International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

CLIMATE CHANGE SENTIMENT ANALYSIS USING NATURAL LANGUAGE PROCESSING AND LSTM MODEL

NAVEEN REDDY NEMILI [1], S. HEMANTH [2], MD. JASMINE [3], V. NITHIN [4], C. RAJESH [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— Climate change remains one of the most pressing global challenges, and understanding public sentiment surrounding this issue is critical for shaping effective policy and response strategies. Social media platforms, particularly Twitter, have become key venues for individuals to express their opinions, concerns, and reactions to climate change-related topics. To capture and analyze this sentiment, this study employs Natural Language Processing (NLP) techniques combined with Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, to analyze tweets related to climate change. The LSTM model, renowned for its ability to capture long-range dependencies in text data, is utilized to classify sentiment and extract meaningful insights from the discourse. By applying advanced NLP techniques and deep learning, this study aims to provide a comprehensive understanding of public sentiment on climate change, enabling stakeholders, including policymakers and environmental organizations, to better grasp public perceptions and inform strategies to tackle the climate crisis.

Index Terms – Climate change, sentiment analysis, Natural Language Processing (NLP), Long Short-Term Memory (LSTM), Twitter analysis, social media analytics, deep learning, text classification, public opinion mining, environmental informatics, opinion mining, recurrent neural networks (RNN).

I. INTRODUCTION

Climate change is one of the most urgent global issues, profoundly impacting ecosystems, economies, and human well-being. Understanding public perception of this complex problem is vital for creating effective mitigation strategies. Social media, especially platforms like Twitter, serves as a dynamic channel for public discourse, providing a wealth of real-time data on opinions and sentiments about climate change. Leveraging this data, advanced Natural Language Processing (NLP) and machine learning techniques can reveal trends, patterns, and sentiments, offering valuable insights into how the public views climate-related issues. This project employs NLP combined with Long Short-Term Memory (LSTM) networks to analyze tweets about climate change. LSTM models are adept at capturing long-term dependencies in textual data, making them an ideal choice for sentiment classification. By analyzing this discourse, the project seeks to support policymakers and environmental advocates in understanding public concerns, fostering targeted communication, and designing informed interventions to address the climate crisis.

II. RELATED WORK

A. Deep learning-based sentiment analysis on COVID-19 news videos

Coronavirus disease (COVID-19) has adversely affected all walks of human life. The whole world is confronting this deadly virus, and no country in this world remains untouched during this pandemic. There are several online news videos related to COVID-19 that are shared on various online platforms such as YouTube, DailyMotion, and Vimeo. There were several arguments on the genuineness of the contents, people watch them, share them, and most importantly express their views and opinions as comments on those platforms. Analyzing these comments can unearth the patterns hidden in them to study people's responses to videos on COVID-19. This paper proposes a deep learning-based sentiment analysis approach to people's response toward online COVID-19 video news. This work implements different deep learning approaches such as LSTM, Bi-LSTM, CNN, and GRU to classify sentiment from the comments collected from YouTube.

B. A multi-task model for sentiment aided stance detection of climate change tweets

Climate change has become one of the biggest challenges of our time. Social media

platforms such as Twitter play an important role in raising public awareness and spreading knowledge about the dangers of the current climate crisis. With the increasing number of campaigns and communication about climate change through social media, the information could create more awareness and reach the general public and policy makers. However, these Twitter communications lead to polarization of beliefs, opinion-dominated ideologies, and often a split into two communities of climate change deniers and believers. In this paper, we propose a framework that helps identify denier statements on Twitter and thus classifies the stance of the tweet into one of the two attitudes towards climate change (denier/believer). The sentimental aspects of Twitter data on climate change are deeply rooted in general public attitudes toward climate change. Therefore, our work focuses on learning two closely related tasks: Stance Detection and Sentiment Analysis of climate change tweets. We propose a multi-task framework that performs stance detection (primary task) and sentiment analysis (auxiliary task) simultaneously. The proposed model incorporates the feature specific and shared-specific attention frameworks to fuse multiple features and learn the generalized features for both tasks. The experimental results show that the proposed framework increases the performance of the primary task, i.e., stance detection by benefiting from the auxiliary task, i.e., sentiment analysis compared to its uni-modal and single-task variants.

