Journal of Informatics and Web Engineering

Vol. 4 No. 2 (June 2025) eISSN: 2821-370X

Aspects-Based Sentiment Analysis of Extreme Weather on Twitter Using Long Short-Term Memory

Tursun Abdurahmonov[1], Muhammad Nabil Toby Abiyyu[2], Dzikru Nur Khayat[3], M. Ary Heryanto[4*]

[1]Faculty of Computer Technology and Computer Science, Termez State University, Termez City, Barkamol Avlod Street, 43, Uzbekistan

[2,3]Faculty of Computer Science, Dian Nuswantoro University, Kota Semarang, Jawa Tengah 50131, Indonesia

[4]Faculty of Engineering, Dian Nuswantoro University, Kota Semarang, Jawa Tengah 50131, Indonesia

**Corresponding author: (m.aryheryanto@dsn.dinus.ac.id; ORCiD: 0009-0005-4228-3672)*

Abstract - This study presents an aspect-based sentiment analysis of tweets related to extreme weather events in Indonesia, utilizing the Long Short-Term Memory (LSTM) model. The dataset was obtained through a Twitter crawling process, followed by a series of preprocessing steps including data cleaning, stop word removal, normalization, tokenization, and stemming. The three primary areas of emphasis in the study were kinds of bad weather forecasts, and the government or society reactions. Using a lexicon-based technique, sentiment labelling generated three groups: positive, neutral, and negative. A random oversampling method was employed to address the data imbalance. The model using the LSTM algorithm was trained individually for aspect and sentiment classification tasks, so reaching high accuracies of 98.94% and 97.53%, respectively. The results indicate that the model effectively categorises talk on extreme weather and the opinions of the public. A word cloud visual representation was additionally created to show frequently occurring terms in the dataset, thereby offering insights into current themes and sentiment expressions. This work provides valuable input for government agencies and legislators in developing communication and disaster response plans, thereby serving to better understand the public's view on climate-related events. Future work could involve improving techniques for preprocessing and using larger, wider-ranging datasets for improving the model's robustness and generalisation.

Keywords— Aspect Based, Sentiment Analysis, Extreme Weather, X(Twitter), Long Short-Term Memory

Received: 23 December 2024; Accepted: 7 May 2025; Published: 16 June 2025

1. INTRODUCTION

Many regions regularly experience extreme weather events such as floods, landslides, droughts, and tornadoes, which pose significant threats to public safety and infrastructure. These events frequently lead to substantial economic losses, adverse medical consequences, and environmental harm. The frequent and continuing character of such incidents highlights the need of successful prevention and response procedures. Extreme weather's public perception and responses are essential in directing such efforts. Reviewing the ways individuals perceive and respond to such events could assist disaster management institutions, interaction strategists, and lawmakers [1] in their work. Extreme

MULTIMEDIA UNIVERSITY

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2025.4.2.28

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

weather is atmospheric conditions that vary substantially from previous records and climatological averages. Frequently relatively rare, exceptionally strong, and occasionally linked with significant implications for the natural environment, facilities and human well-being, these types of events typically remain rare. "Extreme" indicates circumstances outside of the anticipated range of variability such as extremely powerful storms, prolonged periods of drought, excessive rainfall, or humidity anomalies. In meteorology, extreme weather is defined as irregularities that far exceed normal boundaries and have an opportunity to cause damage, disruptive, or risk both individuals and ecosystems. Evaluation of risk, preparation, and policy formulation [2] rely on having an understanding and the identification of such irregularities.

In the present day, using communication channels such as Twitter provides an excellent opportunity for academia to collect enormous quantities of data reflecting the views of individuals, points of view, and responses to extreme weather. Assessing how people respond to extreme weather by obtaining sentiment information via Twitter has become an essential research priority. The artificial neural network technique, particularly LSTM [3] is a means to investigate sentiment in text within Twitter. Unlike the formal vocabulary frequently employed in the news, Twitter lets users pour out their hearts in more casual or every day spoken language, thereby it has been chosen as the primary source of information for this study. Twitter's data is real-time as well as reflects people's genuine reactions to hazardous weather. Hashtags and trending subjects aid quickly recognise important tweets; the Twitter API's straightforward access to information allows rapid gathering of information [4].

LSTM is an instance of Recurrent Neural Network (RNN). LSTM can also gather contextual information for classification purposes as well. For controlling the current condition of both the interval and output at any given point, the LSTM network consists of three gates: input gate, output gate, and forget gate. By ways of determining eliminated data, this improves the lengthy sequence memory of the LSTM networks and allows them to deal with long-term dependencies [5]. Employing the LSTM, this research states that it is frequently recommended in problems with classification and reaches an excellent accuracy rate of 95.38% [6]; LSTM is higher than Neural Network and Convolutional Neural Network (CNN) [7].

Various datasets have already been utilised for LSTM-based sentiment analysis research. Employing the LSTM method, earlier studies concentrated primarily on Twitter-related fire emotions. Employing the Twitter API, the study data was gathered to generate 7,000 tweets on forest fires. These tweets were classified as either "positive" or "negative". The data went through several processing stages: preprocessing, dividing into training and testing sets (80:20, 70:30, and 90:10), extraction of features using TF-IDF, and feature augmentation with GloVe and FastText. A technique of weighting words in documents, TF-IDF considers both how often they occur in one document and how common they are in an entire corpus. The results showed that the LSTM model combined with FastText feature expansion from Common Crawl achieved an accuracy level of 80.59%, slightly better than the GloVe expansion, that reached 80.13%. FastText feature expansion has been demonstrated in this study to have been more successful in increasing the accuracy of analysis of sentiment in forest fire events [8].

