Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 908–917

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model.

Samson Ebenezar Uthirapathy [a*] Domnic Sandanam [b]

[a]School of Computing Science, Galgotias University, Greater Noida, India, 201306
[a,b]Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India, 620015.

Abstract

Nowadays, Climate change is an important environmental factor that affects every living thing on the earth. It is very essential to study the public perceptions regarding the disaster events frequently happening due to climate change. In today's digital era individuals are using social network platforms namely Twitter, Facebook, and Weibo now and then to express their views about any events. In this paper, the climate change Twitter data set was considered for analyzing the topics and the opinions discussed by the public regarding climate change. The Latent Dirichlet Allocation(LDA) method was used to list out the various topics present in the data set and the Bidirectional Encoder Representation from Transformers(BERT uncased) is an efficient deep learning technique used to classify the sentiments present in the data set. Here the sentiments were labelled as pro news, support, neutral and anti. The performance of the proposed topic modelling and sentiment classification model was compared using the precision, recall, and accuracy measures. The BERT uncased model with has shown the best results such as precision of 91.35%, recall of 89.65%, and accuracy of 93.50% compared to other methods.

Keywords: *Climate change, Tweets, Topic modelling, Sentiment analysis, BERT, LDA.*

1. Introduction

Undoubtfully Climate change affects our planet and necessitates international cooperation. It has an impact on humans in terms of food, water, and health. Extreme weather alters wind patterns, raises average temperatures, and alters rainfall amounts [1]. Recently, the Antarctic region recorded the highest temperature, while the California wildfires and the Coronavirus pandemic have drawn attention to global warming and climate change[2]. Many studies contributed to an understanding of the magnitude of disasters caused by climate change [3]. People in today's digital world use social networks such as Facebook, Twitter, and Instagram, among others, to share their thoughts, feelings,

10.1016/j.procs.2023.01.071

and emotions about events that occur in and around them. Twitter, as a widely accessible medium, allows people to express their opinions on global warming and provides data on how public opinion shifts over time and place. [4]

Twitter is a popular online platform with approximately 229 million monetizable users and over 336 active users as of January 2022[5]. Twitter users have been actively discussing various events such as global warming, elections, climate change, and the coronavirus pandemic, among others. Millions of tweets are posted on Twitter every day, providing more data for analysis. As a result, tweets extracted from Twitter are thought to predict elections and analyse the impact of disasters caused by climate change, pandemics, wars, and so on. As a result, social media posts about extreme weather events range from individuals expressing their knowledge and opinions to support agencies in emergencies providing warnings, information, and updates on relief actions. [6]

Several recent studies use data science, natural language processing (NLP), data visualization and social network analysis techniques to develop effective and efficient methods for extracting important, required information from social network data[6]. The information discussed on social media cover several topics and expresses a diversity of opinions, with positive, negative, and neutral views on the topics or incidents. In some other cases, people may express their positive or negative feelings. As a result, topic modelling and sentiment analysis in text mining were combined to identify the topics discussed in the Twitter data set and analyse sentiment polarities. This analysis is used to understand the common man's perceptions regarding climate change and its impacts, Therefore the scientific community can concentrate on analysing the sources of climate change and recommend solutions to the governments to control the activities that cause climate change.

In this paper, a deep learning model was created using Latent Drichilate Analysis (LDA) and the Bidirectional Encoder Representation Technique (BERT). The LDA topic modelling technique is used to identify the topics in the data set, and the BERT model is used to categorise the data. The BERT deep learning Model supports both left-to-right and right-to-left sentence processing. BERT was trained on a large text corpus. Using the masked language prediction approach and the next sentence prediction approach, the model can better understand the language and learn inconsistencies in data patterns, as well as generalise well across a variety of NLP tasks.

The ensuing part of this paper is arranged as follows: related works in section 2, an explanation of the proposed model using LDA and BERT in section 3. the experimental results and discussion in section 4, and Section 5 summarise the paper's conclusion.

