



# Machine Learning-Based Sentiment Analysis of Climate Change Discussions on Twitter

S. Srinivas<sup>1</sup>, Dharmana Akhila<sup>2</sup>

Assistant Professor, Department of CSE, Megha Institute Of Engineering And Technology For Womens, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301<sup>1</sup>

M.Tech Student, Department of CSE, Megha Institute Of Engineering And Technology For Womens, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301<sup>2</sup>

**Abstract:** A more nuanced and complicated public opinion on climate change discussions on Twitter is what we want to shed light on. This study makes use of state-of-the-art ML and NLP tools, most notably Support Vector Machines (SVM), to do sentiment analysis on a worldwide scale using a large dataset acquired from a trustworthy external source. Focusing on both positive and negative emotions, this research aims to identify the many ways in which people express themselves emotionally while discussing climate change. Critically examined are the veracity and practicality of the chosen third-party dataset, which was acquired via Kaggle and made feasible by a Canadian Innovation Foundation JELF grant that was bestowed onto Chris Bausch of the University of Waterloo. With a remarkable F1 score of 0.70, the SVM-based sentiment analysis shows how well the selected approach captures the nuanced nature of climate change arguments on Twitter. Legislators, groups aiming to reach a worldwide audience, and climate change activists may all benefit from the research's communication tactics. The complex link between public opinion on climate change and online discourse is better understood in this research, which makes use of an external dataset and the sentiment analysis component of the Support Vector Machines method. In conclusion, this work adds substantial new information to the expanding field at the intersection of social media dynamics and environmental awareness by demonstrating the efficacy of support vector machines (SVMs) in identifying sentiment subtlety in massive datasets.

**Keywords:** Climate, sentiment analysis, and complexity are related terms.

## INTRODUCTION

Now more than ever, people can voice their thoughts and ideas on pressing digital concerns on various social media platforms. The effects of climate change on ecosystems, economies, and societies are hotly contested in these discussions. In recent years, Twitter has developed into a venue where people can share their views on global warming in real time. Using a database, this article intends to analyse the feelings conveyed in climate change-related Twitter interactions. An in-depth sentiment analysis of climate change-related tweets will provide light on the feelings, opinions, and points of view expressed in these messages. Understanding how people think and talk about climate change is important for understanding how society's dynamics of advocacy and protection are changing.

## LITERATURE REVIEW

The connection between sentiment analysis and climate change-related social media conversations has been the subject of much prior study. They have looked at different kinds of sentiment, how the media affects people's views on the environment, and how well sentiment analysis captures public opinion. To completely understand the intricacies of sentiment on Twitter, however, further research is required. In the context of media conversations about climate change, sentiment analysis has been the subject of a great deal of academic work. • By integrating the benefits of two methods, this methodology [1] may make sentiment analysis more successful. To determine the tone of specific phrases, a lexicon-based method is used; to categorise all tweets, a machine-based method is employed. The authors test their method on datasets consisting of tweets and discover that it achieves better results than methods that rely only on lexicons or machine learning. • An open-source programme for analysing tweets on climate change has been created by the authors [2]. Due to the abundance of information, the general public's understanding of climate change may be difficult to decipher; this toolbox aims to rectify that. Climate change and the COVID-19 pandemic, Twitter attitudes towards climate change data, and social media initiatives for environmental protection are just a few of the research and publications that have used the toolkit. Both the University of Sheffield website and the ACM Digital Library acknowledged the author's work, demonstrating its recognition in academic circles. Researchers interested in public sentiment and climate change activism on social media will find the Toolkit to be an invaluable resource. Citations of this study [3] are common in the academic literature, attesting to its significance for sentiment analysis. The study's main emphasis is on improving point-of-interest