Aiming to categorise emotions expressed through Twitter posts into three groups, positive, negative, and neutral, this paper is focused on sentiment analysis. Comprehensive data preprocessing, text cleaning, removal of extraneous characters, hashtag standardisation, and tokenisation to maintain contextual integrity is part of this analysis. The LSTM network, a type of recurrent neural network appropriate for capturing sequential dependencies and contextual nuance in text, is the main model used. LSTM has been selected for its demonstrated effectiveness for handling the brief, informal, and context-sensitive character of Twitter data. Results from experiments indicate that the suggested approach outperforms baseline models in sentiment task classification on social media data [9] with an F1-Score of 0.93 and an accuracy of 93%.

Another study examines at sentiment analysis as an approach of classifying and analysing a huge volume of unstructured text data generated on social media. This paper attempts to improve the accuracy of existing sentiment analysis models through the integration of several artificial intelligence (AI) and LSTM technologies. This approach also considers the using essential symbols like emojis and special characters to more carefully capture emotional subtleties in text. Though this abstract does not have specific quantitative findings, this work aims to make sentiment analysis algorithms more relevant and practical to employ in the current data-driven the community [10]. The focus is on classifying general sentiment from unstructured data on social media such as Twitter based on previous studies addressing sentiment analysis using the LSTM method. This work takes a more particular approach. Apart from categorising sentiment in general, this work investigates sentiment depending on three primary variables: types of extreme weather, weather forecasts, and government or community responses. This aspect-based method allows the study to find public opinion on specific elements of extreme weather events in the context of augmented intelligence implementation.

Augmented intelligence is a human-focused approach to AI that enhances human intelligence rather than replacing it. AI trains computers to learn human behaviour, such as studying, assessment, and decision-making, and uses machines to mimic intelligent human behaviour [11]. Via the LSTM computation, artificial intelligence plays an essential part in analysing and interpreting vast amounts of Twitter data to identify open conclusions about various angles of unusual weather, such as its effects on community activities and well-being, within the framework of aspect-based estimation research of exceptional weather in Indonesia. Be that as it may, Increased Insights improves this AI-driven study by stressing the partnership between human judgement and machine learning.

2. LITERATURE REVIEW

Several studies have conducted sentiment analysis on social media using machine learning techniques. [12] examined public sentiment towards government regulation on Twitter, using the Naïve Bayes and LSTM methods. The dataset was collected using the Twitter API with keywords such as *"permendikbud30"* (Government Regulation No. 30) and *"kekerasan seksual di kampus"* (sexual harassment on campus), yielding 2,765 tweets. The dataset has been reduced to 471 useful records after preprocessing; those were subsequently labelled and assigned TF-IDF-based feature weighting. Results demonstrated that the LSTM method superior to Naïve Bayes, reaching 84% accuracy, 75% recall, as well as an F1-Score of 80%. By contrast, the Naïve Bayes model achieved 75% accuracy, 75% recall, and an F1-Score of 75%. Though LSTM implemented better, the authors highlighted at the fact that for the best performance it requires greater processing power and greater datasets. [13] further investigations tackled the Support Vector Machine (SVM) algorithm's assessment of public opinion sentiment on variations in the weather. Data from Twitter collected via API using phrases like "Weather Change" was employed in this study. Preprocessing procedures including cleaning, case folding, tokenising, stop word removal, stemming, and labelling processed the initial information of 850 tweets and produced 806 data ready for use. The TF-IDF method of classification followed by transforming the data into vectors. The SVM model's results of assessment indicated 70% accuracy, 39% precision, 39% recall, and 37% F1-Score. The results showed that the model was better at identifying negative sentiment compared to positive or neutral sentiment. Although there currently has been potential for improvement to the data and model, this study discovered SVM to perform reasonably well in sentiment categorisation.

The following study by [14] tackles optimising the performance of sentiment analysis on useful and uninformative tweets from the Twitter account @infoBMKG by applying different machine learning algorithms: Naïve Bayes, Naïve Bayes + Adaboost, SVM, and SVM with Particle Swarm Optimisation (PSO). This study processes 1,000 Indonesian tweets collected by crawling employing the Gata Framework preliminary processing framework. RapidMiner has been employed to test the data and generate evaluation measures including accuracy, precision, recall, and AUC. The results suggested that the technique using SVM surpassed others with an accuracy of 79.25%, a recall of 89.38%, and an AUC of 0.845. SVM is shown in this work to be better in sentiment classification and to be able to precisely examine tweets.

[15] carried out additional study on the three main approaches—lexicon-based, machine learning, and hybrid methods related to climate change. The study intended to determine the way each method evaluates climate change-related tweet sentiment. The lexicon approach uses tools such as VADER, TextBlob, and SentiWordNet. On the opposite hand, the machine learning method uses algorithms including SVM, Naïve Bayes, and Logistic Regression with Bag-of-Words (BoW) and TF-IDF techniques to extract features. Results demonstrate that the hybrid approach combining lexicon and machine learning, yields the most excellent performance with the highest F1-Score of 75.3% from the use of TextBlob and Logistic Regression. Lemmatisation was demonstrated in the study to improve accuracy by 1.6%; TF-IDF surpassed BoW in boosting Logistic Regression performance.