2. Related Work

The tweets in the Turkish language which contain messages about global warming have been collected and analyzed to determine the opinion of the users about climate change [6]. It classifies the tweets into various opinion categories like positive, negative and neutral. The proposed model has been built using machine learning techniques called k-nearest neighbour(K-NN), Naïve Bayes(NB) and support vector machine(SVM). The classifier's efficiency is compared using accuracy measures. Based on accuracy Naïve Bayes, SVM and K-NN have shown 65.43%, 73.51% and 74.63% respectively and there is scope for improvement of the classification accuracy.

A sentiment investigation framework using Bi-directional long short-term memory(BiLSTM) was developed [7] to analyze the sentiments of the tweets that caries information about climate change as well as worldwide heating. The sentiments are classified into various emotional categories namely discrimination, anger, joy and inspiration. The proposed classifier has given the state of art results using accuracy measures which range from 87.01 % to 89.80% for various opinion categories.

In [8] a topic modelling using Latent Dirichlet allocation(LDA) to list the topics which are present in the data set and the NRC(National Research Council of Cannada) emotion word lexicon association method used for sentiment analysis. A topic modelling was proposed [9] using LDA to identify the subjects discussed in tweets or Wikipedia.

A topic document sentence model was created [10] using joint sentiment topic(JST) and LDA. The TDS model was used to determine the polarity of the sentiments present in the product reviews data. VADER was used in [11] to analyse sentiments of the tweets related to COVID19 vaccines and LDA was used to identify topics.

The tweets were extracted from Twitter based on the locations of the Arctic and Alaska [12] to analyze public perception about the climate changes in that regions. The public opinions were analysed using the NRC emotion word lexicon association algorithms. The tweets-based opinion analysis using a fuzzy-based approach is discussed in [13]. The tweets are related to climate change and energy consumption in Alaska. A machine learning and deep neural network-based spam identification and sentiment analysis method for the Twitter data set were proposed in [14]. In this paper, the Naïve Bayes algorithms in machine learning and the LSTM method in deep learning have shown better performance than other methods in terms of accuracy measure.

An aspect-based sentiment extraction method for mobile phone reviews has been proposed [15] using artificial intelligence techniques namely random forest(RF), SVM and K-NN. The authors used the Twitter data set containing the tweets related to iPhone and Samsung phone reviews. In [16] LSTM and aspect-based opinion analysis using fuzzy logic for mobile phone reviews have been proposed. Sentiment analysis for movie reviews using convolutional neural networks(CNN) and LSTM has been developed in [17]. The proposed methods show 87.74% and 88.02% of accuracy than the other methods and there is a scope for performance improvement.

In [18] Bidirectional encoder representation from Transformers(BERT) and Dilated convolutional neural networks(DCNN) have been used to develop a sentiment analysis model. In this method, BERT was used for word embedding and DCNN layers were used for classification and fine-tuning the model. In [19] aspect-based sentiment analysis model was developed using BERT. The BERT model was used to mine the aspect level information from the large collection data to extract the aspect features for further classification and fine-tune the model. A BERT-CNN-based sentiment classification model was proposed [20] to classify the sentiments in a large movie review data set. In this paper, the author used the BERT technique for word embeddings and the convolutional neural network(CNN) to learn the semantics of information present in the text.

A BERT-based language model was proposed [21] for the Turkish language to pre-process the data to fit the machine learning algorithms. This BERT-based technique reduces the steps of data pre-processing. In [22] a BERT-based guided LDA model has been proposed for aspect-based sentiment analysis of the restaurant review data set. A sentiment analysis framework for social networks data was proposed [23] using machine learning techniques namely naïvebase and random forest to analyse the opinion present in the Twitter data set.

3. Proposed Model

The proposed model topic modelling and sentiment classification model includes the following phases: data preprocessing, topic modelling, and classification model. LDA (Latent Dirichlet Allocation) is used for topic modelling, and BERT is used for sentiment classification. Using the well-known topic modelling technique LDA, this work investigates the various climate change topics discussed by the public via tweets. The BERT uncased model is then used to categorise the sentiments present in the climate change tweets data into four different categories.