C. Machine learning for disaster management: Insights from past research and future implications

Natural disasters can cause destructive harm and tremendous loss to lives on earth. It destroys residential and industrial areas and negatively affects the water quality in rivers and streams. Disasters cause substantial financial losses and impact the mental state and emotions of the survivors who lose everything. There is a significant need for proper disaster management systems and strategies to cope with the situations and reduce the impact. The advancement in information technology and communications has made the availability of large datasets that are highly useful for researchers. Machine learning technologies have proven invaluable in disaster and crisis management, specifically pre-disaster and post-disaster stages. Recent works discuss machine learning applications in disaster management, growing exponentially. This paper addresses a detailed discussion on

machine learning models for disaster management and its various stages, such as recovering and consolidating information, searching and rescuing, and post-disaster assessment. This paper gives an insight into the multimodal machine learning framework for planning efficient strategies for disaster management.

D. Climate change sentiment analysis using lexicon, machine learning and hybrid approaches

The emissions of greenhouse gases, such as carbon dioxide, into the biosphere have the consequence of warming up the planet, hence the existence of climate change. Sentiment analysis has been a popular subject and there has been a plethora of research conducted in this area in recent decades, typically on social media platforms such as Twitter, due to the proliferation of data generated today during discussions on climate change. However, there is not much research on the performances of different sentiment analysis approaches using lexicon, machine learning and hybrid methods, particularly within this domain-specific sentiment. This study aims to find the most effective sentiment analysis approach for climate change tweets and related domains by performing a comparative evaluation of various sentiment analysis approaches. In this context, seven lexicon-based approaches were used, namely SentiWordNet, TextBlob, VADER, SentiStrength, Hu and Liu, MPQA, and WKWSCI. Meanwhile, three machine learning classifiers were used, namely Support Vector Machine, Naïve Bayes, and Logistic Regression, by using two feature extraction techniques, which were Bag-of-Words and TF–IDF. Next, the hybridization between lexicon-based and machine learning-based approaches was performed. The results indicate that the hybrid method outperformed the other two approaches, with hybrid TextBlob and Logistic Regression achieving an F1-score of 75.3%; thus, this has been chosen as the most effective approach. This study also found that lemmatization improved the accuracy of machine learning and hybrid approaches by 1.6%. Meanwhile, the TF–IDF feature extraction technique was slightly better than BoW by increasing the accuracy of the Logistic Regression classifier by 0.6%. However, TF–IDF and BoW had an identical effect on SVM and NB. Future works will include investigating the suitability of deep learning approaches toward this domain-specific sentiment on social media platforms.

III. PROPOSED METHODOLOGY

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) architecture designed to model and predict sequential data. Unlike traditional RNNs, LSTMs are specifically engineered to avoid the long-term dependency problem, where the network fails to learn from long sequences due to vanishing gradients. LSTMs are widely used in tasks such as time series forecasting, natural language processing (NLP), speech recognition, and any application involving sequences of data. LSTM networks are designed to solve the vanishing gradient problem that conventional RNNs face when learning long sequences. The architecture of an LSTM network includes a special kind of neuron known as a memory cell, which has three important components.

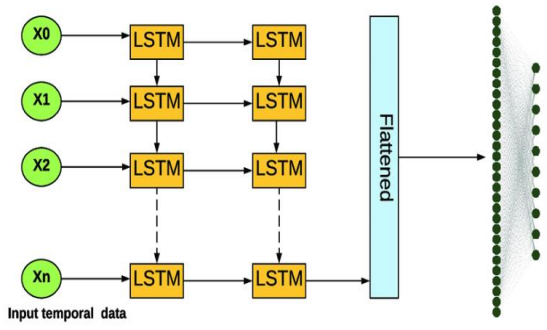


Fig. 1: System Overview

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.