[16] discusses machine learning-based sentiment analysis on tweets associated with Saudi tourism to help with the development of the hospitality industry in line with Saudi Vision 2030. Using a dataset of Arabic and English tweets referred to as FDTST (First Dataset for Tweets Saudi Tourism), this study establishes and evaluates multiple classification models including SVM and Naïve Bayes. This approach improves accuracy through using feature extraction and selection, data pre-processing such as stemming and lemmatisation, and data collecting from relevant accounts and hashtags. The results of the experiment show that, with an accuracy of 92.2% for Arabic tweets and

90.1% for English tweets, the SVM model with RBF kernel surpasses the NB model. Based on insights produced by sentiment analysis, this paper offers suggestions to improve tourism in Saudi Arabia.

Several studies have demonstrated the effectiveness of applying methods based on machine learning, including LSTM, in sentiment analysis throughout a variety of subjects as well as Artificial Intelligence (AI). Earlier research has concentrated on wide analysis of sentiment. On the contrary this paper shows an aspect-based method for examining specific feelings about the impacts of extreme weather conditions, such health or transportation. In line with the benefits of LSTM in capturing temporal setting relationships while offering deeper analysis results, this approach enhances comprehensive understanding of extreme weather issues.

AI plays an important role in the study of aspect-based sentiment analysis of Indonesia's Extreme Weather on Twitter using LSTM. AI enables efficient processing of large amounts of Twitter data to understand public opinion on the impact of extreme weather. With LSTM capabilities, AI can deeply analyse the interactions between terms in the text, identify sentiment patterns, and group data from certain aspects, such as the impact on transportation, health, or economic activities. The application of AI in this study increases the accuracy of sentiment analysis. AI provides relevant insights for policy makers to design more effective responses to the challenges of extreme weather in Indonesia.

3. RESEARCH METHODOLOGY

In this study, the author uses a data collection method by taking tweets from users on Twitter related to extreme weather in Indonesia with the keywords "hot," "flood," "evacuation," and "baking." The data was obtained through crawling techniques on the Twitter platform, specifically involving tweets submitted by tweets from users on Twitter. Documents collected from tweets X will be a test data source, which will then be classified based on specific aspects using the LSTM method.

3.1 Proposed Method

As shown in Figure 1, the classification workflow using LSTM begins with tweet data collection, followed by preprocessing, and culminates in result analysis.

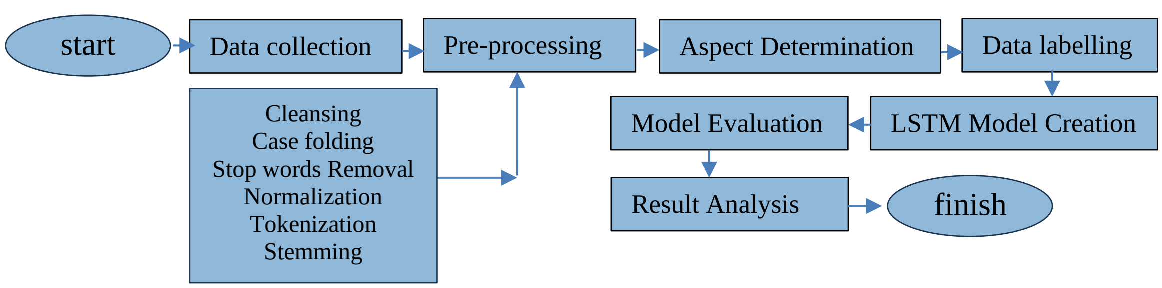

Figure 1. Conceptual Framework of Method

It illustrates the workflow for conducting sentiment analysis using an LSTM model. The process begins with data collection, where tweets are gathered as raw input. Next, pre-processing is performed to clean and prepare the textual data. The primary steps in this procedure are cleansing, case folding, stop words removal, normalisation, tokenisation, and stemming, each of which are essential for ensuring data quality and model readiness. The following step after pre-processing the text is aspect identification, in which specific topics or subjects, aspects that inside the text are identified. The dataset subsequently undergoes data labelling, whereby every text sample becomes a sentiment label usually positive, negative, or neutral. Utilising the labelled data, the learning model for LSTM is trained and created; it can acquire sequential dependencies in text. The trained model is evaluated resulting from training to assess its accuracy, precision, recall, and F1-Score. The results are then investigated to comprehend the performance of the model and to obtain important insights. Finally, the workflow concludes with the result of the analysis, which might assist with shape decisions or further research and thus finish the sentiment analysis phase.

3.2. LSTM Model Architecture

LSTM network often yields good results in the sequential analysis of a long text [17]. The LSTM architecture employed in this study depicted in Figure 2 consists of an embedding layer, bidirectional LSTM, and dual classification outputs (aspect and sentiment). First, sentences or words from tweets are taken and converted into vectors using the Embedding layer. Each vector is then fed to an LSTM cell, which processes sequential information and captures long-term dependencies in the data. The LSTM cells are connected to a hidden layer, where important features for sentiment analysis are extracted. The output layer provides the sentiment analysis results based on these features.

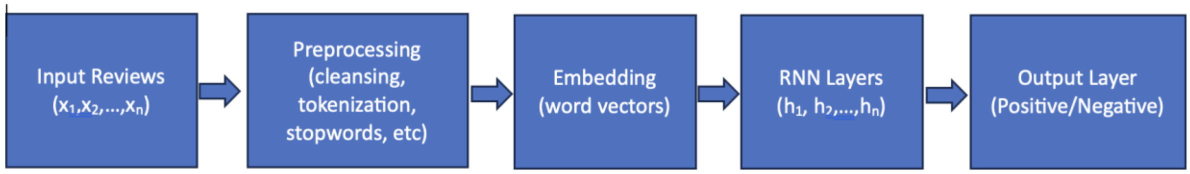

Figure 2. LSTM Model Architecture

It enhanced LSTM architecture for joint aspect and sentiment classification. The model processes raw tweets through preprocessing, embedding, and LSTM layers, culminating in dual softmax outputs for aspect and sentiment prediction. Design highlights bidirectional LSTM and dropout for robust feature learning. In this architecture, the pre-processed tweets are vectorized using an embedding layer. These vectors are passed to an LSTM layer to capture sequence dependencies. The output is passed to a dense layer with a ReLU activation function, and finally to a softmax classifier that outputs either the aspect class or sentiment label. This architecture is duplicated and fine-tuned separately for aspect and sentiment classification tasks.

