



A Machine Learning Based Twitter Sentiment Analysis Framework for Climate Change Discussions

K Naresh¹, Velamburu pavithra²

¹Assistant Professor, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

²Postgraduate, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

How to Cite this Article:

pavithra, V. (2026). A Machine Learning Based Twitter Sentiment Analysis Framework for Climate Change Discussions. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(03). <https://doi.org/10.55041/ijcope.v2i4.063>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i4.063>

Abstract

One of the most talked-about topics in the world today is climate change, which has sparked a lot of public discussion on social media. Particularly on Twitter, users regularly express their opinions, worries, and attitudes regarding climate-related subjects, making it a valuable source of current public opinion. Automated sentiment analysis is a useful solution because manually analysing these massive amounts of textual data is difficult and time-consuming. This study uses Twitter data to analyse public opinion on climate change using a machine learning-based methodology. The suggested system gathers tweets about climate change and uses a variety of natural language processing methods, including stopword removal, tokenisation, text cleaning, and lemmatisation. In order to determine attitudes like positive, negative, or neutral, the processed data is subsequently transformed using a vectorisation technique and classed using a trained machine learning model. Additionally, users can enter text and receive real-time sentiment forecasts using a web-based interface created with Flask. The suggested approach shows how social media analytics and machine learning may be successfully integrated to comprehend how the public views climate change and promote data-driven environmental awareness and decision-making.

Keywords

Climate Change, Sentiment Analysis, Twitter Data, Machine Learning, Natural Language Processing, Social Media Analytics



I. Introduction

One of the most important environmental issues the world is currently facing is climate change. Scientists, decision-makers, and the general public are deeply concerned about environmental degradation, rising sea levels, harsh weather, and rising global temperatures. People are increasingly expressing their thoughts and engaging in online conversations regarding climate change due to the quick development of social media platforms. Among these, Twitter has become a well-liked tool for people to express their ideas, opinions, and responses in real time about environmental issues. The vast amount of textual data produced on Twitter offers insightful information about how the public views and understands climate change. However, it is challenging and ineffective to manually analyse such large volumes of data. This has led to the necessity for automated systems capable of efficiently analysing and interpreting public opinion. A branch of natural language processing (NLP) called sentiment analysis assists in recognising and classifying opinions conveyed in text into various sentiment categories, including neutral, negative, and positive.

Sentiment classification jobs are now much more accurate and efficient thanks to machine learning algorithms. Machine learning algorithms can automatically identify patterns and categorise feelings from textual data by training models on labelled datasets. These methods have been used by researchers more frequently in recent years to examine public opinion on social media platforms.

This study proposes a machine learning-based sentiment analysis method to examine Twitter conversations about climate change. The system transforms textual data into machine-readable formats using vectorisation and text preparation techniques. The sentiment of the input text is then predicted using a trained classification algorithm. Additionally, a web-based interface created with Flask makes it simple for users to enter text and receive sentiment forecasts. The suggested approach seeks to show how machine learning may enhance social media analysis and offer insights into popular opinions regarding climate change.

II. Proposed System / Methodology

Using Twitter data and a machine learning-based framework, the proposed method seeks to assess public opinion regarding climate change. Through a web-based interface, the system is intended to automatically process textual material, categorise feelings, and display the prediction findings. The primary goal of the suggested

system is to offer a productive and automated method for comprehending public perceptions of climate change conversations on social media.

Data preprocessing, feature extraction, machine learning model training, and sentiment prediction are some of the steps that make up the system's entire workflow. First, the system gathers textual data about climate change from a dataset of climate-related tweets.

Preprocessing techniques are used to clean and normalise the text data because raw social media material frequently contains noise such as special characters, URLs, hashtags, and stop words. Preprocessing involves filtering out stop words, removing superfluous letters, and converting the text to lowercase. Words are also reduced to their most basic versions using tokenisation and lemmatisation approaches. These actions improve the machine learning model's performance and the quality of the textual data. Following preprocessing, a vectorisation method like Count Vectorizer or TF-IDF Vectorizer is used to convert the cleaned text input into numerical features. In order for machine learning algorithms to handle textual data, this phase transforms it into a machine-readable format.

The processed dataset is then used to train a machine learning classification model to recognise various sentiment categories, such as neutral, negative, and positive. The model is saved and incorporated into a Flask-based web application after it has been trained. Users can enter a text or tweet about climate change through the web interface, and the technology instantly guesses the attitude that goes along with it. The user interface, the Flask backend server, and the machine learning prediction module are the three primary parts of the suggested system design. Efficient sentiment analysis is made possible by this integrated technology, which also offers a user-friendly platform for examining public sentiments regarding climate change discussions.

III. Dataset Description

Twitter posts pertaining to conversations on climate change make up the dataset used in this study. There is a lot of publicly accessible textual data on social media sites like Twitter, where individuals often share their thoughts and attitudes about environmental issues. These tweets are a useful resource for examining how the general public views climate change.

Numerous tweets with climate-related terms and conversations on sustainability, global warming, and environmental issues are included in the dataset. A sentiment label that reflects the text's emotional tone is

linked to every tweet in the dataset. Positive, negative, and neutral classes are typically included in the sentiment categories. These labels aid in the machine learning model's ability to identify patterns that differentiate between the various emotions conveyed in the tweets.

The sentiment categorisation model is trained and assessed using the dataset. The algorithm can discover how various words and phrases affect sentiment classification by examining this dataset. This makes it possible for the program to forecast the tone of fresh user input pertaining to conversations on climate change. When developing a machine learning model for sentiment analysis, the dataset is crucial.