3.1 Data pre-processing

Pre-processing of data is a significant activity in any sentiment analysis and topic modelling framework. Because the raw data set contains more noisy and unwanted data, the performance of the topic models and sentiment classifiers will suffer. Tweets are short messages that include hashtags, emojis, text, and special symbols to express the user'spoint of view. As a result, the tweet data must be carefully preprocessed. Pattern removal, stop word removal, URL removal, tokenization, and vectorization are some of the preprocessing techniques used in this work.

Pattern removal: Symbols and special characters such as @,#, %, and so on are useless for topic modelling and sentiment analysis and should be removed from the data set.

Stopword removal: Stop words such as 'is,' 'was,' 'and,' 'or,' and so on have no meaning for sentiment analysis and must be removed from the data set.

URL removal: If the URLs in the data set reduce classification accuracy and provide no meaning, they must be removed.

Tokenization: The process of breaking down sentences into individual words or tokens. For instance, the sentence 'climate change is affecting people all over the world will be tokenized as follows. 'the','climate',' is affecting','the','lives',' in','the','world'.

Vectorization: Vectorization is the process of converting each word into a vector format. The documents are converted into matrix format during this process. The TF-IDTF vectorizer was used in this work. It generates a document word matrix, which is then fed into the LDA topic modelling. The proposed topic modelling and sentiment classification model architecture is depicted in Fig. 1.

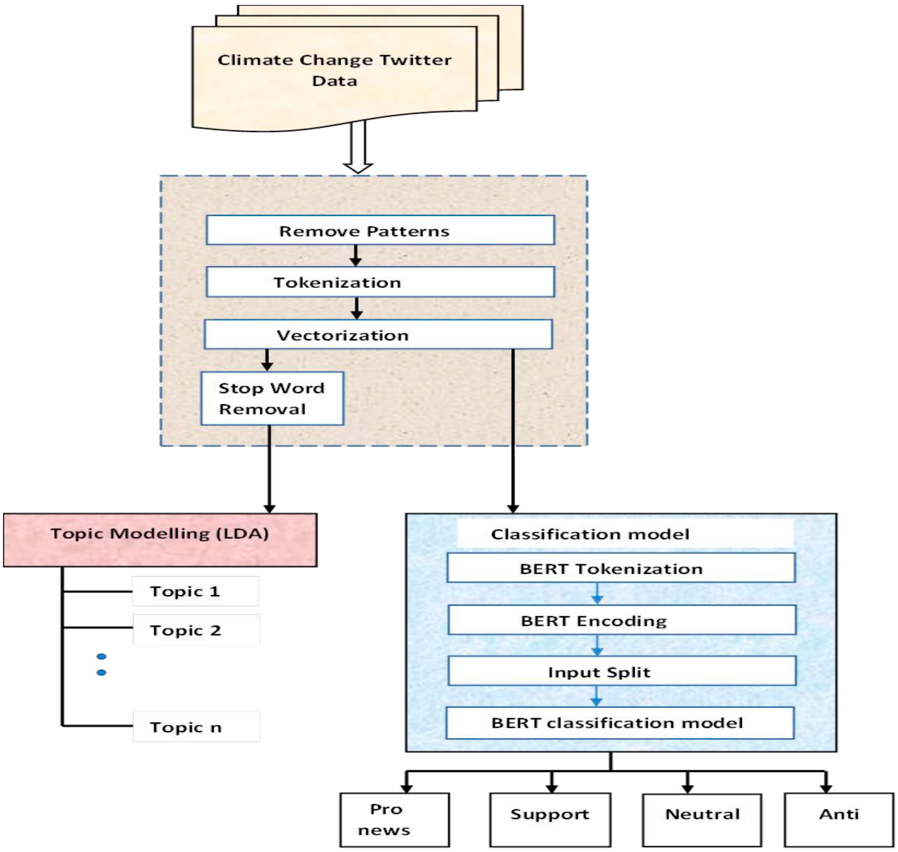

Fig. 1 The Proposed topic modelling and sentiment analysis model architecture and workflow

3.2 LDA topic modelling:

The Latent Dirichlet allocation[8-11] is a generative probabilistic model that identifies the topics present in the documents and then maps relevant words or tokens to the topics. The term "topics" refers to the unknown variable relationships that exist between words in a lexicon and their positions in documents. The collection of various topics comprises a document. LDA searches for hidden themes in the collection and interprets documents with those themes.

