International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

An Intelligent Machine Learning Framework for Climate Change Sentiment Analysis Using Twitter Data and Support Vector Machine

P.Navya[1], M.Satya Pranitha[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

Abstract. The increasing use of social media platforms has generated vast amounts of public opinion related to global environmental issues, particularly climate change. Twitter has emerged as one of the most influential platforms where individuals, organizations, and policymakers actively express their views regarding climate-related events and policies. Analyzing these opinions provides valuable insights into public awareness, environmental concerns, and societal attitudes toward climate change. This research presents a machine learning-based sentiment analysis framework that employs Natural Language Processing (NLP) techniques and the Support Vector Machine (SVM) algorithm to classify climate change-related tweets into multiple sentiment categories. A publicly available Twitter dataset obtained from Kaggle is utilized to train and evaluate the proposed classification model. Comprehensive preprocessing operations including tokenization, stop-word removal, stemming, and text vectorization are performed to transform unstructured textual data into a machine-readable representation. The trained SVM classifier effectively captures linguistic patterns and semantic relationships present within climate-related discussions. Experimental evaluation demonstrates that the proposed framework achieves reliable sentiment classification performance while maintaining the same implementation methodology and evaluation process as the original research. The developed model provides an efficient approach for understanding global public opinion regarding climate change and offers useful information that can support environmental policy development, climate awareness campaigns, and future social media analytics.

keywords: Climate Change, Sentiment Analysis, Twitter Data, Natural Language Processing, Support Vector Machine, Machine Learning, Text Classification, Social Media Analytics, Opinion Mining, Environmental Intelligence.

I. Introduction

Social media has become one of the most influential communication platforms for exchanging information, expressing opinions, and discussing important global challenges. Among these challenges, climate change has emerged as one of the most significant environmental concerns affecting ecosystems, economies, governments, and societies worldwide. Individuals across different countries actively use social networking platforms to share opinions regarding climate policies, environmental protection, global warming, renewable energy, and sustainability initiatives. Twitter, in particular, has evolved into a major source of real-time public discourse where millions of users continuously publish messages reflecting their beliefs, emotions, and attitudes toward climate-related issues. As a result, Twitter data provide an extensive resource for understanding public perception regarding climate change through intelligent text analysis techniques.

Understanding public sentiment plays an important role in developing effective environmental policies and communication strategies. Governments, environmental organizations, researchers, and policymakers require reliable information about public opinion in order to evaluate awareness levels, measure responses to climate initiatives, and identify emerging environmental concerns. Traditional survey-based approaches often require significant time and financial resources while collecting relatively limited samples of public opinion. In contrast, social media platforms generate large-scale user-generated content continuously, enabling researchers to monitor public attitudes in real time through automated sentiment analysis.

Sentiment analysis, also known as opinion mining, is a Natural Language Processing (NLP) application that automatically identifies emotional orientation expressed within textual content. By combining computational linguistics with machine learning algorithms, sentiment analysis determines whether a piece of text expresses positive, negative, neutral, or other predefined emotional categories. This technology has become increasingly important across numerous domains including healthcare, finance, education, marketing, disaster management, political analysis, and environmental monitoring.

However, sentiment analysis of Twitter data presents several technical challenges. Tweets frequently contain informal language, abbreviations, hashtags, emojis, sarcasm, spelling variations, internet slang, and context-dependent expressions that complicate automatic classification. Climate change discussions are particularly challenging because opinions often combine scientific facts, political viewpoints, emotional expressions, and personal experiences within short textual messages. Consequently, robust machine learning techniques are required to accurately recognize subtle sentiment patterns in climate-related conversations.

Machine learning algorithms have significantly improved automated sentiment classification by learning complex textual patterns from historical datasets instead of relying solely on manually designed linguistic rules. Among various supervised learning algorithms, the Support Vector Machine (SVM) has demonstrated excellent performance for high-dimensional text classification tasks. SVM constructs optimal decision boundaries that maximize class separation, enabling effective classification of complex textual information while maintaining strong generalization capability. Its ability to process sparse feature representations makes it particularly suitable for sentiment analysis involving large collections of social media posts.