3.3 Evaluate Model

The next phase of the research involves predicting sentiment on test data, generating classification reports, and evaluating model performance using a confusion matrix. Key evaluation metrics include accuracy, precision, recall, and F1-Score, which collectively measure the model's effectiveness in aspect-based sentiment classification. Precision indicates the proportion of correctly identified positive cases among all predicted positives, while recall reflects the proportion of actual positive cases accurately identified by the model. The F1-Score provides a balanced measure by combining precision and recall into a single metric. Although this study focuses exclusively on the LSTM model, previous research has also employed alternative algorithms such as K-Nearest Neighbors (KNN) [18]. The confusion matrix is a tabular representation used to evaluate classification models, offering a detailed summary of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It helps assess not only overall accuracy but also how well the model distinguishes between sentiment classes.

3.3.1 Accuracy

Accuracy measures the proportion of correct predictions made by a model across all prediction classes. It is calculated by dividing the sum of true positives and true negatives by the total number of predictions, computed using Equation (1).

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \tag{1}$$

3.3.2 Precision

Precision is a metric that indicates how many of the instances predicted as positive by the model are actually correct. It is calculated using Equation (2).

$$\text{Precision} = TP / (TP + FP) \tag{2}$$

3.3.3. Recall

Recall measures the model's ability to correctly identify all actual positive cases in the dataset. It is calculated as depicted in Equation (3).

$$\text{Recall} = TP / (TP + FN) \tag{3}$$

3.3.4 F1-Score

A combined measure that considers both precision and recall (see Equation (4)).

$$\text{F1-Score} = 2\,(\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \tag{4}$$

3.4 Analysis of Results

At this stage, the research highlights evaluating the performance of the sentiment and aspect classification framework on newly produced tweets. The trained LSTM model is employed to predict each tweet's aspect category along with associated sentiment polarity. This method enables the identification of topics and responses to emotion expressed in public discourse [19], [20], [21]. The study suggests to better comprehend how people perceive and respond to severe weather events by using the comprehension of the model's predictions. The resulting research provides valuable information that could potentially use by disaster management agencies, political leaders, and communicators to create more targeted and effective strategies. These results help boost public engagement and preparedness in the face of environmental issues by promoting data-informed decision-making.

4. RESULTS AND DISCUSSIONS

4.1 Data Collecting

This study uses data obtained through the crawling process from X related to extreme weather in Indonesia. The crawling was taken using the tweet-harvest library using Node.js. Furthermore, tweet searches were carried out with keywords such as Panas, Banjir, Evakuasi, and BMKG. The number of tweets taken can be adjusted according to needs, here each keyword is taken 400 data. This process resulted in 1624 tweets that took Indonesian-language tweets from October 1, 2023, to March 1, 2024. The information collected was labelled according to aspects (types of extreme weather, government or community responses, and weather predictions) and sentiment (positive, negative, neutral) as presented in Table 1.

Table 1. Data Distribution

Aspect	Positive	Negative	Neutral	Total
Types of Extreme Weather	778	808	399	1985
Government or Community Response	164	126	239	529
Weather Forecast	0	8	304	312
	942	942	942	2826

The dataset goes through a preprocessing process before being trained on the model and determining aspects based on keywords to automatically categorize and label to choose positive, negative, and neutral labels. With the occurrence of data imbalance, random oversampling is carried out. Both datasets, aspects and sentiments will be divided into training and testing data. The division of training data is 80%, and testing is 20%.

4.2 Aspect Classification

Several architectures and hyperparameters are used. In addition, optimizers and loss functions are also used. In aspect classification, the optimizer used is Adam. This optimizer was chosen because it has been used frequently in previous studies and has shown satisfactory results. The loss function used is sparse categorical cross-entropy. After several experiments and testing with test data, the first experiment was run with seven epochs, namely 97.50%, after which the author conducted a second experiment with the best results of 98.94% when run with nine epochs. The following are the parameters for aspect classification presented in Table 2.

Table 2. The Parameter for Aspect Classification

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 100,100)	500000
Spatial_dropout1d (Spatial Dropout1D)	(None, 100,100)	0
lstm (LSTM)	(None, 100)	80400
dense (Dense)	(None, 3)	303
Total params: 580703 (2.22 MB) Trainable param: 580703 (2.22 MB) Non-trainable params: 0 (0.00 Byte)		

In the next stage, the evaluation of the aspect classification model is carried out by predicting aspect labels on the test dataset, generating a classification report, and presenting the corresponding confusion matrix based on the previously trained model. The confusion matrix is also visualized to provide a clearer view of the model's performance, including its overall accuracy in classifying aspect categories. The model's accuracy is reported as a percentage to facilitate interpretation. Following multiple experimental runs, the model achieved a peak accuracy of 98.94% in aspect classification. As shown in Table 3, both precision and recall scores for the extreme weather types of categories reached 0.99, indicating strong model consistency and reliability. In addition, Figure 3 illustrates minimal misclassification, further validating the robustness of the trained model in handling aspect-specific tweet data.