The LDA is fed the document word matrix generated during the data pre-processing phase. Then, LDA generates two distinct matrices: the document-topic matrix and the topic word matrix. Each tweet is treated as a document in this work. Then, given a set of documents, a topic-specific probability distribution is computed, with each word assigned to one of the topics. It determines the number of topics covered in a text corpus and which topics are primarily discussed. The LDA topic modelling is depicted graphically in Fig. 2.

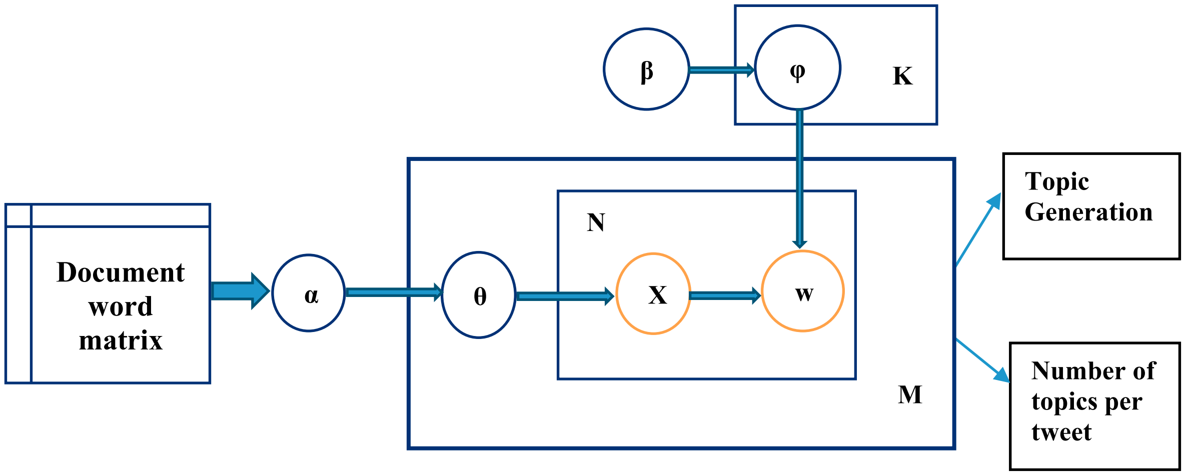

Fig. 2 LDA Topic Modeling

The hidden topics and observed word distribution of the LDA generative process can be mathematically represented using the equation (1).

$$p(\phi_{1:k}, \theta_{1:M}, X_{1:M}, w_{1:M}) = \prod_{i=1}^{k} p(\phi_i) \prod_{m=1}^{M} p(\theta_m) \left(\prod_{n=1}^{N} p((X_{m,n}|\theta_m))\, p((w_{m,n}|\phi_{1:k}, x_{m,n}))\right) \quad (1)$$

The LDA model operates on the assumptions as follows, the corpus D contains k topics and documents M, and the document has N words. The notations α & β are the Dirichlet hyperparameters for topic distribution θ in documents and word distribution φ in topics, where $p(\phi_{k)}$ is the distribution of words for topic k, p($\theta_{m)}$ is the distribution of topics for m^{th} document and $x_{m,n}$ is the topic assigned to the word $w_{m,n}$, here $w_{m,n}$ is m^{th} $word\ in\ n^{th}\ document.$ In equation (1) x and w represent the topics and words. As per equation(1) for every topic i, $p(\phi_{i)}$ is determined then for every document m, $p(\theta_m)$ is determined then a topic x is selected by the topic distribution p(θ_m) from each word position. Finally, a word w is selected by the distribution $p(\phi_k)$. Finally, the LDA maps the topics and their respective documents. Using this the most frequent topics present in the corpus can be listed.