IV. Data Preprocessing

Because raw textual data from social media platforms frequently contains noise and unstructured information, data preparation is a crucial stage in sentiment analysis. Emojis, hyperlinks, hashtags, special characters, and other extraneous components can all be found in tweets, which can have a detrimental effect on how well machine learning algorithms work. Thus, before the model is trained, preprocessing methods are used to clean and normalise the textual input.

To preserve consistency and prevent word duplication with varying case, the first step in preprocessing is to transform all text into lowercase format. The text is then cleared of undesired characters including punctuation, digits, URLs, and special symbols. This aids in the dataset's reduction of superfluous noise.

The text is then tokenised into individual words so that the machine may do word-level analysis. Common words like "the," "is," and "and" are eliminated because they don't provide any meaning to sentiment analysis. Lemmatisation is also used to reduce variants of the same term by transforming words into their base or root form.

The cleaned text input becomes more structured and appropriate for feature extraction and machine learning model training after these pretreatment procedures are finished. The accuracy and effectiveness of the sentiment categorisation system are greatly increased by appropriate preprocessing.

V. System Implementation and User Interface

The suggested sentiment analysis system is put into practice as a web application that enables users to examine material pertaining to climate change and determine the sentiment conveyed in the input. The technology provides real-time sentiment forecasts by combining machine

learning methods with an intuitive user interface. The application's frontend interface is designed using HTML and CSS, and its backend processing is handled by Python and the Flask framework. Through the system interface, users can enter text pertaining to talks about climate change and have its sentiment automatically analysed. Before predicting the sentiment category, the machine learning model preprocesses and extracts features from the text. Confidence scores and the anticipated outcome are then shown on the online interface.

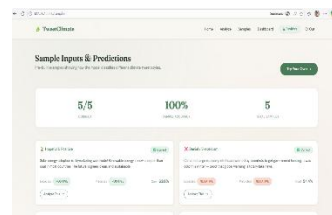


Figure 1: Sample Sentiment Prediction Interface

The trained model's ability to categorise various climate-related tweet kinds is seen in the sample prediction interface. After analysing several sample inputs, the algorithm forecasts the sentiment categories that correlate to those inputs. The interface shows the total sample accuracy as well as the number of accurately predicted samples. Both the expected sentiment and the predicted sentiment produced by the machine learning model are labelled on each sample tweet.

The algorithm can successfully identify positive, negative, and neutral feelings in discussions connected to climate change, according to the results displayed in the interface. For every prediction, the system also offers a confidence score that shows the likelihood of the anticipated mood.

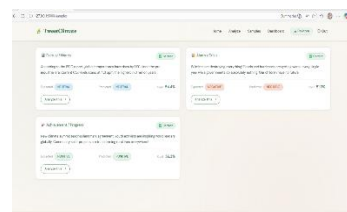


Figure 2: Climate Tweet Sentiment Classification Results

This interface displays a variety of climate-related tweet examples along with the sentiment classification outcomes. The approach examines a variety of communications, including talks about climate policies, environmental concerns, and factual claims. The model divides the sentiment into good, negative, and neutral categories based on the linguistic characteristics taken from the tweets. The algorithm effectively discerns between assertions about climate change that are

informative, sceptical, and optimistic. This illustrates how well the machine learning model recognises a range of opinions stated in social media conversations.

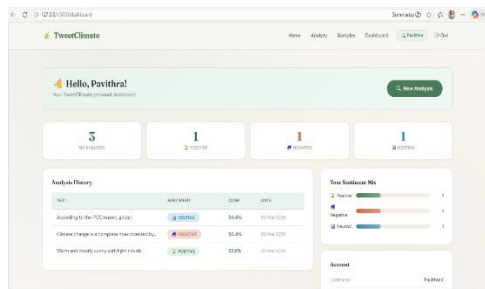


Figure 3: User Dashboard and Sentiment Analysis History

A customised summary of the user's sentiment analysis activity is given by the dashboard interface. It shows the distribution of positive, negative, and neutral feelings as well as the overall number of analyses completed. The analysed text, anticipated sentiment, confidence score, and analysis date are all recorded in the dashboard's history of prior studies.

In order to help users comprehend the general sentiment trends in their examined data, the system also provides a visual depiction of the sentiment distribution. This feature enhances user interaction and increases the system's usability and informativeness.

VI. Conclusion

This study used Twitter data to offer a machine learning-based paradigm for assessing public opinion on climate change. The suggested system prepares the textual data for sentiment classification using natural language processing methods as tokenisation, vectorisation, and text preprocessing. Real-time sentiment prediction for climate-related material is achieved by integrating a Flask-based web application with a trained machine learning model. In order to better comprehend public attitudes on climate change issues, the algorithm effectively classifies tweets into positive, negative, and neutral sentiments. By making it simple for users to analyse text and see prediction results, the designed web interface enhances usability. All things considered, the suggested method shows how well machine learning and social media analytics can be combined to research public opinion and environmental consciousness.

References

- [1] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2012.
- [2] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta, 2010, pp. 1320–1326.
- [3] S. Kiritchenko, X. Zhu, and S. Mohammad, "Sentiment Analysis of Short Informal Texts," *Journal of Artificial Intelligence Research*, vol. 50, pp. 723–762, 2014.
- [4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2013.
- [5] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," *Stanford University Technical Report*, 2009.