6. 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

The experimental results show that the proposed NLP-LSTM model effectively classifies climate change-related tweets based on sentiment. The LSTM network successfully captures contextual relationships in text, improving sentiment classification accuracy over traditional approaches. The model provides meaningful insights into public opinion by analyzing large volumes of Twitter data. Overall, the framework demonstrates the effectiveness of deep learning for climate change sentiment analysis and supports informed decision-making for policymakers and environmental organizations.

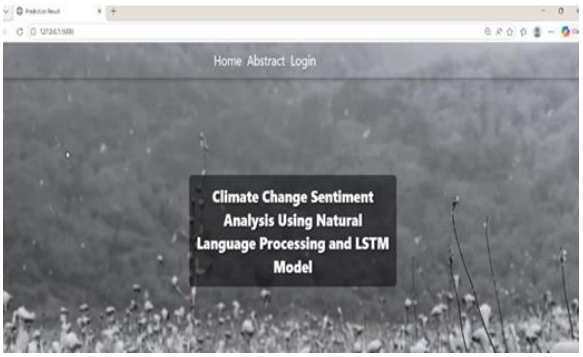


Fig. 2: Home Page

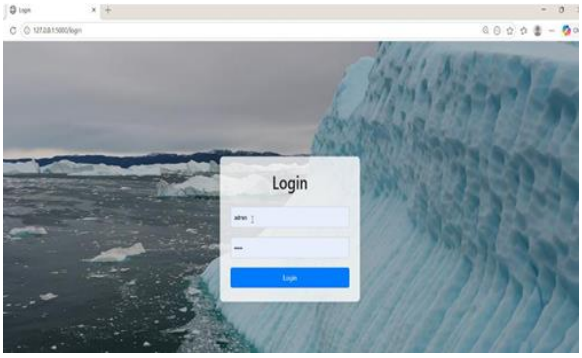


Fig. 3: Login

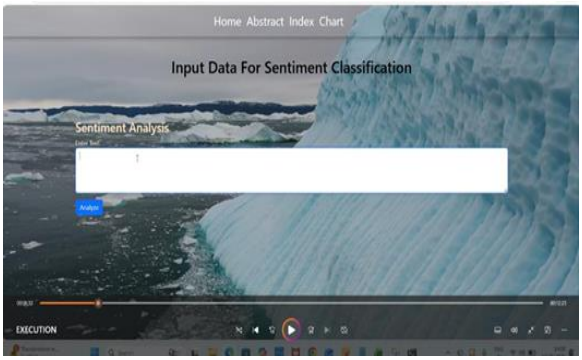


Fig. 4: Text Input Page

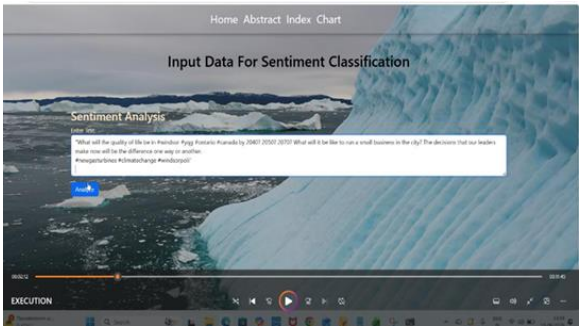


Fig. 5: Sentiment Analysis Input Page

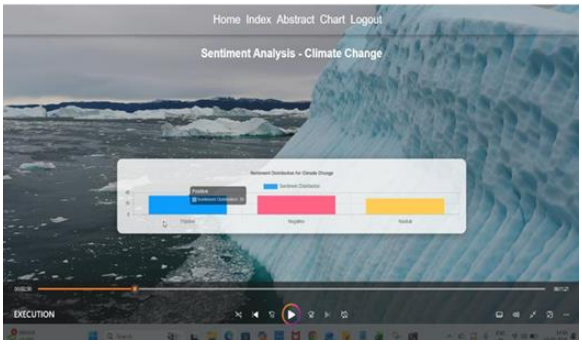


Fig. 6: Sentiment Distribution Chart

V. CONCLUSION

This study successfully employed Natural Language Processing (NLP) and Long Short-Term Memory (LSTM) models to analyze public sentiment on climate change using data from social media platforms. By classifying sentiments into positive, negative, and neutral categories, the project demonstrated the potential of advanced deep learning approaches in understanding public discourse on pressing global issues. The results highlight the importance of leveraging sentiment analysis for informed policy-making, targeted awareness campaigns, and fostering public engagement.

REFERENCES

[1] T. Abdullah and A. Ahmet, ‘‘Deep learning in sentiment analysis: Recent architectures,’’ ACM Comput. Surv., vol. 55, no. 8, pp. 1–37, Aug. 2023.

[2] S. Ali, A. Basit, T. A. Makanda, Inamullah, F. U. Khan, M. Sajid, T. Riaz, H. F. Abbasi, Manzoor, and A. Sohail, ‘‘Improving drought mitigation strategies and disaster risk reduction through MODIS and TRMM-based data in relation to climate change over Pakistan,’’ Environ. Sci. Pollut. Res., vol. 30, no. 14, pp. 40563–40575, Jan. 2023.