3.3 BERT classification Model:

The Bidirectional Encoder Representation from Transformers (BERT) [19-23] is a deep learning method for natural language processing. Normally, language models read input sequences from left -to right otherwise right -to left directions, but BERT reads input sequences from both directions. The BERT is powered by a Transformer and employs an attention mechanism that studies appropriate associations between words in a sentence. An encoder is included in a Transformer. The encoder is used to read the input texts. BERT is used as a sentiment classification model in this paper. The encoder receives input in the form of a token sequence. The inputs are converted into vectors called word embeddings before being passed to the encoder. Here, word embedding is a combination of three different embedding schemes: token, segment, and positional embeddings.

Token embeddings: Token embeddings insert two special tokens [CLS] and [SEP] into the input word tokens. [CLS] is a classification token that comes before the first word in the sentence, and [SEP] comes at the end. [SEP] distinguishes the end of the first sentence from the start of the next sentence in the same input sequence.

Segment embeddings: A marker is added to each token in segment embeddings to allow the encoder to differentiate the sentences. Each token is given a positional embedding to determine its place in the sentence.

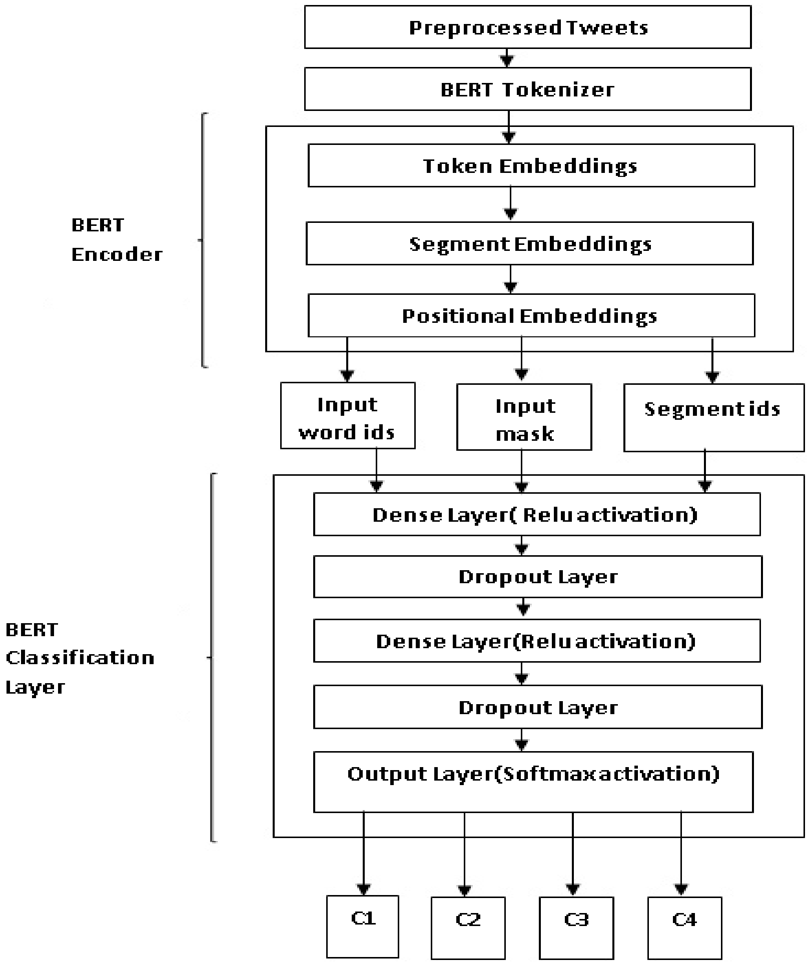

Fig. 3. BERT Classification Model

The proposed BERT model has three stages of processing. After pre-processing, the tweets are tokenized with the BERT tokenizer and sent to the BERT encoder, which has three encoding phases: token embeddings, segment embeddings, and positional embeddings. Vectorized features such as input word ids, input masks, and segment ids are output by the encoder. The BERT predicts the masked tokens using the other tokens. As a result, it is known as a masked language model[15]. From the entire input, BERT predicts only the masked token. The feature vectors are then forwarded to the BERT classification layer.

