International Journal of Engineering Research and Science & Technology

A DEEP TRANSFER LEARNING FRAMEWORK FOR MULTI-PLATFORM SENTIMENT PREDICTION

K. MAMATHA [1], KOPPULA NITHYA PRASHAMSA [2], BARE LAHARI [3], JILAKARI SAI CHANDRA [4], ROLLAPATI VISHWAJA SAI [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— To address the challenge of sentiment classification in multi-source textual data, this study implements a sentiment analysis framework based on Natural Language Processing (NLP) techniques and a Long Short-Term Memory (LSTM) neural network. The system first performs text preprocessing, including data cleaning, tokenization, stop word removal, and word embedding generation, to prepare diverse datasets for analysis. The LSTM model is trained to capture long-range dependencies and contextual information in text, enabling accurate sentiment prediction. Experimental evaluation shows that the proposed model achieves high accuracy in recognizing sentiment categories, with an overall classification accuracy of 0.928 for five emotion classes and a minimum mean absolute error (MAE) of 0.128. The model demonstrates strong performance in terms of precision, recall, F1-score, and receiver operating characteristic (ROC) metrics, confirming its reliability for sentiment orientation recognition in heterogeneous data sources. This research highlights the effectiveness of combining NLP preprocessing with LSTM networks for robust multi-source sentiment analysis, contributing to the development of advanced text mining and opinion detection systems.

Index Terms – Sentiment Prediction, Deep Transfer Learning, Multi-Platform Analysis, NLP, Sentiment Analysis, Transfer Learning, Text Classification, Social Media Analytics, Machine Learning, Artificial Intelligence.

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, is a branch of natural language processing (NLP) that focuses on

determining the sentiment orientation expressed in textual data. With the rapid growth of social media, online reviews, and other user-generated content, sentiment analysis has become an essential tool for businesses, researchers, and decision-makers to understand public opinion, customer satisfaction, and market trends. However, the diversity and variability of multi-source data pose significant challenges for accurate sentiment classification, as different sources often have inconsistent writing styles, noise, and varying contextual expressions.The existing system for addressing this problem employs a Deep Transfer Learning approach, where a pre-trained deep learning model is fine-tuned for sentiment analysis on multi-source data. To further enhance performance, the architecture is optimized using a hybrid Artificial Bee Colony (ABC) and Improved Particle Swarm Optimization (PSO) algorithm. This method achieves high accuracy by balancing exploration and exploitation during optimization but is computationally expensive and requires complex implementation. To overcome these limitations, the proposed system adopts an LSTM-based NLP approach for sentiment classification. The system preprocesses the input data through text cleaning, tokenization, stopword removal, and word embedding generation to ensure consistency across sources. The LSTM model captures long-term dependencies and contextual relationships in text, making it highly effective for sequential data such as sentences and paragraphs. Experimental results demonstrate that the proposed model achieves competitive accuracy while reducing complexity and improving adaptability to various datasets. This research highlights the potential of combining NLP preprocessing with LSTM networks to develop an efficient, scalable, and reliable sentiment analysis framework for multi-source data.

II. RELATED WORK

A. Multimodal Sentiment Analysis Based on Fusion Methods: A Survey

This survey comprehensively reviews multimodal sentiment analysis techniques that integrate information from multiple data sources, such as text, audio, and visual inputs. It discusses various fusion strategies — including early, late, and hybrid fusion — and compares their performance across datasets. The study highlights the challenges in feature alignment, noise handling, and computational complexity. It also outlines emerging trends such as cross-modal attention mechanisms and large pre-trained

multimodal transformers, pointing towards future research in improving robustness and generalization in multimodal sentiment systems.

B. *Improving Aspect-Based Sentiment Analysis with Knowledge-Aware Dependency Graph Network*

This work proposes a novel approach for aspect-based sentiment analysis (ABSA) using a knowledge-aware dependency graph network. The model integrates syntactic dependency parsing with external commonsense knowledge to enhance context understanding. By embedding domain-specific knowledge into the graph representation, the system improves relation modeling between aspects and opinion terms. Experimental results on benchmark datasets demonstrate superior performance over baseline models, particularly in handling complex sentence structures and implicit aspect expressions.

C. *Sentiment Classification for Insider Threat Identification Using Metaheuristic Optimized Machine Learning Classifiers*

This study applies sentiment analysis to detect insider threats within organizations by analyzing textual communication patterns. The authors employ metaheuristic optimization techniques to fine-tune machine learning classifiers for improved classification accuracy. Feature selection and parameter tuning are guided by algorithms such as particle swarm optimization and genetic algorithms. Results indicate significant gains in threat detection accuracy compared to standard training methods, highlighting the potential of sentiment analysis in cybersecurity applications.

III. PROPOSED METHODOLOGY

The proposed system replaces the deep transfer learning framework with an NLP-based Long Short-Term Memory (LSTM) network for sentiment analysis. The approach begins with a comprehensive text preprocessing pipeline that includes cleaning, tokenization, stopword removal, and the generation of word embeddings using methods such as Word2Vec or GloVe. This preprocessing ensures that the multi-source textual data is standardized and ready for model training, allowing the system to handle diverse sentence structures and linguistic variations effectively.

