Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

Sentiment Insights Unveiling Text Emotions Through Machine Learning

Harshad Pratap Sawant
Department of Computer Science
Dr.D.Y Patil Arts,Commerce & Science college pimpri
Pune,india

Sakshi Pandharinath Pokharkar
Department of Computer Science
Dr.D.Y Patil Arts,Commerce & Science college pimpri
Pune,india

Abstract - Our research introduces an innovative and efficient method for incorporating discourse relations into sentiment analysis of tweets. This approach is tailored for web-based applications handling unstructured, noisy text, like tweets, where traditional linguistic tools struggle with the inherent noise. Current methods for micro-blogs, especially on platforms such as Twitter, often rely on simplistic bag-of-words models that overlook the significance of discourse particles like 'but,' 'since,' 'although,' and others. However, our study highlights the importance of integrating discourse relations and conditionals into these models, significantly improving sentiment classification accuracy.

Additionally, we explore how semantic elements like modals and negations influence discourse relations, consequently impacting the sentiment of a sentence. Our approach efficiently identifies discourse relations and associated rules through minimal processing via list lookup. By linguistically describing these relations, we establish conditions in rules and features for Support Vector Machines (SVM). Our findings demonstrate that this discourse-based bagofwords model excels, particularly in noisy mediums like Twitter, surpassing existing Twitter-based applications.

Moreover, our method proves advantageous in structured reviews, showcasing higher accuracy compared to state-of-the-art systems, especially within the travel review domain. In summary, our system not only competes well with existing systems but also presents the benefit of being less resource-intensive, signifying a promising advancement in sentiment analysis.

Keywords - sentiment Analysis,Emotion analysis, Naive Bayes ,Support Vector Machines, Preprocessing, Tokenization, Stemming ,Lemmatization , Twitter.

I. INTRODUCTION

This project idea stems from a deep interest in sentiment analysis within the machine learning domain, a pivotal subset of natural language processing (NLP). Given the growing significance of NLP, this project aims to explore and interpret human language, delving into the captivating landscape of sentiment analysis. Its origins lie in a previous venture involving the categorization of brief music clips based on their emotional attributes. Expanding on this groundwork, the project endeavors to extend the same fundamental concept to a new realm by studying the sentiments conveyed in tweets. The main goal is to identify whether a tweet predominantly expresses positive or negative sentiment.

This initiative highlights the flexibility and relevance of sentiment analysis methodologies across various data sources, offering an opportunity to uncover valuable insights within social media dialogues. By adapting and employing sentiment analysis techniques specifically to tweets, this project seeks to contribute to a deeper comprehension of sentiments expressed in the digital sphere. It holds the potential to impact areas such as social media marketing, brand perception analysis, and public opinion monitoring, offering valuable implications for these fields.

The drive behind delving into a sentiment analysis project arises from the escalating importance of understanding human emotions and viewpoints conveyed through textbased interactions. In today's digital era, an overwhelming amount of unstructured data is generated daily, holding a wealth of undiscovered insights. These insights are not only valuable for businesses and organizations but also for society as a whole. Sentiment analysis, a crucial component of natural language processing (NLP), emerges as a powerful tool in this dynamic landscape. It acts as a means to decode public sentiments, evaluate customer feedback, and monitor online reputation.

II. LITERATURE REVIEW

Almatrafi, Parack, and Chavan (2014) extended sentiment analysis to a geospatial dimension through their study “Application of Location-Based Sentiment Analysis Using Twitter for Identifying Trends Towards Indian General Elections 2014.” Their work employed NLP and machine learning to extract location-specific sentiments from tweets related to the Indian elections. The study showcased the potential of integrating geolocation data to identify regional sentiment variations, enabling policymakers and analysts to interpret public opinion patterns more effectively across different regions.

Advancing beyond traditional machine learning methods, Liu et al. (2018) presented a deep learning-based model in “Sentiment Analysis on Social Media Using Convolutional and Recurrent Neural Networks.” Their approach integrated CNNs for feature extraction and RNNs for sequential dependency modelling, achieving superior performance across multiple sentiment analysis benchmarks. This hybrid architecture captured both semantic and contextual features, reflecting a major evolution toward more intelligent and context-aware sentiment interpretation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

Collectively, these studies trace the progression of sentiment analysis from early machine learning approaches to sophisticated deep learning frameworks. The research evolution demonstrates a continuous movement toward greater contextual awareness, real-time processing, and application in diverse domains such as politics, marketing, and social behaviour analysis. As social media continues to shape global communication, sentiment analysis remains an essential tool for understanding collective human emotions, trends, and decision-making processes.

III. ALGORITHMS

A. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is categorized as a supervised machine learning algorithm commonly utilized for classification and regression tasks. Its effectiveness in managing high-dimensional spaces renders it applicable across diverse domains, including image recognition, text classification, and bioinformatics. In the context of binary classification, SVM's primary objective is to identify a hyperplane that effectively separates data into two classes, emphasizing the importance of maximizing the margin between them. This margin, denoting the distance between the hyperplane and the nearest data point from either class, is a key focus for optimal performance.