Table 3. Classification Report Aspects

Aspect	Precision	Recall	F1- Score	Support
Types of Extreme Weather	0.99	0.99	0.99	398
Government or Community Response	0.98	0.98	0.98	106
Weather Forecast	1.00	0.97	0.98	62
Accuracy	-		0.99	566
Macro Avg	0.99	0.98	0.99	566
Weighted Avg	0.99	0.99	0.99	566

Table 3 illustrates the evaluation of the classification model performance for three aspects: Types of Extreme Weather, Government or Community Response, and Weather Forecast. The model shows excellent performance with precision, recall, and F1-Score of 0.99 for Types of Extreme Weather (398 data), 0.98 for Government or Community Response (106 data), and 0.98 for Weather Forecast (62 data), respectively. The overall accuracy of the model reaches 99% of 566 data. The macro average and weighted average scores for precision, recall, and F1-Score are also notably high, each reaching 0.99. This demonstrates the model's consistent performance in accurately classifying all aspects.

From this classification report, the trained model is then evaluated using a confusion matrix which can be seen in Figure 3.

As confirmed by Figure 3 (aspect confusion matrix), the model made only 2 misclassifications for 'government response' aspects, indicating excellent performance. Figure 3 shows the confusion matrix used to evaluate the classification model based on three factors: extreme weather type, weather forecast, and government or community response. In the extreme weather type factor, the model accurately predicted 396 data items, with only 2 data items incorrectly assigned to government or community response. In terms of weather forecast, the model correctly identified 60 data items, while 2 other data items were incorrectly categorized as extreme weather type. In terms of government or community response, the model correctly predicted 104 data items, and 2 data items were incorrectly assigned to the extreme weather type category. This matrix shows that the model performance is very satisfactory with few misclassifications across all factors.

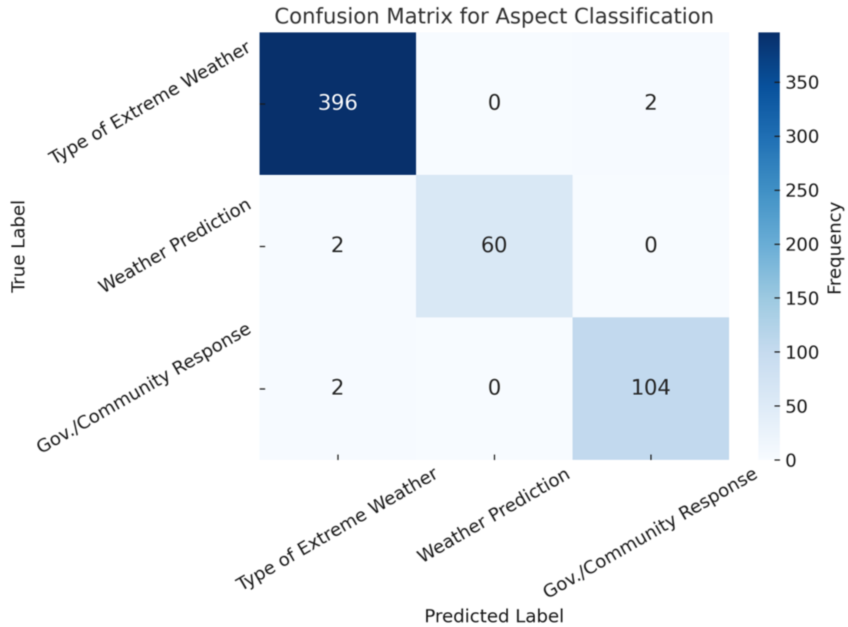

Figure 3. Confusion Matrix Aspects

After testing, we can monitor the loss from training this model, which is then depicted as in Figure 4.

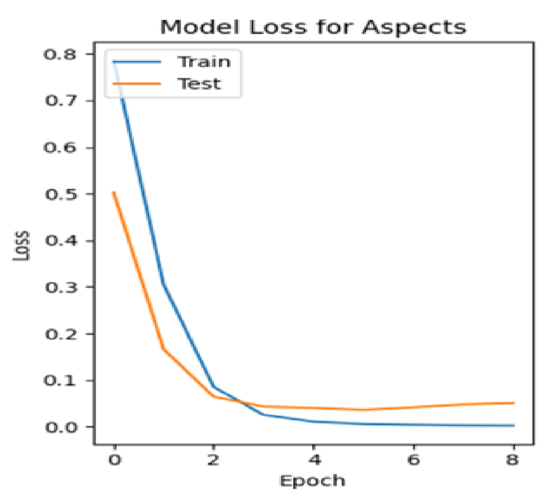

Figure 4. Aspect Loss Model

Figure 4 illustrates the loss graph of the model during the training and evaluation process at several epochs for aspect classification. This chart demonstrates that the misfortune on the preparing and assessment information diminished essentially within the to begin with few epochs, indicating an increment within the adequacy of the show as the preparing advanced. After the 4th age, the misfortune on the preparing information nearly come to zero, showing that the demonstrate might get it the information exceptionally well. On the other hand, the loss on the evaluation data also showed a low value but started to show a slight increase after a certain epoch, which could indicate the possibility of overfitting. Overall, this graph shows that the model can reduce the error on the training and evaluation data. The difference in loss fluctuation patterns between Figure 3 (aspect classification) and Figure 6 (sentiment classification) stems from the inherent complexity of sentiment expression in tweets. While aspect labels are straightforward (keyword-based), sentiment labels rely heavily on context and expression styles. Slight overfitting in the sentiment model may cause more fluctuation, as seen in the testing loss trend.