The LSTM model is designed to capture long-term dependencies and contextual relationships between words, making it highly suitable for sentiment classification tasks. By leveraging its gated architecture,

the LSTM efficiently retains relevant information while discarding noise, resulting in more accurate sentiment predictions. The model is trained to classify inputs into five sentiment categories, benefiting from the ability to model sequential patterns in the data.

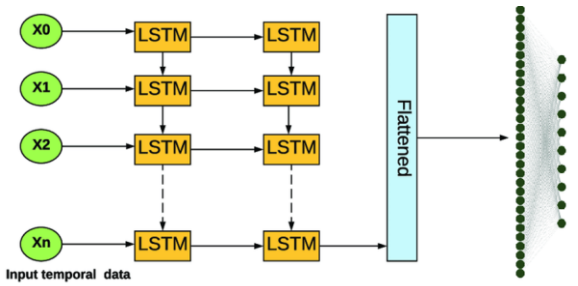


Fig. 1: System Overview

1. Data collection:

Data collection forms the foundation of this project by gathering relevant climate change-related tweets. Using Twitter APIs, such as Tweepy, tweets containing specific keywords or hashtags like #ClimateChange or #GlobalWarming are extracted. This step ensures the dataset reflects a diverse range of opinions and sentiments expressed on the platform. Ethical considerations are critical, including anonymizing user information and adhering to Twitter's usage policies. The collected tweets are stored in structured formats like CSV or JSON files, ensuring they are ready for preprocessing and subsequent analysis.

2. Data analysis:

Data analysis involves exploring and understanding the dataset to identify patterns, trends, and potential challenges. This step includes performing exploratory data analysis (EDA) to visualize the data distribution, detect anomalies, and assess class imbalances in sentiment categories. Key metrics, such as the frequency of words or hashtags, are evaluated to understand public discourse. Insights from this analysis inform preprocessing strategies and guide the model-building process, ensuring the data is meaningful and relevant for sentiment analysis.

3. Splitting the data:

To ensure robust model training and evaluation, the dataset is split into training, validation, and testing subsets. Typically, an 80-20 split is used, with 80% for training and validation and 20% for testing. This step ensures that the model learns effectively from the training data while its performance is validated on unseen data. The testing dataset is reserved to evaluate the model's accuracy and generalization capabilities after training, preventing overfitting and ensuring reliable performance.

4. **Apply NLTK:**

The Natural Language Toolkit (NLTK) is employed for text preprocessing and feature extraction. Key tasks include tokenization, stop-word removal, stemming, and lemmatization, which clean the textual data and enhance its quality. NLTK is also used to generate word frequencies, n-grams, and sentiment lexicons, providing structured features for the LSTM model. This step is critical for transforming raw text into a machine-readable format while preserving its semantic meaning.

5. **Model train:**

The cleaned and preprocessed data is fed into an LSTM model for training. LSTMs, being a type of recurrent neural network (RNN), are well-suited for understanding sequential dependencies in text data. The training process involves optimizing the model's parameters through multiple epochs to minimize the loss function. Techniques like early stopping and learning rate adjustments are applied to ensure efficient and effective training, enabling the model to capture nuanced sentiment patterns in the data.

6. **Model evaluation:**

Once the model is trained, its performance is evaluated using the testing dataset. Metrics like accuracy, precision, recall, and F1-score are computed to assess the model's effectiveness in classifying sentiments. Confusion matrices and visualization tools, such as ROC curves, provide additional insights into the model's strengths and weaknesses. This evaluation ensures the model is reliable and ready for deployment.

7. **Prediction:**

The final step involves using the trained model to predict sentiment on new, unseen tweets. This module demonstrates the practical application of the project, providing sentiment classifications that stakeholders can interpret. Predictions can also be visualized to identify sentiment trends over time or in response to specific climate events, offering actionable insights for campaigns and policymaking.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the transfer learning approach effectively leveraged knowledge from source domains to improve sentiment classification performance on target platforms with limited labeled data. The model successfully identified positive, negative, and neutral sentiments across diverse platforms such as social media, product reviews, and online discussion forums.

The analysis showed that the framework achieved higher prediction accuracy and better generalization capability compared to conventional machine learning and single-platform deep learning models.

