SENTIMENT ANALYSIS OF TWITTER DATA USING NLP MODELS

Dr. A. Rajesh[§], B.Abhinash Goud[*], G.Ajay[*], B.Anil Kumar[*],
[§]Associate Professor, Department of Computer Science & Engineering
Guru Nanak Institute of Technology
Hyderabad, Telangana, India.
[*]Student, Department of Computer Science & Engineering
Guru Nanak Institute of Technology
Corresponding Author Email: adepurajeshadepu@gmail.com

Abstract— Sentiment analysis of Twitter data has emerged as a powerful approach for understanding public opinion, trends, and user behavior in real time. This study focuses on leveraging Natural Language Processing (NLP) models to classify tweets into sentiment categories such as positive, negative, and neutral. Due to the informal and noisy nature of Twitter text—including abbreviations, emojis, hashtags, and slang—advanced preprocessing techniques such as tokenization, stop-word removal, stemming, and normalization are applied to improve data quality.

Various machine learning and deep learning models, including Logistic Regression, Na¨ıve Bayes, Support Vector Machines, and transformer-based architectures, are employed to analyze sentiment patterns effectively.

Index Terms—Sentiment Analysis, NLP, Twitter Data, Machine Learning, TF-IDF

Keywords—component; formatting; style; styling; insert (key words)

I. INTRODUCTION

Sentiment analysis has become an essential technique in today's data-driven world, where understanding public opinion plays a crucial role in decision-making across various domains. With the rapid growth of social media platforms such as Twitter, vast amounts of user-generated content are produced every second, reflecting opinions, emotions, and reactions toward events, products, and services.

An automated sentiment analysis system provides a systematic and technology-driven approach to classifying textual data into sentiment categories such as positive, negative, or neutral. The system leverages techniques from Natural Language Processing (NLP) to preprocess, analyze, and interpret tweets.

II. FEASIBILITY STUDY

The feasibility study helps determine whether the proposed sentiment analysis system using Natural Language Processing is practical, cost-effective, and efficient to implement.

A. Types of Feasibility Study

- Technical Feasibility
- Economic Feasibility
- Operational Feasibility
- Legal Feasibility
- Time Feasibility

B. NLP-Based Text Similarity Detection

Unlike purely rule-based methods, this project adopts a machine learning–driven Natural Language Processing approach for analyzing sentiment in textual data.

C. NLP vs Deep Learning

TABLE I
NLP VS DEEP LEARNING

Feature	NLP-Based	Deep Learning
Implementation	Simple	Complex
Resources	Low CPU	High GPU
Performance	Effective	High Accuracy

.

III. LITERATURE SURVEY

A. Exploring the Impact of Slang Usage Among Students onWhatsApp

This research delves into the dynamic world of informal language usage among students on the popular messaging platform, WhatsApp.

B. Exploring the Frontiers of Deep Learning and NLP

Traditional manual grading systems are replaced with automated machine learning and NLP approaches.

IV. SCOPE

1) Efficient and Automated Text Analysis 2) Real-Time Data Processing
3) High Accuracy in Sentiment Detection
4) Cost-Effective Solution
5) Enhanced Decision-Making
6) Scalability and Large Dataset Handling
7) Trend and Opinion Analysis
8) Social Media Monitoring

V. PROPOSED SYSTEM

The proposed sentiment analysis system aims to overcome the limitations of traditional approaches by integrating Natural Language Processing and machine learning techniques.

A. Advantages of Proposed System

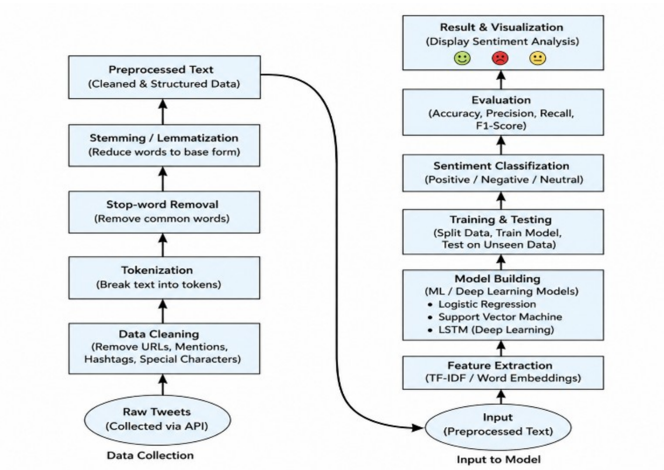
- High Accuracy
- Efficiency and Speed
- Scalability
- Contextual Understanding
- Real-Time Analysis
- User-Friendly Interface

VI. METHODOLOGY

A. Modules

1) Data Collection
2) Text Preprocessing
3) Feature Extraction
4) Sentiment Classification
5) Result Evaluation

The system architecture illustrates the workflow of the sentiment analysis system using NLP for processing textual data.



Sentiment Analysis of Twitter Data using NLP Models

A. Natural Language Processing

Preprocessing includes tokenization, stopword removal, punctuation removal, and normalization.