4.3 Sentiment Classification

In sentiment classification, 1D convolution is applied because this type of convolution is ideal for analysing text data. The activation function used in 1D convolution is RELU. This function produces zero for negative values and provides an increase for positive values. At the end of the network, softmax is added for sentiment classification purposes. In addition, an optimizer and loss function are also applied. For sentiment classification, the optimizer used is Adam. The choice of this optimizer is based on its consistent experience in previous research and its satisfactory results. The loss function used is sparse categorical cross entropy. After conducting several experiments and evaluations with test data, the initial test was carried out for 8 epochs with a result of 95.20%. Then, the author conducted a second experiment which produced the best number of 97.53% when run for 10 epochs. Figure 5 shows the parameters used for sentiment classification.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, 100, 100)          500000

 conv1d (Conv1D)             (None, 92, 200)           180200

 lstm_1 (LSTM)               (None, 92, 128)           168448

 lstm_2 (LSTM)               (None, 64)                49408

 dense_1 (Dense)             (None, 3)                 195

=================================================================
Total params: 898251 (3.43 MB)
Trainable params: 898251 (3.43 MB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 5. Sentiment Classification Parameters

Entering the classification report and confusion matrix from the previously created model training, visualizing the confusion matrix, providing the overall accuracy value of the model in classifying sentiment, and displaying the model accuracy in percentage form. The best sentiment classification accuracy was obtained, producing an accuracy of 97.53%. As seen in Table 4, the model achieved an F1-Score of 0.98 for positive sentiment, with overall accuracy of 97.53%. Meanwhile, Figure 6 illustrates the model's consistency through its confusion matrix.

The macro average represents the unweighted mean of the metrics across all classes, while the weighted average accounts for class imbalance. Both metrics confirm the model's strong generalizability. Table 4 presents the measurements to survey the model's execution in classifying opinions into three categories: Negative, Unbiased, and Positive. The accuracy, review, and F1-Score for each category are exceptionally tall, extending from 0.95 to 1.00. In particular, the Positive category performs excellently, with precision, recall, and F1-Score all at 1.00, indicating that all predictions for this category are correct. The model achieves an overall accuracy of 98% across the 566 samples analysed, indicating an outstanding performance. In addition, the macro-average and weighted average values for precision, recall, and F1-Score all reach 0.98, indicating consistent performance across categories considering the proportion of each sample. These findings emphasize the model's ability to classify sentiments effectively and equally across all three categories.

Table 4. Classification Report Sentiment

Label	Precision	Recall	F1-Score	Support
Negative	0.95	0.97	0.96	188
Neutral	0.97	0.95	0.96	193
Positive	1.00	1.00	1.00	185
Accuracy			0.98	566
Macro Avg	0.98	0.98	0.98	566
Weighted Avg	0.98	0.98	0.98	566

This classification report evaluates the trained model using a confusion matrix as shown in Figure 6.

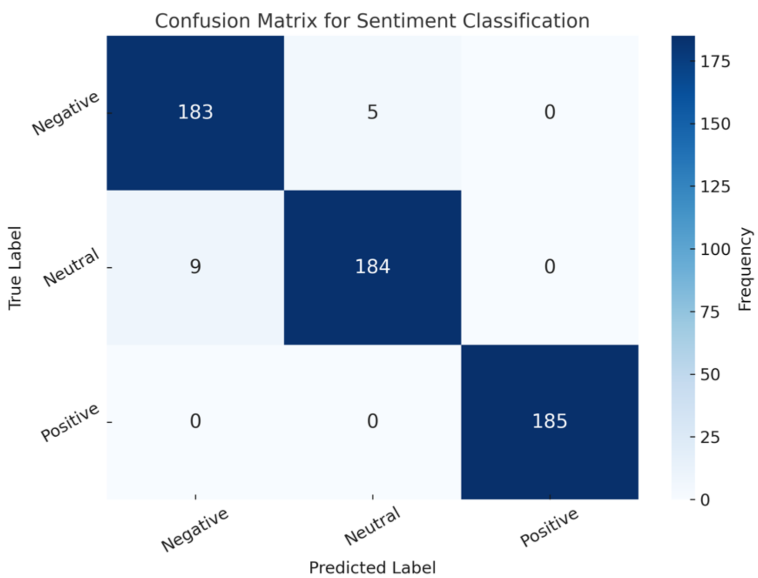

Figure 6. Confusion Matrix Sentiment

The confusion matrix used to evaluate the performance of the model in sentiment classification. The network appears the number of precise and erroneous forecasts in each estimation category, to be specific Negative, Unbiased, and Positive. The numbers found on the most corner to corner demonstrate rectify expectations, specifically 183 for the Negative category, 184 for Impartial, and 185 for the Positive category. The numbers outside the diagonal reflect errors in prediction, such as nine Neutral examples identified as Negative and five Negative examples incorrectly marked as Neutral. For the Positive category, there were no prediction errors. This matrix shows that the model performed very well, with most of the predictions falling within the correct category.

After testing, we can monitor the loss from training this model, depicted in Figure 7.