classification using support vector machine (SVM), which entails locating a hyperplane in an N-dimensional space. By maximising the interval between classes, this strategy seeks to identify the best choice limit. Among the many uses for the study's results are searches for very unfavourable views, domain-independent sentiment categorisation, and analysis of informal political debate. Deep learning for Arabic sentiment analysis has also made mention of this study. To improve aspect-based sentiment analysis, the author [4] suggests a new hierarchical method. Their approach successfully captures the hierarchical structure of user evaluations, which are composed of many characteristics and are linked to distinct moods. This is the fundamental novelty. A methodology is presented by the authors that analyses the attitude expressed towards each feature of a review after identifying the relevant factors. By including this hierarchical structure, the model enhances the level of detail in sentiment analysis, allowing for a more comprehensive comprehension of user sentiment from many angles. In analytics in particular, a thorough comprehension of user sentiment necessitates a sophisticated link between elements and emotions, which this study helps to build. • The public's evolving views on climate change and its anthropogenic origins are the subject of the study [5]. Researchers in New Zealand discovered that more and more individuals are buying into the idea that human activity is causing climate change. This change is more pronounced in years when public acceptability is on the rise. A number of reasons, including increased scientific agreement on the reality of climate change and widespread media coverage of related weather occurrences, have contributed to this shift, according to the authors. Taken together, these results point to a growing public knowledge and support for taking action on climate change, as well as an improvement in popular trust in the science behind it. Research on sentiment analysis, social media content analysis, and prediction algorithms has included references to the study [6]. A few examples include research on using recurrent neural networks for sentiment analysis on Twitter and an analysis of customer reviews using CNN-Bi LSTM. Another article discusses the use of deep learning for sentiment analysis. Pieces such as A Two-Stage Voting Method for Ensemble Learning in Sentiment Classification with a Sensitive Grass Hopper Optimisation Algorithm and Context-Aware Analysis for Twitter Post Effects as well as Ensemble Learning in Sentiment Classification in Social Networks make reference to this work.

A thorough study and review of sentiment analysis techniques for Twitter data was proposed by the author [7]. The authors cover a wide range of methods, from deep learning models to algorithms based on lexical information. Utilising an extensive dataset of annotated Twitter data, they assess the efficacy of different methods. According to the results, deep learning models outperform more conventional approaches when it comes to sentiment categorisation. Depending on the application and database, the article demonstrates, is crucial when selecting a sensitivity analysis approach. The ever-changing nature of language and social media platforms further highlights the need of constantly evaluating and improving sentiment analysis methodologies. • The goal of the author's analysis of tweets on climate change was to determine the attitudes represented in such tweets [8]. The study employs a dual strategy, integrating topic modelling and sentiment analysis methods. To begin, the study finds the most common threads and topics of conversation among the climate change tweets by using topic modelling. The breadth of discussion that characterises internet conversation may be better grasped with this first step. Afterwards, the research uses sentiment analysis to understand how people feel about climate change by analysing the tone of these tweets. Kumar hopes to shed light on the complexities of the climate change discussion in online communities by integrating topic models with sentiment analysis; this will allow us to better grasp the underlying themes and nuanced emotions surrounding this pressing international problem. • In order to decipher the subtle undertones and rhetorical devices used in tweets about climate change, the research [9] applies a sophisticated method to sentiment analysis. The writers shed light on the complicated dynamics of online debates regarding climate change by conducting a thorough and lucid analysis of a variety of Twitter exchanges. They studied the rhetorical techniques of sarcasm and unsociable speech. We gain a better understanding of the complex nature of public engagement with scientific subjects on social media by learning about the nuanced language and sentiment on Twitter, which in turn sheds light on how people communicate and discuss climate change. To better capture the subtleties in tweets, deep learning modified neural networks (DLMNNs) are being used to enhance sentiment analysis [10]. To improve understanding and interpretation of sentiment patterns, the DLMNN model incorporates state-of-the-art neural network designs and makes use of deep learning methods. This work makes a valuable contribution to the ever-changing field of sentiment analysis by examining the difficulties caused by the peculiarities of Twitter data, including the prevalence of informal and slang terms. His study lays forth a novel strategy for enhancing the precision and efficacy of sentiment analysis techniques within the ever-changing environment of online discourse.

## METHODOLOGY

Sentiment analysis makes use of state-of-the-art natural language processing (NLP) methods and support vector machine (SVM) to achieve its objective. This dataset, compiled by Chris Bausch of the University of Waterloo with funding from the Canadian Innovation Foundation JELF and made available on Kaggle, is a fantastic illustration of the wide range of global viewpoints on the climate crisis. Using its excellent grasp of sentiment subtlety, SVM is a major player in the field of sentiment identification in tweets. More specifics on the methodology and approach utilised to do sentiment analysis



on the dataset will be provided later on. We will also discuss the consequences of applying sentiment analysis to social media posts on climate change. Understanding people's perspectives on social media is the driving force behind this study. Section A. Acquiring Datasets: Chris Bauch of the University of Waterloo was able to access the Kaggle dataset used in this study thanks to a JELF grant from the Canadian Innovation Foundation. Chris Bauch has accumulated a large collection of tweets addressing many topics, including climate change. He is a reputable source. The tweets in the dataset have labels given to them from the following categories: 2. This tweet offers a safe source of climate change information. • 1 (Pro): The human-caused climate change argument is bolstered by this tweet. • 0 (Neutral): The existence of human-caused climate change is not affirmed or denied in this tweet. • -1 (Anti): This tweet denies that human activity is a major contributor to global warming. Analysing Data

