AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 3(2026)

SENTIMENT ANALYSIS OF SOCIAL MEDIA PRESENCE

[1]Dr. Ratna Raju Mukiri,[2]Thuraka. Alekhya,[3]Seelam. Sirisha,[4]Sanaka. Naga Mounika

[1]Professor, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

[2,3,4]U. G Student, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

ABSTRACT-*This paper presents a Sentiment Analysis of Social Media Presence system using Machine Learning (ML) and Natural Language Processing (NLP). The main objective of this work is to identify whether social media posts or comments express positive, negative, or neutral sentiments. With the rapid growth of social media platforms, people share a large number of opinions, reviews, and experiences every day. Analyzing this information manually is difficult, time-consuming, and may produce inaccurate results. To overcome this problem, the proposed system automatically analyzes text using Machine Learning techniques. The input text is first cleaned and preprocessed using the NLTK library by removing unwanted words and symbols. Next, the TF-IDF Vectorizer converts the processed text into numerical features. The Logistic Regression algorithm is then used to classify the sentiment of the given text as positive, negative, or neutral. The system is developed using the Flask framework, which provides a simple and user-friendly interface and displays the predicted sentiment along with its confidence score. The proposed system provides a fast, accurate, and efficient solution for sentiment analysis. It can be applied in customer feedback analysis, product review analysis, brand reputation monitoring, and market research to help organizations make better decisions based on public opinion.*

KEYWORDS: *Machine Learning (AI), Natural Language Processing (NLP), Logistic Regression, TF-IDK, NLTK, Flask, Text-Classification, Opinion Mining.*

INTRODUCTION

Social media has become an important platform where people share their opinions, feelings, and experiences through posts, comments, and reviews. Every day, millions of users post their thoughts on platforms like Twitter, Facebook, Instagram, and YouTube. Analyzing all this information manually is difficult and takes a lot of time. Sentiment Analysis helps solve this problem by using Machine Learning (ML) and Natural Language

Processing (NLP) to automatically identify whether a post or comment is positive, negative, or neutral. In this project, the input text is first cleaned and processed to remove unwanted words and symbols. Then, the TF-IDF Vectorizer converts the text into numerical data, and the Logistic Regression algorithm predicts the sentiment of the text. The system also shows the analysis results in graphical form, making them easy to understand. This helps businesses, organizations, and researchers know what people think about their products, services, or brands. By automating the sentiment analysis process, the system saves time, reduces manual work, and supports applications such as customer feedback analysis, product review analysis, brand reputation monitoring, and market research.

RELATED WORK

Several researchers have contributed to the field of sentiment analysis with the objective of understanding public opinions, emotions, and attitudes expressed through social media platforms. Early studies focused on rule-based and lexicon-based approaches, where predefined dictionaries containing positive and negative words were used to determine sentiment. These methods were simple and easy to implement but often failed to capture the actual meaning of sentences, especially when dealing with sarcasm, slang, and contextual variations commonly found in social media content. To address these limitations, researchers introduced Machine Learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), Decision Trees, Random Forests, and Logistic Regression for sentiment classification. These approaches improved accuracy by learning patterns from labeled datasets rather than relying solely on predefined rules. More recently, Natural Language Processing (NLP) techniques such as text preprocessing, tokenization, stemming, and feature extraction methods like TF-IDF have been widely adopted to enhance sentiment prediction. The proposed work builds upon these advancements by integrating NLP techniques, TF-IDF Vectorization, and the Logistic Regression algorithm to efficiently classify social media comments into positive, negative, and neutral sentiments, providing meaningful insights through an interactive and user-friendly interface.

LITERATURE SURVEY

Sentiment analysis has been extensively studied by researchers to understand opinions expressed through social media and online platforms. Pang and Lee (2008) introduced the concept of opinion mining using machine learning techniques and demonstrated that text classification

algorithms can effectively identify positive and negative sentiments. Bing Liu (2012) further expanded this research by presenting both lexicon-based and machine learning approaches for analyzing customer reviews and social media content. His work highlighted the importance of sentiment analysis in business intelligence and decision-making. Go, Bhayani, and Huang (2009) proposed a Twitter sentiment analysis model using distant supervision and compared algorithms such as Naïve Bayes, Support Vector Machine, and Maximum Entropy. Their study showed that machine learning methods significantly improve classification accuracy for social media data. Hutto and Gilbert (2014) developed the VADER sentiment analysis model, which is specifically designed for social media text and effectively handles slang, emoticons, and abbreviations. Based on the findings of these studies, the proposed system adopts Natural Language Processing (NLP) techniques, TF-IDF Vectorization, and the Logistic Regression algorithm to achieve accurate, efficient, and reliable sentiment classification. The system provides meaningful insights into public opinion by classifying social media comments into positive, negative, and neutral categories, making it suitable for customer feedback analysis, brand monitoring, and market research.