The distinctiveness of SVM lies in its ability to handle non-linear decision boundaries through the application of the kernel trick. This technique involves the mapping of input features into a higher-dimensional space, facilitating the identification of a hyperplane within this transformed space. Common kernel functions encompass linear, polynomial, radial basis function (RBF), and sigmoid. The pivotal role of support vectors, defined as the data points closest to the decision boundary, cannot be overstated, influencing the position of the hyperplane and, consequently, the overall performance of SVM.

B. Random Forest Algorithm

Random Forest is a powerful and flexible machine learning ensemble method designed for tasks such as classification and regression. Falling under the category of ensemble learning, Random Forest builds multiple decision trees during training and integrates their predictions for more robust and accurate results.

An important characteristic of Random Forest is the introduction of randomness in the training process. This involves selecting random subsets of features for each decision tree and utilizing bagging, a technique that creates diverse trees through bootstrap sampling with replacement. This randomness serves to prevent overfitting and enhances the model's resilience to noise in the data.

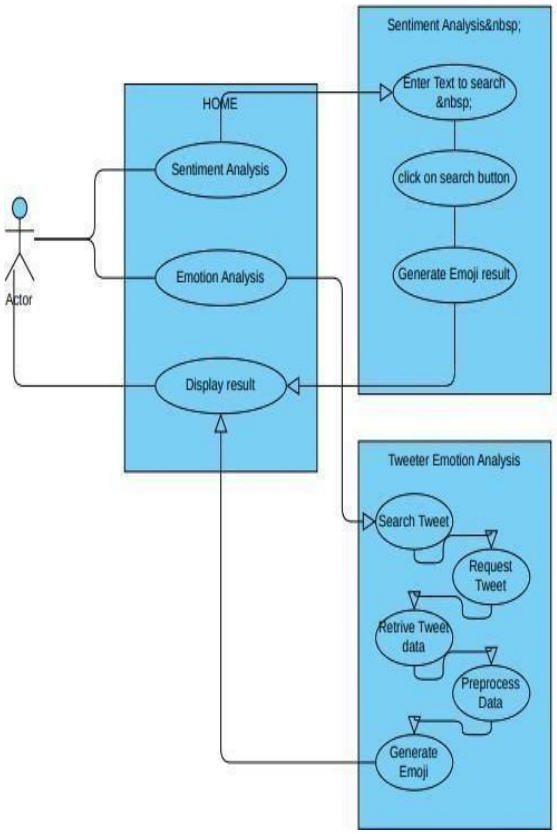
The construction of each decision tree within the Random Forest involves recursively splitting nodes based on the most informative features at each step. This splitting process continues until a specified stopping criterion, such as a maximum depth or a minimum number of samples per leaf, is met. In classification tasks, Random Forest combines the predictions of individual trees through a voting mechanism, while in regression tasks, predictions from each tree are averaged to obtain the final output.

A distinctive feature of Random Forest is its ability to provide a measure of feature importance, indicating the contribution of each feature to the model's predictive performance. Additionally, Random Forest exhibits robustness and generalization capabilities, being less susceptible to overfitting compared to individual decision trees. Its versatility allows it to handle both categorical and numerical features with minimal hyperparameter tuning, making it widely utilized in various domains, including finance, healthcare, and remote sensing.

IV. DIAGRAMS

A. Use Case Diagram

A use case diagram is a visual representation in Unified Modelling Language (UML) that depicts the interactions between actors (users or external systems) and a system, highlighting various ways users can interact with the system and the corresponding functionalities provided by the system in response to those interactions