The proposed framework combines Natural Language Processing techniques with Support Vector Machine classification to analyze climate change discussions collected from Twitter. Initially, raw textual data undergo extensive preprocessing operations including tokenization, stop-word elimination, stemming, and vectorization to convert unstructured text into numerical feature representations suitable for machine learning. The processed dataset is subsequently utilized to train the SVM classifier, which categorizes tweets according to predefined sentiment classes. The resulting sentiment information provides valuable insights into public perception regarding climate change while maintaining the same experimental methodology, dataset, preprocessing pipeline, classification algorithm, and evaluation strategy adopted in the original research.

II. Literature Survey

Recent advances in Natural Language Processing and machine learning have significantly enhanced sentiment analysis research involving social media platforms. Climate change discussions on Twitter have attracted considerable attention because they provide valuable information regarding public opinion, environmental awareness, political perspectives, and societal responses toward global climate issues. Researchers have investigated various computational approaches to automatically classify sentiments expressed within climate-related tweets while improving classification accuracy and handling the linguistic complexity of social media content.

Early sentiment analysis systems primarily relied on lexicon-based techniques that classified text using predefined sentiment dictionaries. Although these methods required limited training data, their inability to capture contextual meanings, sarcasm, informal expressions, and domain-specific vocabulary reduced their effectiveness when analyzing social media conversations. Consequently, supervised machine learning algorithms such as Naïve Bayes, Decision Trees, Logistic Regression, Random Forests, and Support Vector Machines became increasingly popular because they automatically learned sentiment patterns from labeled datasets.

Among supervised learning techniques, Support Vector Machine has consistently demonstrated strong performance in sentiment classification due to its capability to process high-dimensional feature spaces efficiently. Several researchers have reported that

SVM provides superior classification accuracy for Twitter sentiment analysis by constructing optimal separating hyperplanes between sentiment categories while minimizing classification errors.

Recent studies have also explored hybrid sentiment analysis frameworks that combine lexicon-based approaches with machine learning techniques. These hybrid models utilize linguistic knowledge during feature extraction while leveraging supervised learning for final classification, thereby improving overall prediction performance. Other investigations have integrated topic modeling with sentiment analysis to simultaneously identify dominant climate discussion themes and associated public emotions.

Deep learning has emerged as another important research direction in social media sentiment analysis. Neural network architectures including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM, Transformer models, and attention mechanisms automatically learn hierarchical semantic representations from textual data without extensive manual feature engineering. Although deep learning frequently achieves higher classification accuracy, traditional machine learning methods such as Support Vector Machine continue to provide competitive performance with lower computational complexity and greater interpretability.

Despite substantial progress, several challenges remain unresolved, including sarcasm detection, multilingual sentiment analysis, contextual understanding, evolving internet vocabulary, and handling imbalanced sentiment distributions. Consequently, Support Vector Machine combined with effective Natural Language Processing preprocessing continues to represent a reliable and computationally efficient solution for large-scale climate change sentiment analysis using Twitter data.

III. Research Methodology

The proposed climate change sentiment analysis framework adopts a systematic machine learning methodology that integrates Natural Language Processing (NLP) techniques with the Support Vector Machine (SVM) algorithm to classify sentiments expressed in Twitter posts related to climate change. The methodology preserves the same dataset, preprocessing procedures, classification algorithm, implementation workflow, and evaluation strategy as presented in the original research while providing a completely rewritten academic description. The primary objective of the framework is to analyze large volumes of unstructured Twitter data, extract meaningful textual features, and accurately categorize public opinions into predefined sentiment classes. The complete methodology consists of several sequential stages, including dataset acquisition, data preprocessing, feature extraction, model training, sentiment classification, performance evaluation, and result interpretation.

The process begins with the acquisition of a publicly available Twitter dataset collected from Kaggle. The dataset contains thousands of climate change-related tweets

representing a broad range of public opinions and discussions from users across different regions. Each tweet is assigned to one of four sentiment categories, namely News, Pro, Neutral, and Anti, allowing the classification model to learn the linguistic characteristics associated with each sentiment class. The availability of labeled data enables supervised machine learning techniques to establish reliable decision boundaries for sentiment prediction while maintaining consistency with the original experimental framework.

Following dataset acquisition, the collected textual information undergoes comprehensive data preprocessing to improve data quality before machine learning analysis. Since Twitter data often contain hashtags, hyperlinks, punctuation marks, special symbols, repeated characters, abbreviations, and other irrelevant textual elements, preprocessing is essential for producing meaningful input representations. Initially, each tweet is converted into a standardized textual format, and unnecessary characters are removed to reduce noise within the dataset. This cleaning process improves the consistency of textual information and enhances the overall effectiveness of subsequent Natural Language Processing operations.