The classification layer is made up of two dense layers that are activated by the Rectified linear unit (Relu). The dense layer consists of a fully connected neural network, two drop-out layers, and one output layer activated with softmax. The activation function determines the output of each dense layer in this case. The activation function adds nonlinearity to each neuron's output in the neural network. In equation (2), Relu's mathematical expression is mentioned

$$f(x) = \max(0,x) \tag{2}$$

In equation (2), x is the input vector, and the Relu produces output when the input vector is greater than zero. The softmax is used in the classification output layer for categorical outputs in multiclass classification. The softmax is represented by equation (3).

$$f(\bar{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}} \tag{3}$$

In equation (3) $\bar{x}$ is the input vector to the softmax function, x_i is an element of the input vector i. The denominator in equation(3) is the normalization term and k is the number of classes. In this work, the tweets are classified into four different classes: news, support, neutral and anti.
The sentiment classification model using BERT is depicted in Fig. 3. It categorises tweets into four categories. In Fig. 3, c1,c2,c3, and c4 represent the four output classes, which are news, support, neutral, and anti, respectively.

4. Result and discussion:

For the experiments and discussion in this paper, a Twitter data set containing tweets about climate change was used. The data set was obtained from the Kaggle repository[24], and Tweets were gathered between 27 April 2015, and 21 February 2018. Since the climate change tweets were considered, the majority of people support or oppose the actions taken by governments or local governments to control global warming. The tweets in the data set were classified into four sentiment classes. Tweets labelled as News contain accurate information about climate change. Tweets labelled as support include tweets that support beliefs about man-made climate change. Those tweets labelled as neutral neither support nor oppose man-made climate change. The anti-climate-change tweets express their opposition to man-made climate change. There are 43944 tweets in the climate change data set. This includes 9277 tweets labelled as News, 22963 tweets labelled as Support, 7716 tweets labelled as Neutral, and 3991 tweets labelled as Anti. The experiments were carried out in Python.

4.1 Most frequent topics in the data set:

Using LDA topic modelling, we identified the most frequently occurring topics in the climate change data set. Table 1 displays some of the frequently discussed topics in various categories.

Table 1 lists the sample topics discussed in climate change tweets in various categories such as News, Support, Neutral, and Anti. According to the number of tweets in each category, more people are becoming aware of and debating global warming and climate change. People want the government to take collective action to control global warming, and the majority of them agree that it is a man-made phenomenon. Only a few people deny that climate change is a result of human activity.

Table 1 Climate change topics in various sentiment categories

Categories	Topics
News	weather Mediterranean to become desert unless global warming limited
News	years of Living Dangerously wants to make climate change a voting issue
Support	how climate change affects sea-level rise (and flood cities) as the world warms
Support	It's vital that the public health community addresses climate change
Neutral	is this climate change or just weather?
Neutral	Thank you global warming for giving us nice weather in November
Anti	you do understand that climate change is natural, not necessarily caused by humans. It snowed in Miami, and the Atlantic froze
Anti	Too bad this normally intelligent man believes that the world is ending due to global warming which is the fault…

4.2 Sentiment classification model performance:

The classification accuracy of the proposed topic modelling and sentiment analysis model was compared with existing classification approaches namely K-nearest neighbour(KNN)[6], Support vector machine(SVM)[6] and the deep learning methods BiLSTM[7] and LSTM-CNN[17]. The precision and recall along with accuracy measures were considered for the performance comparison.

Precision: It is the ratio of true positive to total predicted positive values. It can be represented mathematically using equation(4)

$$\text{Precision} = \frac{True\_Positive}{True\_Positive + False\_Positive} \tag{4}$$

Recall: It is the proportion of true positive to predicted positive values. It can be represented mathematically using equation(5).

$$Recall = \frac{True\_Positive}{True\_Positive + False\_negative} \tag{5}$$

Accuracy: It is the ratio of properly predicted values to the number of values present in the document. It is mentioned in equation(6).

$$Accuracy = \frac{True\_Positive + True\_Negative}{True\_Positive + True\_Negative + False\_Positive + False\_Negative} \tag{6}$$

Parameter settings of the BERT base uncased model are shown in Table 2.