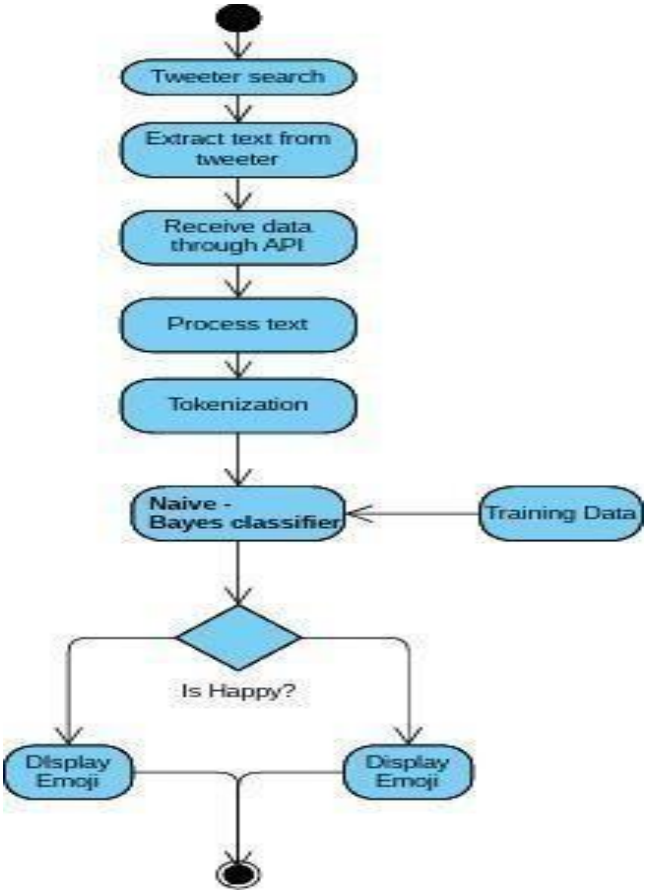


Figures A. Use Case Diagram

A. Activity Diagram

An activity diagram is a visual representation within the Unified Modelling Language (UML) that illustrates the dynamic flow of activities and actions in a system or business process. It employs symbols like rounded rectangles for activities, arrows for transitions, diamonds for decision nodes, and bars for forks and joins. Activity diagrams are instrumental in modelling the sequential and parallel aspects of processes, offering insights into system dynamics and assisting in the analysis and design of workflows and software systems.

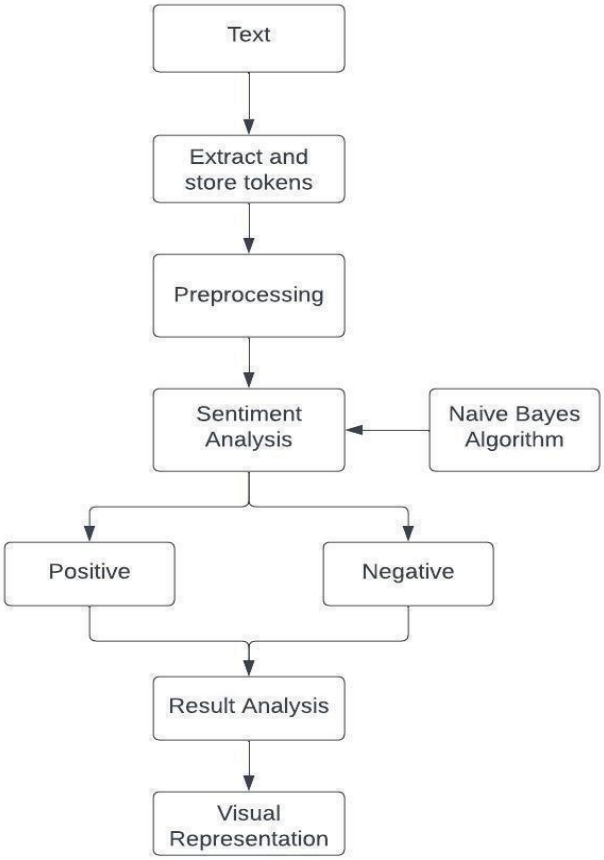
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026



Figures B. Activity Diagram

B. Architecture Diagram

An architecture diagram serves as a visual representation, outlining the structural organization of a system. Widely employed in software development and engineering, this diagram provides a high-level overview of the system's components, illustrating their relationships and interactions. It typically showcases key building blocks such as servers, databases, and applications, highlighting connections, interfaces, and dependencies between them.



Figures C. Architecture Diagram

CONCLUSION

Sentiment analysis is technique employed to analyze and discern the opinions, attitudes, and emotional states expressed by individuals, ranging from positive to negative on a spectrum. Commonly, linguistic features such as parts of speech are utilized to extract sentiment from text, with adjectives playing a particularly significant role in this process. However, challenges may arise when words encompass both adjectives and adverbs in their usage, potentially complicating the accurate identification of sentiment and opinions.

In the context of sentiment analysis on tweets, the proposed system commences by extracting Twitter posts based on user input. Furthermore, the system calculates the frequency of each term within the tweet. To generate results, the system adopts a supervised machine learning approach. This methodology contributes to achieving precise sentiment analysis outcomes.

REFERENCES

[1] C.D. Manning, P. Raghavan and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, pp. 234-265, 2008

[2] T. Wu, C. Lin and R. Weng, "Probality estimates for multi-class classification by pairwise coupling", Proc. JMLR-5, pp. 975-1005, 2004

[3] "Support Vector Machines" [Online], http://scikit-learn.org/stable/modules/svm.html#svm-classification, Accessed Jan 2016

[4] P. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques", Proc. ACL-02 conference on Empirical methods in natural language processing, vol.10, pp. 79-86, 2002

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

[5] P. Pang and L. Lee, "Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval", vol. 2(1-2), pp.1- 135, 2008

[6] 7. E. Loper and S. Bird, "NLTK: the Natural Language Toolkit", Proc. ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics ,vol. 1,pp. 63-70, 2002

[7] 8. H. Wang, D. Can, F. Bar and S. Narayana, "A system for real-time Twitter sentiment analysis of 2012 U.S.presidential election cycle", Proc. ACL 2012 System Demostration, pp. 115-120, 2012