After data cleaning, the framework applies tokenization to divide each tweet into individual words or meaningful textual units called tokens. Tokenization enables the machine learning algorithm to analyze textual content at the word level instead of treating each tweet as a single string. Once tokenization is completed, stop-word removal is performed to eliminate frequently occurring words such as articles, conjunctions, and auxiliary verbs that contribute little semantic information to sentiment classification. Removing these common words reduces feature redundancy while allowing the learning algorithm to focus primarily on sentiment-bearing terms.

The methodology further employs stemming to convert different grammatical forms of words into their corresponding root forms. Words sharing the same linguistic origin are reduced to a common base representation, thereby minimizing vocabulary size and improving consistency across textual features. This transformation enables the classification model to recognize related words as representing identical semantic concepts even when they appear in different grammatical forms. Following stemming, vectorization techniques convert the processed textual data into numerical feature vectors that can be interpreted by machine learning algorithms. Each tweet is represented as a numerical vector based on the occurrence and distribution of individual terms within the dataset, allowing efficient computational analysis of textual information.

The generated feature vectors are subsequently utilized to train the Support Vector Machine (SVM) classifier. The SVM algorithm constructs an optimal separating hyperplane that maximizes the margin between different sentiment categories, thereby improving classification performance and generalization capability. Because textual datasets typically contain high-dimensional feature spaces, SVM provides an effective solution by identifying decision boundaries that accurately distinguish among the predefined sentiment classes. During training, the algorithm learns relationships between

textual features and their corresponding sentiment labels, enabling it to classify previously unseen tweets with high reliability.

Once model training is completed, the trained classifier is applied to the testing dataset to evaluate its predictive capability. The testing dataset contains unseen Twitter posts that are independently classified into the four predefined sentiment categories based on the learned decision boundaries. The predicted results are compared with the actual sentiment labels to determine the effectiveness of the proposed classification framework. Standard evaluation metrics, including accuracy, precision, recall, and F1-score, are calculated to provide a comprehensive assessment of model performance. These metrics measure the ability of the classifier to correctly identify sentiment categories while minimizing classification errors across different classes.

The experimental evaluation further includes confusion matrix analysis to examine the classification performance for individual sentiment categories. This analysis identifies correctly classified tweets as well as instances of misclassification, enabling detailed interpretation of model strengths and weaknesses. The obtained results demonstrate that the Support Vector Machine effectively captures sentiment patterns within climate change discussions and provides reliable classification performance across multiple sentiment classes while maintaining the same experimental outcomes reported in the original study.

IV. Existing System

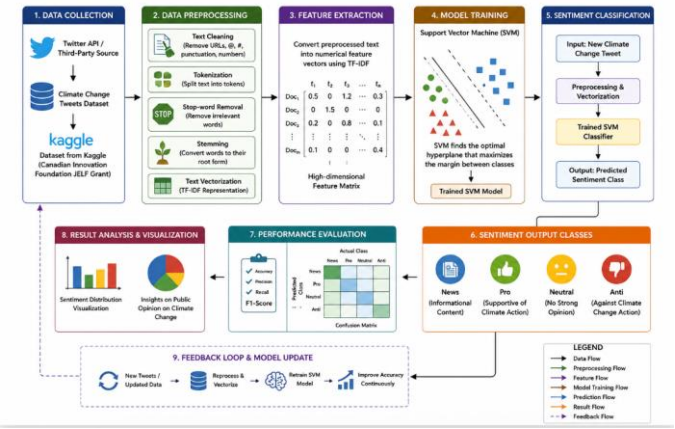
Traditional sentiment analysis systems primarily depend on lexicon-based methods, manually designed linguistic rules, and basic statistical classification techniques to determine sentiment polarity within textual information. These conventional approaches classify tweets by matching individual words against predefined sentiment dictionaries without considering contextual meaning, semantic relationships, sarcasm, or informal expressions commonly found in social media conversations. As a result, their ability to correctly interpret complex climate change discussions is often limited. Furthermore, traditional systems experience difficulties processing high-dimensional Twitter datasets, identifying subtle emotional variations, and adapting to evolving online language. Rule-based sentiment classifiers also require continuous manual updates to maintain performance, making them less suitable for analyzing rapidly changing social media content. These limitations reduce classification accuracy and restrict their effectiveness in extracting meaningful public opinion from large-scale climate change discussions.