Sentiment Distribution of the Tweet Dataset

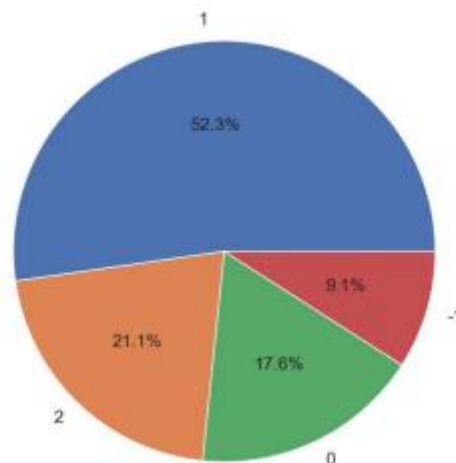


Fig.1 Dataset consists of 43949 entries with no null counts.

Subject: B. Preparing Data To ensure that raw, unprocessed text data is clean and available for further analysis, data preparation is a crucial part of sentiment analysis. This method comprises several essential steps to transform unstructured data into a structured format that permits substantial insights. To begin, all texts are tokenized and broken down into its component pieces. This sets the stage for further filtering and analysis. Some examples of fundamental operations in natural language processing (NLP) include tokenization, stemming, vectorization, and stop word removal. To better grasp the text's underlying sentiment, various natural language processing techniques work together to prepare the data for the application of machine learning models. • Tokenization: This method involves breaking down large pieces of text into smaller ones, called tokens, which are often phrases or words. In sentiment analysis, tokenization is breaking down each passage or tweet into its own set of terms. This technique prepares the text for further text processing, which facilitates the subsequent analysis. As an example, the line "Climate change is a global challenge" may be tokenized as follows: ["global", "challenge", "is", "climate", "is"] Cut Out Superfluous Words: "Stop words" are common terms that are often removed from text analyses due to their lack of semantic significance and their ability to introduce noise into the study. The first step in stop word removal is to construct a dictionary of often used terms (like "and," "the," etc.). • Rooting By removing suffixes and returning words to their base, or root form, stemming makes it possible to regard words that share the same root as equivalent. In sentiment analysis, stemming helps to reduce the dimensionality of the data by bringing word variants back to their common root, which benefits in consolidation. A root-reduction analysis would reveal that the words "running" and "runner" both originate from the word "run." • Vectoring: Machine learning numerical forms of text information, such as Bag-of-Words (BoW), may be understood and analysed by algorithms for sentiment analysis when vectorized. One vector for every document (tweet) in a three-dimensional space. As an example of a Bag-of-Words vector, the following may be expressed as "Climate change is a global challenge": one, one, one, one, one. Through the implementation of these data preparation techniques, we transformed raw text data into a format suitable for sentiment analysis. Vectorization changes, stemming minimises variations, tokenization divides the text, and stop word removal focusses on significant terms. When evaluating the model, we utilise the testing set—which accounts for 25% of the dataset—as a benchmark. Precision, accuracy, recall, and F1 score are the included parameters that are used to quantify the model's accuracy in attitude classification.



- Accuracy : 0.7230111050427818
- Precision : 0.7242155563925576
- Recall : 0.7230111050427818
- F1 : 0.7054239413690935

## RESULT AND DISCUSSION

The overall accuracy of the algorithm for sentiment analysis is 76.25%. For this dataset, this means that the system accurately recognised 76.25 percent of the tweets. The following table displays the model for emotion analysis based on the confusion matrix:

Actual	Predicted	Count
News	News	5000
News	Pro	343
News	Neutral	178
News	Anti	444
News	Pred -1	53
Pro	News	33
Pro	Pro	806
Pro	Neutral	955
Pro	Anti	126
Pro	Pred -1	1
Neutral	News	1
Neutral	Pro	21
Neutral	Neutral	246
Neutral	Anti	5134
Neutral	Pred -1	378
Anti	News	0
Anti	Pro	2
Anti	Neutral	12
Anti	Anti	39
Anti	Pred -1	558
Anti	Pred 0	1660

