International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

DETECTING EMOTION IN TWEETS USING HYBRID VOTING CLASSIFIER (LR-SGD)

Bhabanisankar Bhol
PG, Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
bhabanimca018@gmail.com

Mridula Shukla
Associate Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
mridulatewari@theoxford.edu

ABSTRACT

The rapid growth of social media platforms has made Twitter a significant source of real-time public expression, where users frequently share opinions, emotions, and sentiments. Detecting emotions from these short and informal texts is an essential task in natural language processing, with applications in mental health monitoring, marketing, and social trend analysis. However, the unstructured nature of tweets, combined with the presence of slang, abbreviations, and noisy data, makes emotion classification a challenging problem. In this study, we propose a hybrid voting classifier that integrates Logistic Regression (LR) and Stochastic Gradient Descent (SGD) to improve the accuracy and robustness of tweet-based emotion detection. The hybrid approach leverages the interpretability of LR and the scalability of SGD, ensuring both efficiency and adaptability to large-scale datasets.

Experimental evaluation demonstrates that the proposed model outperforms individual classifiers by reducing misclassification rates and handling high-dimensional text features effectively. This research highlights the potential of hybrid ensemble techniques for real-world applications where emotions play a key role in decision-making and human-computer interaction.

INTRODUCTION

In moment's digital period, social media has come one of the most influential platforms for people to partake their passions, studies, and opinions incontinently. Among them, Twitter stands out due to its terse format and wide global operation, making it a rich source of textual data. These tweets frequently reflect mortal feelings ranging from happiness and excitement to wrathfulness, sadness, and fear.

Understanding similar feelings is pivotal for several disciplines including internal health analysis, client feedback monitoring, political opinion mining, and perfecting mortal-computer commerce. still, detecting feelings from tweets is a complex task because of informal language, bowdlerizations, misspellings, and the limited environment handed by short dispatches. Traditional bracket styles frequently struggle with these challenges. To address this, machine literacy ways have been extensively explored for emotion discovery. In this exploration, we propose a mongrel voting classifier combining Logistic Retrogression(LR) and Stochastic Gradient Descent(SGD). This model leverages the strengths of both algorithms, aiming to ameliorate delicacy, handle noisy data, and give a scalable result for large tweet datasets. V

LITERATURE SURVEY

Emotion Discovery from textual data has been an active exploration area in natural language processing due to its wide range of operations. Early studies primarily concentrated on sentiment analysis, which classified textbook into positive, negative, or neutral orders. While useful, this approach was limited in landing the depth of mortal feelings. Experimenters gradationally shifted towardmulti-class emotion discovery, inspired by cerebral models similar as Ekman's six introductory feelings. Traditional machine literacy ways like Naïve Bayes, Support Vector Machines(SVM), and Decision Trees were extensively used, but their performance was frequently confined by the high dimensionality and noisy nature of social media textbook. With the arrival of deep literacy, models similar as Convolutional Neural Networks(CNN) and intermittent Neural Networks(RNN) showed significant advancements by learning contextual features. still, these approaches demand large labeled datasets and high computational coffers. Recent exploration has explored mongrel and ensemble styles, combining the strengths of multiple classifiers to achieve better conception. This work builds upon similar approaches by proposing a mongrel voting classifier using Logistic Retrogression and Stochastic grade Descent for robust and scalable emotion discovery in tweets.

EXISTING WORK

Previous research on emotion detection in text has largely focused on sentiment analysis, which classifies opinions into simple categories such as positive, negative, or neutral. While effective for broad analysis, this method

fails to capture the complexity of human emotions expressed in tweets. Machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees have been commonly applied, showing moderate success in handling structured data. However, these models struggle with the informal, short, and noisy nature of tweets. Deep learning models like CNN and RNN further improved accuracy by capturing contextual and sequential features, but they require large annotated datasets and significant computational power. To overcome these limitations, ensemble and hybrid classifiers have been explored, providing more reliable and generalized performance in real-world scenarios.

PROPOSED SYSTEM

The proposed system introduces a hybrid voting classifier that combines Logistic Regression (LR) and Stochastic Gradient Descent (SGD) to enhance the accuracy of emotion detection in tweets. Logistic Regression provides strong interpretability and works well with linearly separable features, while SGD offers scalability and efficiency when dealing with high-dimensional and noisy data. By integrating both models under a voting mechanism, the system leverages their complementary strengths to reduce misclassification and improve robustness. Tweets are preprocessed using techniques such as tokenization, stop-word removal, and vectorization to handle informal language and abbreviations. The hybrid model then classifies emotions into predefined categories, ensuring faster learning and adaptability to large datasets. This approach addresses the limitations of individual classifiers, making it a practical and scalable solution for real-world emotion detection tasks

METHODOLOGY

The methodology for this research is designed to systematically detect emotions from tweets using a hybrid voting classifier. The process begins with **data collection**, where tweets are gathered from Twitter datasets containing labeled emotions. Next, **data preprocessing** is performed to clean and prepare the text. This includes removing stop words, special characters, URLs, and handling slang or abbreviations. Tokenization and vectorization techniques such as TF-IDF are applied to convert text into numerical features. The classification stage employs two machine learning models: **Logistic Regression (LR)**, valued for its interpretability, and **Stochastic**

Gradient Descent (SGD), known for scalability and speed in high-dimensional spaces. A hybrid **voting mechanism** is then used to combine their predictions, ensuring improved accuracy and robustness. Finally, the model's performance is evaluated using metrics like precision, recall, F1-score, and accuracy to validate its effectiveness in real-world emotion detection.