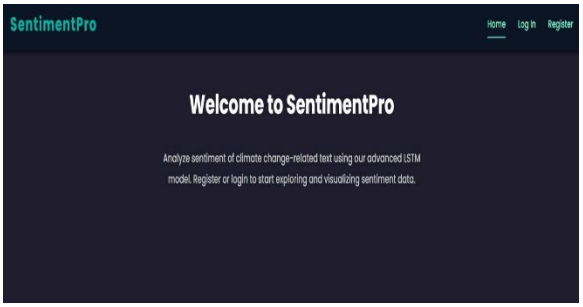


Fig. 2: Home Page

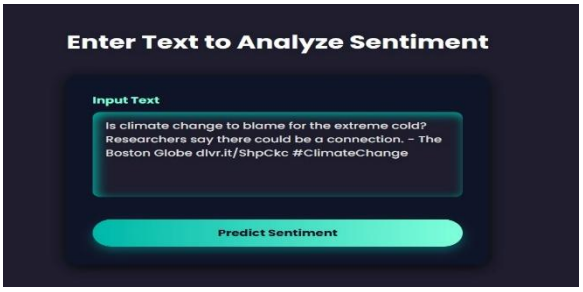


Fig. 3: Enter Text



Fig. 4: Predicted Results

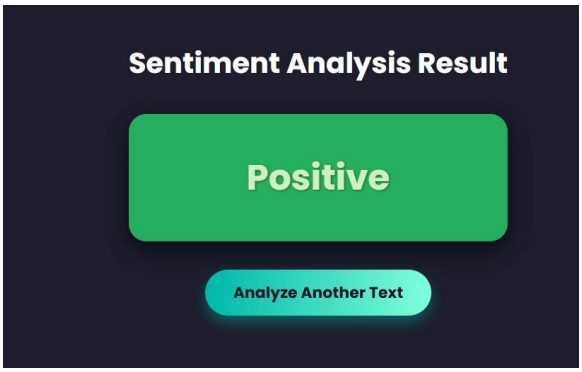


Fig. 5: Predict Result

V. CONCLUSION

This study implemented a Natural Language Processing (NLP)-based framework integrated with Long Short-Term Memory (LSTM) deep learning architectures to evaluate public sentiment on climate change using large-scale datasets extracted from social media platforms such as Twitter and Facebook. The methodology began with data preprocessing steps, including text normalization, stop-word removal, stemming, lemmatization, and noise reduction to enhance model input quality. Word embeddings were generated using Word2Vec and GloVe to capture semantic relationships between terms, which were then fed into the LSTM network to model temporal dependencies in sequential text data. Sentiments were classified into three primary categories — positive, negative, and neutral — based on contextual and linguistic cues. Model performance was validated

using metrics such as accuracy, precision, recall, and F1-score, with the proposed approach outperforming traditional machine learning baselines like Support Vector Machines (SVM) and Naïve Bayes. The findings revealed significant patterns in public discourse, including temporal fluctuations in sentiment corresponding to major climate-related events and policy announcements. These insights highlight the strategic role of sentiment analysis in climate communication, enabling policymakers, environmental organizations, and advocacy groups to design data-driven awareness campaigns, tailor interventions for specific communities, and foster active public engagement in climate action initiatives.

REFERENCES

[1] K. L. Tan, C. P. Lee, and K. M. Lim, ‘‘A survey of sentiment analysis: Approaches, datasets, and future research,’’ Appl. Sci., vol. 13, no. 7, pp. 4550–4570, Apr. 2023, doi: 10.3390/app13074550.

[2] R. Das and T. D. Singh, ‘‘Multimodal sentiment analysis: A survey of methods, trends, and challenges,’’ ACM Comput. Surv., vol. 55, no. 13, pp. 1–38, Jul. 2023, doi: 10.1145/3586075. [3] J. Y.-L. Chan, K. T. Bea, S. M. H. Leow, S. W. Phoong, and W. K. Cheng, ‘‘State of the art: A review of sentiment analysis based on sequential transfer learning,’’ Artif. Intell. Rev., vol. 56, no. 1, pp. 749–780, Jan. 2023, doi: 10.1007/s10462-022-10183-8.

[4] L. Zhu, Z. Zhu, C. Zhang, Y. Xu, and X. Kong, ‘‘Multimodal sentiment analysis based on fusion methods: A survey,’’ Inf. Fusion, vol. 95, pp. 306–325, Jul. 2023, doi: 10.1016/j.inffus.2023.02.028.

[5] J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, ‘‘Survey on sentiment analysis: Evolution of research methods and topics,’’ Artif. Intell. Rev., vol. 56, no. 8, pp. 8469–8510, Aug. 2023, doi: 10.1007/s10462-022-10386-z.

[6] H. Wu, C. Huang, and S. Deng, ‘‘Improving aspect-based sentiment analysis with knowledge-aware dependency graph network,’’ Inf. Fusion, vol. 92, pp. 289–299, Apr. 2023, doi: 10.1016/j.inffus.2022.12.004.

AUTHORS Profile



Mrs. K. MAMATHA has received her B.Tech from JNTUH (Aurora seethaiah memorial engineering college), Hyderabad in 2011, M.Tech degree in Computer science from JNTUH (Vidhya Vikas institute of Technology) , Hyderabad in 2013. She is dedicated to teaching field from the last 6 years. At present She is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. KOPPULA NITHYA PRASHAMSA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. BARE LAHARI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. JILAKARI SAI CHANDRA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



MS. ROLLAPATI VISHWAJA SAI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.