Table 1: Emotion Analysis

The system achieves the highest performance (81.33%) when determining if tweets are genuine news. On top of that, the algorithm accurately classifies tweets as either pro- or anti-climate change (76.69% and 70.46% in that order, respectively). Alternatively, the neutrality of tweets is not accurately identified by the algorithm (55.58%). According to our results, the sentiment analysis system is capable of accurately categorising climate change-related tweets. And yet, the system does worse when it comes to classifying neutral tweets. This is a common issue in sentiment analysis that researchers are currently trying to resolve. To address this issue, a hybrid strategy combining SVM's strong classification



skills with TextBlob's and VADER's subtle sentiment analysis capabilities may be developed. This approach will take use of the best features of each tool. An improved sentiment analysis model that takes context into account is possible as a result of this merging of approaches.

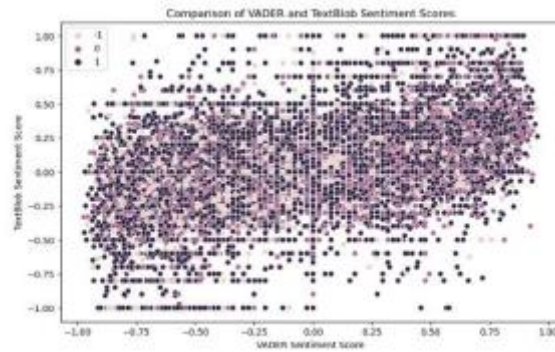


Fig.2 Sentiment scores comparison using scatter plot

This graph displays the correlation between the climate change-related tweets' Vader and TextBlob sentiment ratings. The emotion score for Vader is shown on the x-axis, while the sentiment score for TextBlob is shown on the y-axis. The size of the point indicates how many tweets had that particular emotion score.

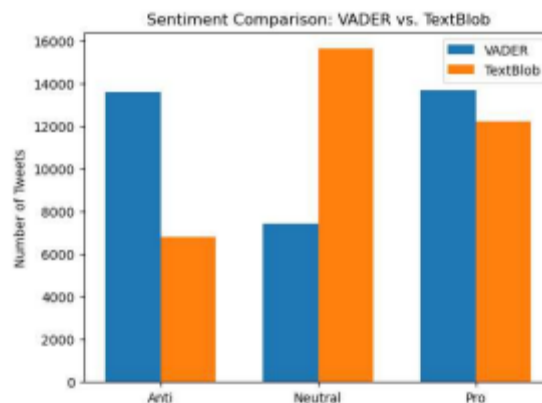


Fig.3 Sentiment comparison using bar graph

According to the graph, the two sentiment ratings are positively correlated with one another. As a result, a higher Vader sentiment score is indicative of a more positive TextBlob sentiment score for the same tweet. While there is a connection, it is not perfect. Tweets with a high Vader sentiment score don't always have a high TextBlob sentiment score. However, the Vader sentiment score is often low in tweets with a high TextBlob sentiment score. Comparing TextBlob and Vader sentiment ratings, we find that they are 46% accurate. One possible explanation is that the methods used to determine the final ratings by various sentiment analysis programmes vary. Although TextBlob is a general-purpose sentiment analysis tool, Vader is used for the detection of negative sentiment. It follows that Vader is more prone to award high scores to tweets with negative emotion, regardless of the tweet's positivity. Combining Vader and TextBlob sentiment scores to forecast sentiment may further enhance the SVM classifier. Our knowledge of Vader's strength in negative emotions and TextBlob's strength in general sentiment make them an ideal pair. This SVM classifier ensemble, which also includes Vader and TextBlob, may therefore improve the model's accuracy.

## FUTURE WORK

A remarkable 76.25% accuracy has been attained using the Support Vector Machine (SVM). This research clarifies the use of Support Vector Machines (SVMs) for sentiment analysis. As new opportunities and paths emerge in the future, the project has a lot of room to grow and develop. Here are a few suggestions for strengthening and enhancing the sentiment analysis model. ## Adjusting the Model's Settings Additional investigation into the model parameters may be possible in future studies. The SVM model's performance might be enhanced by adjusting the parameters. • Workflow Design To collect more complicated sentiment data, it is helpful to use more sophisticated feature extraction algorithms.