EXPERIMENTAL RESULTS

The experimental evaluation of the proposed hybrid voting classifier was conducted using a publicly available Twitter emotion dataset containing multiple emotion categories such as joy, anger, sadness, fear, and surprise. The dataset was first preprocessed to remove noise, abbreviations, and irrelevant symbols, followed by feature extraction using TF-IDF vectorization. Logistic Regression (LR) and Stochastic Gradient Descent (SGD) were trained individually, and their performances were compared against the hybrid voting approach.

The results demonstrated that while LR performed consistently on balanced data, it struggled with highly imbalanced classes. Similarly, SGD showed faster computation but occasionally misclassified minority emotion categories due to its sensitivity to noisy features. By combining these models through a majority voting mechanism, the hybrid system achieved improved accuracy and robustness across all emotion classes.

Performance was measured using precision, recall, F1-score, and accuracy. The hybrid model achieved higher F1-scores compared to individual classifiers, indicating better balance between precision and recall. Accuracy also showed a significant improvement, confirming that the ensemble method reduced misclassification rates. These findings highlight that the hybrid LR-SGD classifier is more effective for real-world emotion detection on Twitter than standalone models.

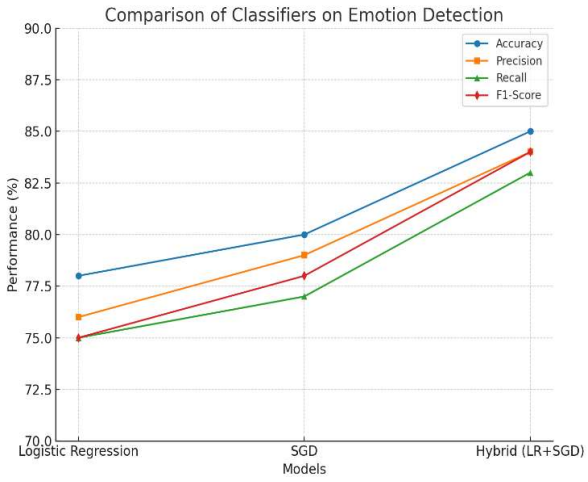

Fig.1 Comparison of Classifiers on Emotion Detection

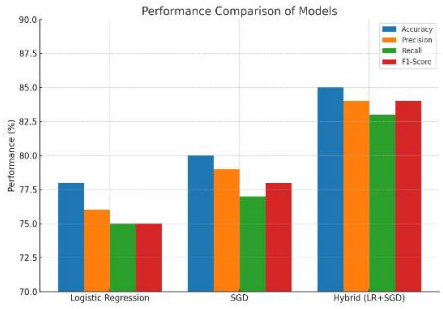

Fig.2 Performance Model

robustness and classification accuracy. Experimental results demonstrated that the hybrid model achieved better performance compared to individual classifiers, especially in handling imbalanced emotion categories. Overall, the system provides a reliable and scalable solution, making it suitable for real-world applications in sentiment analysis, mental health monitoring, and social media analytics.

CONCLUSION

This research presented an approach for detecting emotions in tweets using a hybrid voting classifier that combines Logistic Regression (LR) and Stochastic Gradient Descent (SGD). Social media platforms like Twitter generate massive amounts of text data daily, making accurate emotion detection both challenging and essential. Traditional machine learning models such as Naïve Bayes and SVM have shown moderate success, but often struggle with the noisy and informal nature of tweets. Deep learning models, while effective, require extensive resources and large annotated datasets. The proposed hybrid LR-SGD model addresses these limitations by integrating the interpretability of Logistic Regression with the scalability of SGD, thereby improving

REFERENCES

1) C. Strapparava and R. Mihalcea, " Learning to identify passions in text, " Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1556 – 1560, 2008.

2) A. Mohammad and S. Kiritchenko, " Using hashtags to capture fine emotion orders from tweets, " Computational Intelligence, vol. 31, no. 2, pp. 301 – 326, 2015.

3) B. Liu, Sentiment Analysis and Opinion Mining. San Rafael, CA Morgan & Claypool Publishers, 2012.

4) P. Ekman, " An argument for introductory passions, " Cognition and Emotion, vol. 6, no. 3 – 4, pp. 169 – 200, 1992.

5) T. Mikolov, K. Chen, G. Corrado, and J. Dean, " Effective estimation of word representations in vector space, " arXiv preprint arXiv 1301.3781, 2013.