EXISTING SYSTEM

The existing sentiment analysis systems mainly rely on manual analysis, rule-based methods, or lexicon-based approaches to identify the sentiment of social media text. These methods use predefined dictionaries of positive and negative words to classify user opinions. Although they are simple to implement, they often fail to understand the context, sarcasm, abbreviations, emojis, and informal language commonly used on social media platforms. Some existing systems use machine learning algorithms such as Naïve Bayes, Decision Trees, and Support Vector Machines (SVM) for sentiment classification. While these approaches improve prediction accuracy, they may require extensive feature engineering and large amounts of labeled data. In addition, many traditional systems are unable to process large volumes of real-time social media data efficiently and often lack interactive visualization of sentiment results. They also struggle to maintain high accuracy when dealing with noisy and unstructured text. These limitations reduce the effectiveness of existing sentiment analysis systems in real-world applications. Therefore, there is a need for a more efficient, accurate, and user-friendly solution that can automatically classify social media comments into positive, negative, and neutral sentiments while

providing meaningful graphical representations of the analysis results.

PROPOSED SYSTEM

The proposed system uses Machine Learning (ML) and Natural Language Processing (NLP) to analyze the sentiment of social media text. It automatically identifies whether a user's post, comment, or review expresses a positive, negative, or neutral opinion. First, the input text is cleaned by removing unwanted words, punctuation marks, special characters, and other unnecessary symbols using the NLTK library. After preprocessing, the TF-IDF Vectorizer converts the text into numerical features that can be understood by the machine learning model. The Logistic Regression algorithm is then used to predict the sentiment of the given text accurately. The system is developed using the Flask framework, which provides a simple and user-friendly web interface. Users can enter any text and instantly view the predicted sentiment along with its confidence score. The system also displays the results in graphical form, making them easy to understand. This approach reduces manual work, improves the accuracy of sentiment prediction, and supports quick analysis of social media data. It can be used for customer feedback analysis, product review analysis, brand reputation monitoring, and market research, helping organizations make better decisions based on public opinion.

SYSTEM ARCHITECTURE

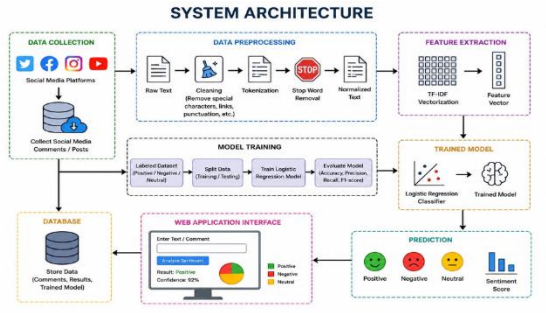

Fig1: System Architecture

MODULE DESCRIPTION

The methodology of the Sentiment Analysis of Social Media Presence system is designed to analyze and classify user opinions shared on social media using Machine Learning (ML) and Natural Language Processing (NLP). The system follows a step-by-step process that includes data collection, text preprocessing, feature extraction, model training, sentiment prediction, and result visualization. First, social media comments or reviews are collected from the dataset. The input text is then cleaned using the NLTK library by removing stop words, punctuation marks, URLs, special characters, and other unwanted symbols. After preprocessing, the TF-IDF Vectorizer converts the cleaned text into numerical features that can be used by the machine learning model. These features are used to train the Logistic

Regression algorithm to identify different sentiment categories. Once the model is trained, it predicts whether a new comment is positive, negative, or neutral. The application is developed using the Flask framework, where users can enter text and view the predicted sentiment along with its confidence score. The results are also displayed in graphical form, making them easy to understand. This methodology provides accurate, fast, and reliable sentiment analysis of social media data.

RESULTS AND DISCUSSION

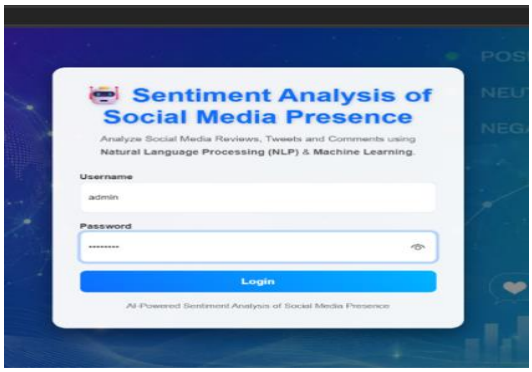

Fig2: Login Page

The login page authenticates users by validating their username and password before granting access to the system

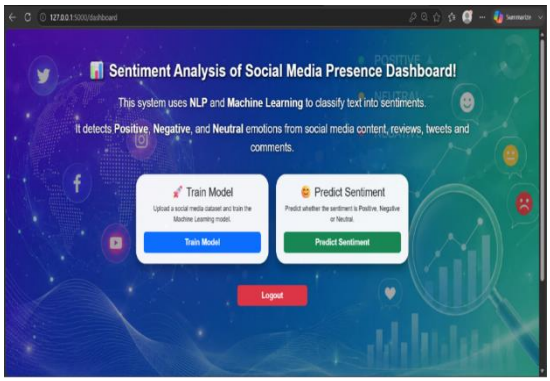

Fig3: Home dashboard page

The dashboard provides options to train the sentiment analysis model and predict the sentiment of social media text.

