of Sentiment & Social Media Trend Analysis

IV.PROPOSED SYSTEM

The project, “Sentiment & Social Media Trend Analysis Using Big Data,” connects raw data input to actionable intelligence. This system captures live social media streams using Apache Kafka as a high-throughput

communications backbone, unlike typical methods. Apache Spark is used for distributed processing to speed up text sanitization and feature engineering across a cluster. In the proposed system used hybrid deep learning to overcome keyword-based analysis's limitations. The system parses human speech's subtle context and sequential dependencies using BERT and LSTM networks.

The system also included Latent Dirichlet Allocation (LDA) for automatic topic modelling to group huge datasets into subject clusters and trending hashtags without manual involvement. Integrating an interactive visualization suite is the final system layer for accessibility. The dashboard shows changing sentiment scores and "hashtag clouds". The framework may analyze historical archives and monitor live events using Lambda architecture (batch and streaming). The tool is ideal for brand monitoring and crisis management due to its dual-stream functionality and automated alert system for emotional spikes.

ADVANTAGE OF PROPOSED SYSTEM

There are many advantages to the suggested system over traditional systems. One of those advantages is the real-time analysis of social media with the help of stream processing. Since the system has a distributed architecture, it means that the system will have infinite scalability with the capacity to handle millions of social media posts without compromising on its efficiency.

Deep learning techniques can be applied to improve the sentiment accuracy of social media posts. In the proposed system, trends, sentiments, hashtags, events, and keyword analysis of social media posts will be studied. Multilingualism and code-switching of social media posts in India will also be analyzed. Big Data solutions will be employed for data analysis in the proposed system. Dashboard tools will be used for visualization purposes.

V.EXPIMENTAL RESULTS

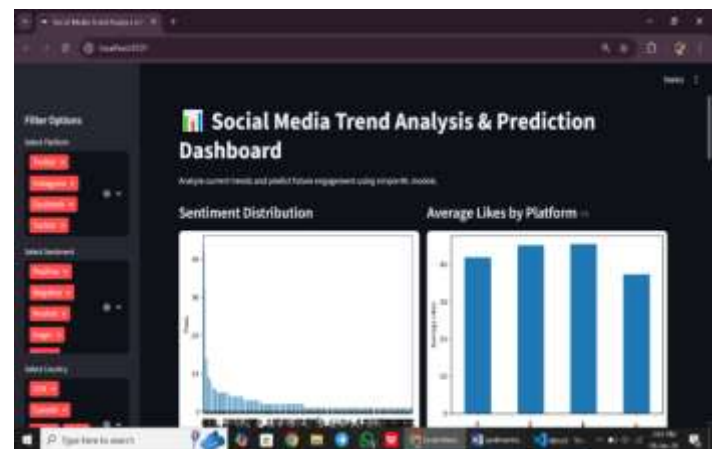
Based on the experimental results, it is evident that the Social Media Sentiment and Trend Analysis System is quite efficient when it comes to analyzing huge amounts of social media content. In this regard, it has been possible for the system to categorize all post content depending on

whether they are positive, negative or neutral through analysis based on the provided dataset. The graphical