Table 2 Parameter Settings

Parameters	Values
Nodes in dense Layer one	128
Nodes in dense Layer two	64
Optimizer	Adam
Dropout	0.2
Loss function	Categorical cross-entropy
Activation Functions	Relu, Softmax(output layer)
Learning rate	2e-5
Batch size	64
No of Epochs	7

To improve performance, we used the k-fold cross validation [25] approach to fine-tune our proposed BERT classification model. We used k=10 in this case. The data set is divided into k groups in the k-fold cross validation strategy. Then, among the k groups, a single group g is chosen as the test data, whereas the remaining k-1 groups are used as the training data. These steps will be repeated for each distinct group. Finally, the mean of all the results is used to calculate the final performance value. Table 3 compares the classification efficiency of the proposed BERT classification model to that of other methods.

Table 3 Classification performance comparative statements

Methods	Precision	Recall	Accuracy
SVM	74.25	72.35	76.34
K-NN	76.35	71.25	78.60
BiLSTM	83.35	80.40	84.50
LSTM+CNN	85.50	82.00	86.45
BERT(uncased) (10-fold cross validation)	**91.35**	**89.65**	**93.50**

According to the results in table 3, the proposed BERT-based sentiment classification model outperforms the other methods in terms of precision with a 6% improvement, recall with a 7% improvement and accuracy with a 7% improvement.

5. Conclusion:

Social networks such as Twitter enable users to share their perspectives on events and incidents that occur in and around them. Climate change and global warming are significant global phenomena. It is critical to examine people's perspectives on climate change events. In this paper, the LDA method was used to identify the most frequently discussed topics in the data set, and a BERT-based sentiment classification model was proposed to categorise the tweets into four different categories: News, Support, Neutral, and Anti. This analysis is used to understand the general public's perceptions of climate change and its consequences. As a result, the scientific community can focus on analysing the root causes of climate change and recommending solutions to governments to control the activities that cause climate change.

The proposed topic and sentiment analysis model using LDA and BERT outperforms the existing approaches in terms of precision (91.35 %), recall (89.65 %), and accuracy (93.50 %). In the future, we plan to concentrate our efforts on identifying extreme viewpoints that can serve as precursors to ferocious behaviour. There is some interest on social media platforms in the real-time vigilance of extreme opinions and the forecast of radicalization.

References

[1] Klingelhöfer, D., Müller, R., Braun, M., Brüggmann, D., & Groneberg, D. A. (2020). Climate change: Does international research fulfill global demands and necessities? In Environmental Sciences Europe (Vol. 32, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/s12302-020-00419-1

[2] Kulp, S. A., & Strauss, B. H. (2019). Author Correction: New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. In Nature Communications (Vol. 10, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41467-019-13552-0

[3] Jang, S. M., & Hart, P. S. (2015). Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. In Global Environmental Change (Vol. 32, pp. 11–17). Elsevier BV. https://doi.org/10.1016/j.gloenvcha.2015.02.010

[4] Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll. In S. Lehmann (Ed.), PLOS ONE (Vol. 10, Issue 8, p. e0136092). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0136092

[5] https://www.statista.com/statistics/303681/twitter-users-worldwide

[6] Kirelli, Y., & Arslankaya, S. (2020). Sentiment Analysis of Shared Tweets on Global Warming on Twitter with Data Mining Methods: A Case Study on Turkish Language. In Computational Intelligence and Neuroscience (Vol. 2020, pp. 1–9). Hindawi Limited. https://doi.org/10.1155/2020/1904172

[7] El Barachi, M., AlKhatib, M., Mathew, S., & Oroumchian, F. (2021). A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. In Journal of Cleaner Production (Vol. 312, p. 127820). Elsevier BV. https://doi.org/10.1016/j.jclepro.2021.127820

[8] Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. In Social Network Analysis and Mining (Vol. 9, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1007/s13278-019-0568-8

[9] Tong, Z., & Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. In Computer Science & Information Technology (CS & IT). The Sixth International Conference on Computer Science, Engineering and Information Technology. Academy & Industry Research Collaboration Center (AIRCC). https://doi.org/10.5121/csit.2016.60616

[10] Farkhod, A., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-Based Topic Modeling Sentiment Analysis Using Topic Document Sentence (TDS) Model. In Applied Sciences (Vol. 11, Issue 23, p. 11091). MDPI AG. https://doi.org/10.3390/app112311091