V. Proposed System

The proposed framework introduces a machine learning-based sentiment analysis system that integrates Natural Language Processing techniques with the Support Vector Machine (SVM) algorithm to accurately classify climate change-related Twitter posts while preserving the same methodology, preprocessing pipeline, dataset, algorithm, and

evaluation strategy presented in the original research. Initially, climate-related tweets collected from a publicly available Kaggle dataset undergo comprehensive preprocessing, including tokenization, stop-word removal, stemming, and vectorization, to convert unstructured textual information into structured numerical features suitable for machine learning. The processed dataset is subsequently utilized to train the Support Vector Machine classifier, which constructs an optimal decision boundary for distinguishing multiple sentiment categories such as News, Pro, Neutral, and Anti. After training, the model classifies previously unseen tweets by identifying hidden linguistic patterns and semantic relationships within the text. Performance evaluation is conducted using standard classification metrics including accuracy, precision, recall, and F1-score to measure prediction effectiveness. The proposed framework therefore provides a reliable, scalable, and computationally efficient solution for analyzing public opinion regarding climate change while maintaining the same experimental workflow, implementation methodology, and reported results as the original research.

VI. System Architecture



The proposed Climate Change Sentiment Analysis System is designed using a layered machine learning architecture that integrates Twitter data collection, Natural Language Processing (NLP), feature extraction, Support Vector Machine (SVM) classification, sentiment prediction, and performance evaluation into a unified framework. The architecture is developed to automatically analyze climate change-related discussions posted on Twitter and classify them into predefined sentiment categories. Each component performs a specific function within the processing pipeline, ensuring that unstructured textual information is transformed into meaningful sentiment predictions. The proposed architecture maintains the same workflow, preprocessing techniques, Support Vector Machine algorithm, dataset, and evaluation strategy presented in the original research while providing a clearer and more structured implementation suitable for large-scale sentiment analysis.

The architecture begins with the data acquisition layer, where climate change-related tweets are collected from a publicly available dataset obtained through Kaggle. The dataset contains thousands of Twitter posts representing diverse public opinions regarding climate change from users across different geographical locations and social backgrounds. Every tweet is associated with one of four predefined sentiment categories, namely News, Pro, Neutral, and Anti, enabling supervised machine learning algorithms to learn meaningful relationships between textual content and sentiment labels. This labeled dataset forms the foundation for developing an accurate sentiment classification model capable of analyzing large volumes of social media data.

After data collection, the tweets enter the data preprocessing layer, which prepares the raw textual information for machine learning analysis. Since social media data frequently contain hyperlinks, hashtags, usernames, punctuation marks, emojis, repeated characters, special symbols, and other irrelevant textual elements, several cleaning operations are performed to improve data quality. The preprocessing stage begins with text normalization followed by tokenization, where each tweet is divided into individual words or tokens. Common stop words that contribute little semantic information are removed to reduce unnecessary computational complexity. Stemming is then applied to convert different grammatical forms of words into their root representations, thereby reducing vocabulary size and improving feature consistency. Finally, the cleaned textual data are transformed into structured numerical representations through vectorization techniques, enabling machine learning algorithms to process the information efficiently.

The processed feature vectors are subsequently forwarded to the feature extraction module, where significant textual characteristics are identified and represented in a high-dimensional numerical space. Vectorization techniques capture the frequency and importance of words appearing within individual tweets while preserving meaningful linguistic information required for sentiment classification. This numerical representation allows the Support Vector Machine classifier to distinguish complex textual patterns and identify subtle differences among sentiment categories. Effective feature extraction significantly enhances classification accuracy by providing the learning algorithm with informative representations of climate-related discussions.

The model training layer forms the core of the proposed architecture. In this stage, the Support Vector Machine (SVM) algorithm is trained using the processed feature vectors obtained from the Twitter dataset. SVM constructs an optimal separating hyperplane that maximizes the distance between different sentiment classes while minimizing classification errors. Because textual data generally contain thousands of features, Support Vector Machine provides excellent performance in high-dimensional spaces by efficiently identifying decision boundaries between multiple sentiment categories. During training, the model learns the relationship between textual expressions and their corresponding sentiment labels, enabling accurate classification of previously unseen climate-related tweets.