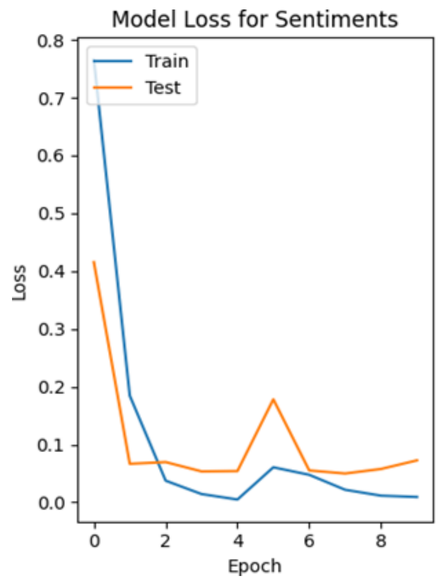

Figure 7. Sentiment Model Loss

It shows the graph of the change in the model loss value during training (Train) and testing (Test) for sentiment classification against the number of epochs. This graph illustrates how the model performance improves over time. At the beginning of training (epoch 0-1), the loss value is high for both Train and Test but quickly decreases, indicating that the model learns well from the data. After several epochs, the Train loss continues to decrease towards zero, while the Test loss remains low with slight fluctuation, indicating good generalization to the test data. However, small fluctuations in the Test loss after a specific epoch may indicate potential mild overfitting, although overall, the model shows stable and effective performance. The difference in loss fluctuation patterns between Figure 4 (aspect classification) and Figure 7 (sentiment classification) may stem from the inherent complexity of sentiment expression in tweets. While aspect labels are more straightforward (based on keywords), sentiment labels rely heavily on context and expression styles. Additionally, slight overfitting in the sentiment model may cause more fluctuation, as seen in the testing loss tren.

4.4 Results Analysis

The example of result (Figure 8) and word cloud (Figure 9) highlight dominant terms like 'flood' and 'heat,' which correlate with negative sentiment in Table 4. The macro average (unweighted mean across classes) and weighted average (accounting for class imbalance) both show consistent performance (0.98), confirming the model's generalizability

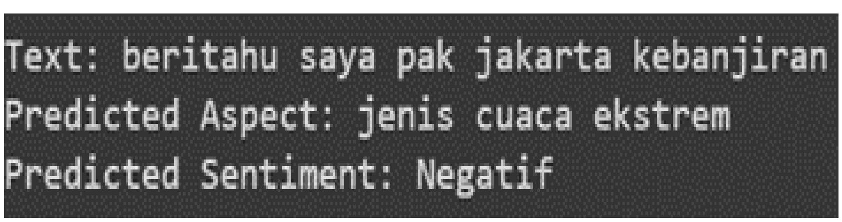

Figure 8. Results example

Word Cloud (Figure 9) highlights key themes, such as public focus on election fairness ('jujur') and criticism of corruption ('korupsi'), validating the LSTM's sentiment classification results in Table 4.

Figure 9. Word Cloud

5. CONCLUSION

Based on the results of the research conducted, it can be concluded that the dataset was obtained through crawling in X and obtained a total of 1624 tweets divided into three aspects (types of extreme weather, government or community responses, weather predictions) and three sentiments (positive, neutral, negative). Most public sentiment towards extreme weather in Indonesia is negative, especially in the context of flooding and government responses. The LSTM model successfully classified sentiment with a high level of accuracy, namely 98.56% for aspects and 97.33% for sentiment. As for the suggestion, many of the datasets used for research have not been appropriately processed due to time constraints, and additional research can try using more complex deep learning models, for example, using Transformer-based models such as BERT for sentiment analysis.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Tursun Abdurahmonov: Conceptualization, Methodology, Writing – Original Draft Preparation;
Muhammad Nabil Toby Abiyyu: Writing – Review & Editing;
Dzikru Nur Khayat: Project Administration, Writing – Review & Editing;
M. Ary Heryanto : Project Supervision, Validation, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] Y. E. N. Nugraha, I. Ariawan, and W. A. Arifin, “WEATHER FORECAST FROM TIME SERIES DATA USING LSTM ALGORITHM,” *JURNAL TEKNOLOGI INFORMASI DAN KOMUNIKASI*, vol. 14, no. 1, pp. 144–152, Mar. 2023. doi: 10.51903/jtikp.v14i1.531.

[2] H. Hikmah, A. Asrirawan, A. Apriyanto, and N. Nilawati, “Peramalan Data Cuaca Ekstrim Indonesia Menggunakan Model ARIMA dan Recurrent Neural Network,” *Jambura Journal of Mathematics*, vol. 5, no. 1, Feb. 2023. doi: 10.34312/jjom.v5i1.17496.

[3] M. Z. Rahman, Y.A. Sari, and N. Yudistira, (2021), “Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(11), 5120-5127.

[4] A. F. Wijaya, and H. Faradian, "Sentiment Analysis of Indonesia’s Extreme Weather Using LSTM," B.S. thesis, Universitas Pancasila, Jakarta, 2024.

[5] L. Li, T.-T. Goh, and D. Jin, “How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis*,” Neural Computing and Applications*, vol. 32, no. 9, pp. 4387–4415, Nov. 2018. doi: 10.1007/s00521-018-3865-7.

[6] W. K. Sari, “Klasifikasi Teks Menggunakan Long ShortTerm Memory Dengan Fitur Word Embedding Glove Dan Word2Vec,” 2020.

[7] L. R. Krosuri, and R. Satish, “Sentiment Analysis using Neural Network and Sentiment Analysis using Neural Network and LSTM,” *IOP Conf. Ser. Mater. Sci. Eng.*, 2021. doi: 10.1088/1757-899X/1074/1/012007.

[8] A. A. W. Maharani, “Sentiment Analysis of Forest Fires on Social Media Networks Twitter Using the Long ShortTerm Memory (LSTM) Method,” *Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, 2023. doi: 10.30865/klik.v3i6.866.