[11] Yin, H., Song, X., Yang, S., & Li, J. (2022). Sentiment analysis and topic modeling for COVID-19 vaccine discussions. In World Wide Web (Vol. 25, Issue 3, pp. 1067–1083). Springer Science and Business Media LLC. https://doi.org/10.1007/s11280-022-01029-y

[12] Bergstedt, H., Ristea, A., Resch, B., & Bartsch, A. (2018). Public Perception of Climate Change in Alaska: A Case Study of Opinion-Mining using Twitter. In GI_Forum (Vol. 1, pp. 47–64). Osterreichische Akademie der Wissenschaften. https://doi.org/10.1553/giscience2018_01_s4

[13] Abdar, M., Basiri, M. E., Yin, J., Habibnezhad, M., Chi, G., Nemati, S., & Asadi, S. (2020). Energy choices in Alaska: Mining people's perception and attitudes from geotagged tweets. In Renewable and Sustainable Energy Reviews (Vol. 124, p. 109781). Elsevier BV. https://doi.org/10.1016/j.rser.2020.109781

[14] Rodrigues, A. P., Fernandes, R., A, A., B, A., Shetty, A., K, A., Lakshmanna, K., & Shafi, R. M. (2022). Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques. In M. Ahmad (Ed.), Computational Intelligence and Neuroscience (Vol. 2022, pp. 1–14). Hindawi Limited. https://doi.org/10.1155/2022/5211949

[15] Naramula, V., & A., K. (2021). Sentiment analysis in aspect term extraction for mobile phone tweets using machine learning techniques. In International Journal of Pervasive Computing and Communications. Emerald. https://doi.org/10.1108/ijpcc-06-2021-0143

[16] Sivakumar, M., & Uyyala, S. R. (2021). Aspect-based sentiment analysis of mobile phone reviews using LSTM and fuzzy logic. In International Journal of Data Science and Analytics (Vol. 12, Issue 4, pp. 355–367). Springer Science and Business Media LLC. https://doi.org/10.1007/s41060-021-00277-x

[17] Gandhi, U. D., Malarvizhi Kumar, P., Chandra Babu, G., & Karthick, G. (2021). Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). In Wireless Personal Communications. Springer Science and Business Media LLC. https://doi.org/10.1007/s11277-021-08580-3

[18] Jain, P. K., Quamer, W., Saravanan, V., & Pamula, R. (2022). Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis. In Journal of Ambient Intelligence and Humanized Computing. Springer Science and Business Media LLC. https://doi.org/10.1007/s12652-022-03698-z.

[19] Pang, G., Lu, K., Zhu, X., He, J., Mo, Z., Peng, Z., & Pu, B. (2021). Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model. In Z. Duan (Ed.), Wireless Communications and Mobile Computing (Vol. 2021, pp. 1–13). Hindawi Limited. https://doi.org/10.1155/2021/5534615

[20] Huang, P., Zhu, H., Zheng, L., & Wang, Y. (2021). Text sentiment analysis based on Bert and Convolutional Neural Networks. *2021 5th International Conference on Natural Language Processing and Information Retrieval (NLPIR)*. https://doi.org/10.1145/3508230.3508231

[21] Özçift, A., Akarsu, K., Yumuk, F., & Söylemez, C. (2021). Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish. In Automatika (Vol. 62, Issue 2, pp. 226–238). Informa UK Limited. https://doi.org/10.1080/00051144.2021.1922150

[22] Venugopalan, M., & Gupta, D. (2022). An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. In Knowledge-Based Systems (Vol. 246, p. 108668). Elsevier BV. https://doi.org/10.1016/j.knosys.2022.108668

[23] Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment Analysis of Twitter Data. In International Journal of Healthcare Information Systems and Informatics (Vol. 14, Issue 2, pp. 1–16). IGI Global. https://doi.org/10.4018/ijhisi.2019040101

[24] https://www.kaggle.com/datasets/edqian/twitter-climate-change-sentiment-dataset

[25] Pal, K., & Patel, Biraj. V. (2020). Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE. https://doi.org/10.1109/iccmc48092.2020.iccmc-00016