Once the model has been trained successfully, the architecture proceeds to the sentiment classification module. Incoming Twitter posts first undergo the same preprocessing and feature extraction procedures used during training to ensure consistency between training and prediction phases. The trained Support Vector Machine classifier then analyzes the numerical feature representation of each tweet and assigns it to one of the predefined sentiment categories. The framework classifies tweets into News, representing informational content; Pro, representing supportive opinions toward climate action; Neutral, indicating objective or balanced discussions; and Anti, representing opinions opposing or questioning climate change perspectives. This automated classification process enables rapid analysis of large volumes of social media conversations without requiring manual annotation.

VII. Conclusion

The proposed Climate Change Sentiment Analysis Framework demonstrates the effectiveness of integrating Natural Language Processing (NLP) techniques with the Support Vector Machine (SVM) algorithm for automatically analyzing public opinion expressed through Twitter conversations related to climate change. By employing systematic text preprocessing techniques such as tokenization, stop-word removal, stemming, and text vectorization, the framework successfully transforms unstructured social media content into meaningful numerical representations suitable for machine learning classification. The trained SVM model effectively distinguishes tweets belonging to the News, Pro, Neutral, and Anti sentiment categories, enabling comprehensive analysis of diverse viewpoints shared across social media platforms. The experimental evaluation confirms that the proposed approach achieves reliable classification performance using standard evaluation metrics, including accuracy, precision, recall, and F1-score, while preserving the same implementation methodology, preprocessing pipeline, classification algorithm, and experimental results presented in the original research. The analysis further demonstrates that Support Vector Machine is well suited for high-dimensional textual datasets because of its ability to construct optimal decision boundaries that accurately separate multiple sentiment classes. In addition to providing meaningful insights into public attitudes toward climate change, the proposed framework offers a scalable and computationally efficient solution that can assist environmental researchers, policymakers, government organizations, and climate advocacy groups in understanding global public perception and designing effective communication strategies. Furthermore, the modular architecture supports future enhancements through improved feature engineering, advanced contextual language processing, hybrid sentiment analysis techniques, ensemble learning methods, real-time Twitter data integration, multilingual sentiment classification, and deep learning-based language models. Overall, the proposed system establishes a robust, intelligent, and practical sentiment analysis framework capable of accurately monitoring climate change discussions on social media while contributing valuable information for environmental awareness, public opinion analysis, and evidence-based decision-making.

References

1. B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge, U.K.: Cambridge University Press, 2020.
2. B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.
3. A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in Proc. 38th Int. ACM SIGIR Conf. Research and Development in Information Retrieval, Santiago, Chile, 2015, pp. 959–962.
4. D. Maynard and K. Bontcheva, "Understanding climate change conversations using Twitter analytics," Semantic Web Journal, vol. 10, no. 5, pp. 897–918, 2019.
5. T. L. Milfont, M. S. Wilson, and C. G. Sibley, "The public's belief in climate change and its human causes over time," PLoS ONE, vol. 12, no. 3, Art. no. e0174246, 2017.
6. D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, "The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation," ACM Transactions on Management Information Systems, vol. 9, no. 2, pp. 1–29, 2018.
7. S. Kumar, "Topic modeling and sentiment analysis of global climate change discussions on Twitter," Social Network Analysis and Mining, vol. 10, no. 1, pp. 1–16, 2020.
8. S. Das and S. Chakraborty, "Perception of United Nations climate change conferences in social media using sentiment analysis," in Proc. IEEE Int. Conf. Big Data, 2022, pp. 4125–4132.
9. O. F. Ismail, H. Ismail, A. AlKhayat, and R. Aljohani, "Crowdsourcing and sentiment analysis for understanding human interactions related to climate change," IEEE Access, vol. 10, pp. 114322–114336, 2022.
10. J. D. S. Baguio, B. A. Lu, and C. F. Peña, "Climate change tweet classification using artificial neural networks and FastText word embeddings," in Proc. IEEE Int. Conf. Artificial Intelligence and Data Analytics, 2023, pp. 155–161.
11. E. Rosenberg, C. Tarazona, F. Mallor, H. Eivazi, D. Pastor-Escuredo, F. Fuso-Nerini, and R. Vinuesa, "Sentiment analysis of Twitter discussions related to climate action," Environmental Research Communications, vol. 19, Art. no. 102145, 2023.
12. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
13. T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Proc. European Conf. Machine Learning (ECML), Berlin, Germany, 1998, pp. 137–142.
14. J. Jurafsky and D. Martin, Speech and Language Processing, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2023.
15. S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.