Predicting the emotion by analysing each word is not the best approach. It's preferable to consider the larger picture or additional specifics in the text. The model may get a better grasp and provide more precise results if this is done. Combinatorial Analysis To improve performance and provide more accurate results, ensemble methods are seen to be the way to go. Therefore, SVM may be used in conjunction with other algorithms or models to boost the performance and efficiency of our model. Subjective Evaluation in Context Recent events or the environment might influence the tone of tweets in some instances. Therefore, the tone conveyed might be affected by this. Applying Contextual Sentimental Analysis will fix this. • Auxiliary Message Processing By including other information, such as visual elements like photographs or videos, sentiment prediction may be further enhanced. Insight into emotional expressions may be enhanced in this way. • Analysing Emotions in Real Time The model may be more useful in the actual world if we looked at its ability to do real-time sentiment analysis and get insights into shifting sentiment trends. • Methods with Humans Involved For more precise results, particularly in ambiguous contexts, it may be worthwhile to investigate person-in-the-loop approaches, where the model collaborates with human annotators. In conclusion, the proposed research directions aim to build upon the current study's foundational work by giving academics and practitioners actionable steps to improve sentiment analysis using SVM and to encourage the ongoing advancement of sentiment analysis methods.

### CONCLUSION

In this case, we employed Support Vector Machines (SVM) to do sentiment analysis for our study, and we got a 76% success rate. Along the way, we learned a lot about the nuances of sentiment expression in textual data via thorough data preparation, model training, and assessment. Our findings demonstrate that SVM may successfully detect sentiment patterns; nonetheless, the area might benefit from more research and development. This paper lays forth a plan for future research to strengthen and broaden the sentiment analysis model's use. Despite the useful outcomes produced by our SVM model, this effort is just the beginning of what will hopefully be a much more comprehensive and advanced strategy for sentiment analysis in the future.

### REFERENCES

- [1]. Koncz, P, Paralic, J “An approach to feature selection for sentiment analysis” In Proceedings of the 2011 15th IEEE International Conference on Intelligent Engineering Systems, Poprad, Slovakia, pp. 357–362 (23–25 June 2011)
- [2]. D. Maynard, K. Bontcheva “Understanding climate change tweets: an open source toolkit for social media analysis” (2015)
- [3]. P.M. Landwehr, W. Wei, M. Kowalchuck, K.M. Carley “Using tweets to support disaster planning, warning and response” *Saf. Sci.* 90 (2016) 33–47.
- [4]. S. Ruder, P. Ghaffari, J. Breslin “A Hierarchical Model of Reviews for Aspect-Based Sentiment Analysis” (2016).
- [5]. T. L. Milfont, M. S. Wilson, and C. G. Sibley “The public’s belief in climate change and its human cause are increasing over time” *Volume 12*, no. 3, p. e0174246 (2017).
- [6]. S. Yoo, J. Song, O. Jeong, “Social media contents based sentiment analysis and prediction system” *Expert Syst. Appl.* 105 (2018) 102– 111
- [7]. D. Zimbra, A. Abbasi, D. Zeng “The state-of-the-art in Twitter sentiment analysis: a review and benchmark evaluation” *ACM Trans. Manag. Inf. Syst.*, 9 (2) (2018), pp. 1-29
- [8]. Sathish Kumar “Topic modeling and sentiment analysis of global climate change tweets” (2019)
- [9]. H. Varshney, P. Shishodiya, and S. Sriramulu, "Classifying Tweets Based on Climate Change," *IEEE* (2022)
- [10]. N. S, D. Paulraj, P. Ezhumalai, M. Prakash “A deep learning modified neural network (dlmnn) based proficient sentiment analysis technique on Twitter data” *J. Exp. Theor. Artif.Intell* (2022)
- [11]. S. Das and S. Chakraborty, "Perception of United Nations Climate Change Conference in Social Networks," *IEEE* (2022)
- [12]. O. F. Ismail, H. Ismail, A. AlKhayat, and R. Aljohani, "Crowdsourcing, Sentiment Analysis and Content Analysis as an Ethnographic Reflection of Social Media Human/Climate Interaction," *IEEE* (2022)
- [13]. J. D. S. Baguio, B. A. Lu, and C. F. Peña, "Text Classification of Climate Change Tweets using Artificial Neural Networks, FastText Word Embeddings, and Latent Dirichlet Allocation," *IEEE* (2023)
- [14]. Emelie Rosenberg a, Carlota Tarazona b, Fermín Mallor c, Hamidreza Eivazi c, David Pastor-Escuredo d, Francesco Fuso-Nerini e f, Ricardo Vinuesa “Sentiment analysis on Twitter data towards climate action” *Volume 19*, September 2023.