[3] O. Alqaryouti, N. Siyam, A. A. Monem, and K. Shaalan, ‘‘Aspect-based sentiment analysis using smart government review data,’’ Appl. Comput. Informat., vol. 20, nos. 1–2, pp. 142–161, Jan. 2024.

[4] A. Alsayat, ‘‘Improving sentiment analysis for social media applications using an ensemble deep learning language model,’’ Arabian J. Sci. Eng., vol. 47, no. 2, pp. 2499–2511, Feb. 2022.

[5] A. Alslaity and R. Orji, ‘‘Machine learning techniques for emotion detection and sentiment analysis: Current state, challenges, and future directions,’’ Behav. Inf. Technol., vol. 43, no. 1, pp. 139–164, Jan. 2024.

[6] V. Anoop, ‘‘Sentiment classification of diabetes-related tweets using transformer-based deep learning approach,’’ in Proc. Int. Conf. Adv. Comput. Data Sci. Cham, Switzerland: Springer, 2023, pp. 203–214.

[7] V. S. Anoop and S. Asharaf, ‘‘Extracting conceptual relationships and inducing concept lattices from unstructured text,’’ J. Intell. Syst., vol. 28, no. 4, pp. 669–681, Sep. 2019.

[8] V. S. Anoop and S. Sreelakshmi, ‘‘Public discourse and sentiment during Mpox outbreak: An analysis using natural language processing,’’ Public Health, vol. 218, pp. 114–120, May 2023.

[9] S. S. Biswas, ‘‘Potential use of chat GPT in global warming,’’ Ann. Biomed. Eng., vol. 51, no. 6, pp. 1126–1127, Jun. 2023.

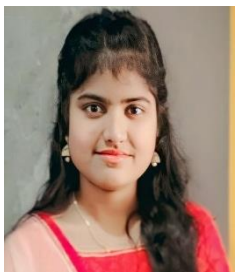
AUTHORS Profile



Dr. Naveen Reddy Nemili has received his MCA from Osmania University, Hyderabad in 2010, M.Tech degree in Computer science from JNTUH, Hyderabad in 2014 and Ph.D degree awarded in Computer Science from SSSUTMS, Madhyapradesh in 2024 He is dedicated to teaching field from the last 12 years. His research area Cloud Computing. At present he is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



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



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



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



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