[9] N. O. Aljehane, “A New Approach to Sentiment Analysis on Twitter Data with LSTM,” in *2023 3rd International Conference on Computing and Information Technology (ICCIT)*, 2023, pp. 657–663. doi: 10.1109/ICCIT58132.2023.10273876.

[10] A. Singh, J. Rawat, S. Bhandari, U. Negi, and S. Pahwa, “Sentiment Analysis on Twitter Data Using LSTM with Emoji Embedding,” In *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, pp. 681–686, 2022. doi: 10.1109/ICIEM54221.2022.9853058.

[11] C. Zhang, and Y. Lu, “Study on artificial intelligence: The state of the art and future prospects,” *Journal of Industrial Information Integration*, vol. 23, 2021. doi: https://doi.org/10.1016/j.jii.2021.100224.

[12] Y. Romadhoni, and K. F. H.Holle, “Analisis Sentimen Terhadap PERMENDIKBUD No. 30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 7, no. 2, pp. 118–124, 2022. doi: 10.30591/jpit.v7i2.3191.

[13] S. Y. Kolo, and S. Supatman, “Analisis Sentimen Terhadap Opini Masyarakat Terkait Perubahan Cuaca Di Indonesia Menggunakan Algoritma Support Vector Machine,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1412–1416, 2024.

[14] M. Y. Hidayatulloh, A. Sunanto, A. Armansyah, M.F.A. Gevin, and D.D. Saputra, “Optimasi Sentimen Analisis Informatif dan Tidak Informatif dari Tweet di BMKG Menggunakan Algoritma Naive Bayes dan Metode Teknik Pengambilan Sampel Minoritas Sintetis,” *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 7, pp. 1–12, 2023. doi: 10.30645/j-sakti.v7i1.565.

[15] N. M. Sham and A. Mohamed, “Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches,” *Sustainability*, *14*(8), 2022. doi: https://doi.org/10.3390/su14084723.

[16] S.M. ALRASHIDI, F.N. ALANAZI, H.A. ALBALAWI, O.M. ALBALAWI, and A.M. AWADELKARIM, “Machine Learning-Based Sentiment Analysis for Tweets Saudi Tourism,” *Journal of Theoretical and Applied Information Technology* 100, no.16, pp. 5096-5109.

[17] R. K. Behera, M. Jena, S.K. Rath, and S. Misra, “Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data,” *Information Processing & Management*, vol. 58, no. 1, 2021. doi: https://doi.org/10.1016/j.ipm.2020.102435.

[18] I. H. Kusuma, and N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 3, pp. 302–307, 2023. doi:

10.30591/jpit.v8i3.5734.

[19] T. Zhao, L. Meng, and D. Song, 'Multimodal Aspect-Based Sentiment Analysis: A survey of tasks, methods, challenges and future directions', *Information Fusion*, vol. 112, p. 102552, Dec. 2024. doi: 10.1016/j.inffus.2024.102552.

[20] X. Shi, M. Yang, M. Hu, F. Ren, X. Kang, and W. Ding, 'Affective knowledge assisted bi-directional learning for Multi-modal Aspect-based Sentiment Analysis', *Computer Speech & Language*, vol. 91, p. 101755, Apr. 2025. doi: 10.1016/j.csl.2024.101755.

[21] A. Hariz, I. Bin, A. Azrir, P. Naveen, and S.-C. Haw, "Sentiment Analysis using Machine Learning Models on Shopee Reviews," *Journal of System and Management Sciences*, vol. 14, no. 2, Jan. 2024, doi: 10.33168/jsms.2024.0213.

[22] A. O. Aremu and I. Muhammad, "Sentiment analysis in social media: A case study of hike in university school fees in selected Nigerian universities," *Journal of Informatics and Web Engineering*, vol. 3, no. 2, pp. 98–104, June 2024, doi: 10.33093/jiwe.2024.3.2.7.

BIOGRAPHIES OF AUTHORS

	**Tursun Abdurahmonov** is from faculty of Computer Technology and Computer Science at Termez State University. He earned a master's in computing and informatics with a research focus in the field of Implementation of Elliptic Curve Exponentiation over Prime Field (Fp) in Global Smart Cards. He can be contacted at email: tursun.abdurahmono07@gmail.com.
	**Muhammad Nabil Toby Abiyyu** is affiliated with Computer Science from Dian Nuswantoro University in Semarang, Indonesia. With a strong foundation in computer science principles, he specializes in data science, focusing on data analytics, and machine learning. In addition, he has expertise in website development, combining technical skills with creative problem-solving to build user-friendly and efficient web applications. He can be contacted at email: 111202012988@mhs.dinus.ac.id.
	**Dzikru Nur Khayat** is affiliated with Electronic Engineering from Dian Nuswantoro University in Semarang, Indonesia. With a strong foundation in computer science principles, he specializes in IoT (Internet of Things), focusing on artificial intelligence. In addition, he has expertise in electrical development. He can be contacted at email: 511202301226@mhs.dinus.ac.id.
	**M. Ary Heryanto** is a researcher and academic who earned his doctorate in Electrical Engineering from the University of Indonesia. He specializes in robotics, control systems, and machine learning, with a focus on applications such as PID controllers, fuzzy logic, and neural network-based solutions. His notable works include research on quadcopter control, facial skin classification using GLCM and neural networks, and the development of portable rehabilitation devices utilizing advanced control mechanisms. He is also actively involved in community engagement initiatives, such as robotics training for special education schools. His contributions to electrical and control engineering are well-documented in various impactful publications. He can be contacted at email: m.aryheryanto@dsn.dinus.ac.id.