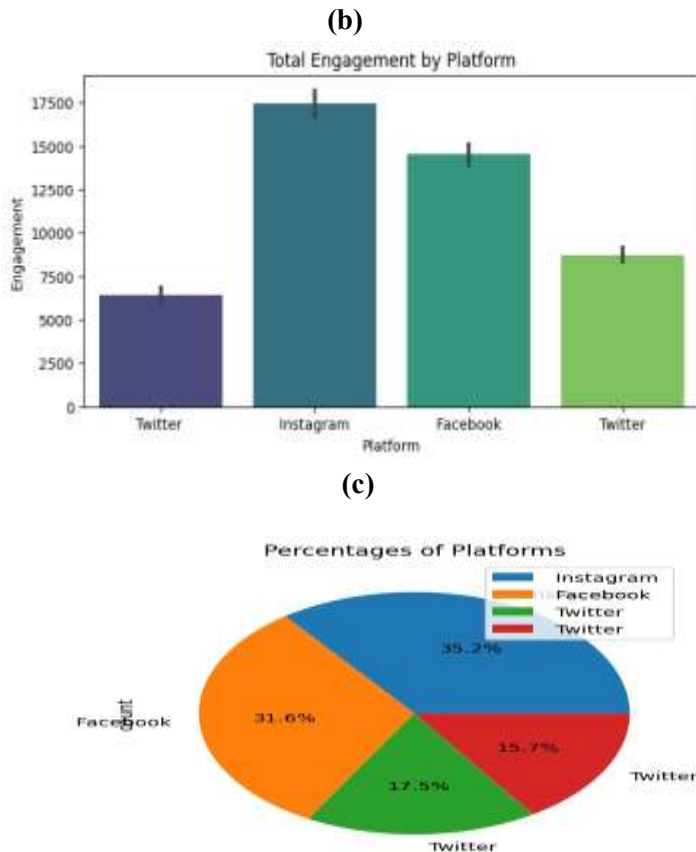
depiction of the sentiments distribution facilitated by the system makes it easier for the different sentiments to be compared among various countries and social media platforms.

The system also gave valuable information regarding the behaviour of users. This was done by analyzing the average number of likes on various platforms, enabling the identification of platforms that get more attention from the users. An hourly analysis of posts also enabled us to establish the optimal time to post messages, when the number of users is highest. This shows that the system not only analyses sentiment but also gives detailed information about the behavior of users.

The system's ability to make predictions using Predictive Analysis, conducted by the means of the Linear Regression method, showed that predictions of the future level of engagement based on the previous one can be made. Using analysis of daily average likes, it became possible to get short-term predictions that corresponded to the current trends, provided that there was enough information for that. Though the methodology used in this case is rather simple, yet it gives quite accurate predictions, as compared to more advanced ones.



(a)



VI.CONCLUSION

The Social Media Sentiment and Trend Analysis System has demonstrated the effectiveness of applying NLP and Big Data technologies for analyzing social media content on a massive scale. It efficiently analyses textual data, recognizes sentiments as positive, negative, or neutral, and finds trends and patterns within them. Adhering to the defined workflow that involves data pre-processing, feature extraction, training a machine learning model, and visualization, the project guarantees the accuracy and reliability of results. The application of machine learning algorithms and the use of efficient tools for real-time processing of large datasets prove the ability of the proposed solution to be implemented in practice for marketing research, customer opinion monitoring, and other purposes.

In the future, the system will become more efficient with the inclusion of several regional languages that will allow it to comprehend various kinds of content. Additionally, some advanced functions should be integrated, such as sarcasm detection and emotions' recognition because people tend to conceal their emotions. This will enhance

the accuracy and credibility of predictions, making them closer to actual human perception.

Moreover, to make the system more efficient, its functionality should be expanded to cover visual elements alongside textual content. Location-based prediction and trending analysis, along with push notifications for real-time updates, are other important components that could be included. As a result, the project will evolve into an efficient, sophisticated, and industrial-grade product.

VII.REFERENCES

- [1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, 2008.
- [2] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," Stanford University Technical Report, 2009.
- [3] Y. Kim, "Convolutional Neural Networks for Sentence Classification," EMNLP, 2014.
- [4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.
- [5] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2019.
- [6] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," ICLR, 2013.
- [7] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, 2003.
- [8] M. Zaharia et al., "Apache Spark: A Unified Engine for Big Data Processing," Communications of the ACM, 2016.
- [9] H. Kwak et al., "What is Twitter, a Social Network or a News Media?" WWW Conference, 2010.
- [10] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," LREC, 2010.
- [11] F. Barbieri et al., "Emoji Sentiment Ranking," LREC, 2016.