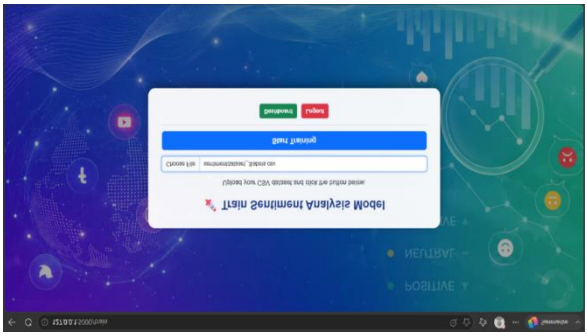

Fig4: Model Train page

Upload the sentiment dataset and train multiple machine learning models for sentiment classification.

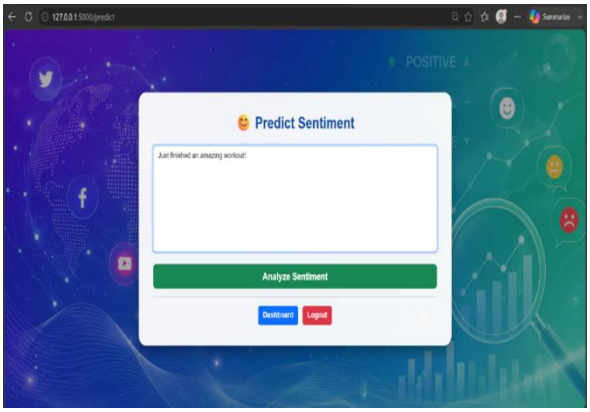

Fig5: Predict Sentiment page

Allows users to enter social media text and analyze its sentiment using the trained machine learning model.

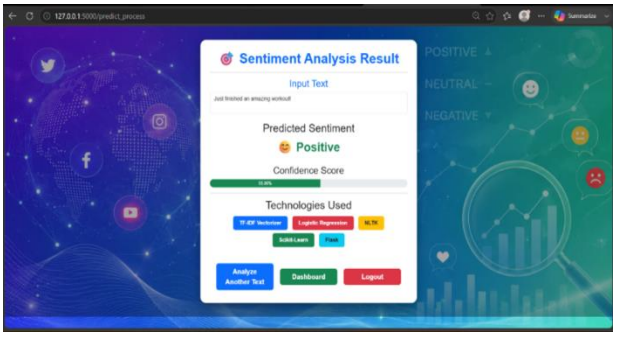

Fig6: Result page

Displays the input text, predicted sentiment, confidence score, and technologies used for the prediction.

CONCLUSION

The Sentiment Analysis of Social Media Presence system helps understand public opinions by automatically classifying social media posts, comments, and reviews as positive, negative, or neutral. It uses Machine Learning (ML), Natural Language Processing (NLP), and the Logistic Regression algorithm to predict sentiments accurately. The system provides a simple and efficient way to analyze customer feedback and public opinions. It can be used to monitor brand reputation, analyze product reviews, and support market research. With its user-friendly interface and reliable sentiment prediction, the system helps organizations make better decisions based on social media data.

FUTURE SCOPE

The proposed Sentiment Analysis of Social Media Presence system can be improved by supporting real-time analysis, multiple languages, and emotion detection. Advanced Machine Learning and Deep Learning techniques can also be used to improve prediction accuracy. These enhancements will make the system more efficient and useful for business and research applications.

REFERENCES

1. D. P. Harini, “Electronic Health Record Stage Identification using Principal Component Analysis,” *Machine Intelligence Research,* vol. 18, no. 01, pp. 864- 873, 17 8 2024.
2. D. P. Harini, “Two Level Intrusion Detection For Detecting Intruders in Multitier Web Applications,” *International Journal of Engineering & Science Research,* vol. 3, no. 9, pp. 472-478, 2013.
3. D. P. Harini, “Analyze and Predict of Human Cyber Attackers using Artificial Neural Network,” *Journal of Basic Science and Engineering,* vol. 21, no. 1, pp. 1529 - 1536, 8 7 2024.
4. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," in Proceedings of NAACL-HLT, 2014.
5. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge University Press, 2008.
6. T. M. Mitchell, Machine Learning. McGraw-Hill, 1997.
7. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
8. S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. O'Reilly Media, 2009.

9. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
10. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," Proceedings of the First Instructional Conference on Machine Learning, 2003.
11. G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.
12. T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Springer, 1998.
13. Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proceedings of EMNLP, 2014.
14. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, 2019.
15. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.