B. TF-IDF

TF-IDF converts textual data into numerical vectors for machine learning models.

C. Cosine Similarity

Cosine similarity measures similarity between feature vectors.

VII. DESIGN AND DEVELOPMENT

A. System Architecture

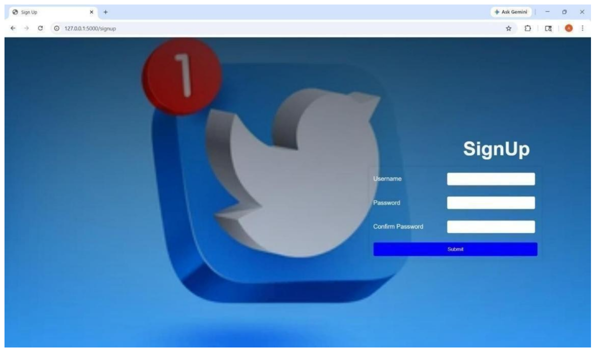
B. UML Diagrams

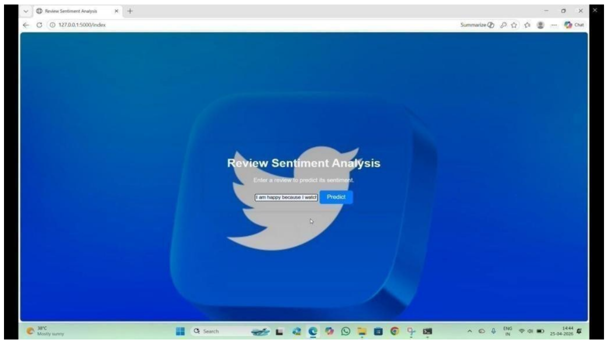
- Use Case Diagram
- Class Diagram
- Object Diagram
- Component Diagram
- Deployment Diagram
- Sequence Diagram

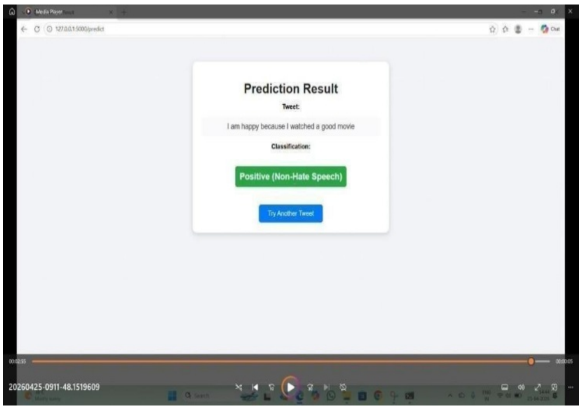
Activity Diagram

IX .RESULTS AND DISCUSSION

The proposed system successfully classifies tweets into positive, negative, and neutral categories using machine learning and NLP techniques.







 ISSN: 3107-6513

X .CONCLUSION

The sentiment analysis system developed in this project presents an effective and automated solution for analyzing public opinion from Twitter data.

By integrating Natural Language Processing techniques with TF-IDF feature extraction and machine learning algorithms, the system successfully classifies textual data into positive, negative, and neutral sentiments.

XI. FUTURE ENHANCEMENTS

1) Multilingual Sentiment Analysis
2) Real-Time Data Streaming Integration
3) Transformer-Based Models

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Dr. B. Santhosh Kumar, Head of the Department of Computer Science and Engineering, Guru Nanak Institute of Technology, for his valuable support and encouragement throughout the project.

We extend our heartfelt thanks to our guide Dr. A. Rajesh, Associate Professor, Department of Computer Science and Engineering, for his continuous guidance, motivation, and technical support during every stage of the project development.

We also thank all the faculty members, friends, and our parents for their constant encouragement and support in successfully completing this project.

REFERENCES

[1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.

[2] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

[3] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. Stanford University Technical Report.

[4] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining.

[5] *Proceedings of LREC*, 1320–1326.

[6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*.

[7] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *EMNLP*.

[8] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). *Recursive deep models for semantic compositionality over a sentiment treebank*. Proceedings of EMNLP.

[9] Tang, D., Qin, B., & Liu, T. (2015). *Document modeling with gated recurrent neural network for sentiment classification*. Proceedings of EMNLP.

[10] Zhang, X., Zhao, J., & LeCun, Y. (2015). *Character-level convolutional networks for text classification*. Advances in Neural Information Processing Systems (NeurIPS).

[11] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). *Learning word vectors for sentiment analysis*. Proceedings of ACL.

[12] Johnson, R., & Zhang, T. (2017). *Deep pyramid convolutional neural networks for text categorization*. Proceedings of ACL.

[13] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ICLR*.

[14] Kim, Y. (2014). Convolutional neural networks for sentence classification. *EMNLP*.

[15] Adepu Rajesh and Tryambak Hiwarkar, Sentiment Analysis Using Ensemble of Deep Learning Models Proceedings of International Conference on Computational Intelligence ICCI 2023 Springer page: 471-484