ISSN 2355-3286

Long Term Prediction of Extreme Weather with Long Short-Term Memory (LSTM) Model: Effect of Climate Change

Nurulita Purnama Putri[1], Adhi Harmoko Saputro[2]
[1,2]Department of Physics, Faculty of Mathematics and Science, Universitas Indonesia, Depok, Indonesia
[1]nurulita.purnama@ui.ac.id
[2]adhi.harmoko@ui.ac.id

Accepted January 11, 2025
Approved June 24, 2025

***Abstract* — Increasingly intense climate change has increased the frequency and intensity of extreme weather, making weather prediction critical for mitigation and adaptation. This research focuses on long-term prediction of extreme weather using the Long ShortTerm Memory (LSTM) model, as well as evaluating the influence of climate change on prediction accuracy. In this study, historical weather data is used to train and test an LSTM model combined with a RandomForestClassifier. Analysis was carried out using the Mean Squared Error (MSE) evaluation technique for 50 epochs and 8 trials at various threshold values (26, 29, 32, 35, 38, 41, 44, 47). The research results show that the LSTM model is able to predict extreme weather with an accuracy of up to 100%. Apart from that, this research also predicts daily rainfall in Bandung City through the process of data collection, preprocessing, normalization and evaluation using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). This model produces an RMSE of 4.24 mm and an MAE of 2.72 mm, indicating quite good predictions. It is hoped that this research can make a significant contribution to the field of meteorology and can be developed further by adding parameters or other methods to improve the quality of predictions. Suggestions are given to increase the amount of data used to obtain better prediction results in the future.**

Index Terms— Weather Prediction; Long Short Term Memory (LSTM); Climate Change; Extreme Weather.

I. INTRODUCTION

Technology is developing very quickly, producing new knowledge that is useful in areas such as education, economics, and meteorology [1]. These developments and changes aim to facilitate human activities [2]. Artificial intelligence has emerged as an algorithmic system designed and developed to have the ability to innovate in various fields, be it computers or machines, so that it is able to think like humans or even surpass them [3].

As the amount of data increases, accuracy also increases. Weather is the atmospheric conditions within a limited and specific time span [4]. Understanding the weather situation is very useful because it provides knowledge for early preventive action when extreme weather occurs [5]. The development of increasingly sophisticated technology opens up opportunities to increase the accuracy of weather predictions [6]. Weather prediction is an effort to understand the development of atmospheric conditions from the past, present and future to anticipate abnormal conditions [7].

Understanding weather analysis and prediction is critical, and research in this area continues to grow. The accuracy obtained using ensemble learning reached 81.21% with an MSE of 18.79%. With the decision tree algorithm forest method, accuracy reached 82.38% with an MSE of 17.62%, while the deep learning method produced an accuracy of 82.92% with an MSE of 17.08%. MSE (Mean Squared Error) is a metric used to measure the extent to which the model matches actual observation data, as well as how accurate the weather predictions produced by the model are [8]. Apart from that, the process of integrating weather radar monitoring throughout Indonesia still requires quite a long time [9].

Predicting extreme weather is a challenge for modeling experts in Indonesia and around the world [10]. Extreme weather is a complex problem due to the small probability of its occurrence, which causes the models developed to often have low accuracy [11]. The main aim of this research is to increase accuracy in predicting and analyzing extreme weather in order to provide more precise predictions in the future [12].

This research uses a deep learning method, namely the Long Short-Term Memory (LSTM) algorithm, which was further developed based on data. LSTM is an artificial neural network method that utilizes information from the past to be processed in the form of sequential data [13]. By using the LSTM algorithm, the accuracy of weather analysis can be improved, supporting various sectors in making more informed decisions. The high accuracy of extreme weather predictions can provide more accurate information, thereby reducing the risk of extreme weather impacts. The use of artificial intelligence technologies such offer learning and data processing with Random Forest

Classifier models for weather analysis and prediction offers the possibility of better planning. The results of this research have great potential to improve understanding of weather information, as well as the contribution of artificial intelligence technology with deep learning methods and LSTM algorithms in better weather analysis and prediction.

II. LITERATURE REVIEW

A. *Extreme Weather and Climate Change*

Extreme weather is an atmospheric event that is outside the historical norms of a region and can cause significant impacts on the environment and human life [14]. Phenomena such as storms, floods, heat waves and droughts are examples of extreme weather [15]. Global climate change, caused in large part by human activities such as burning fossil fuels and deforestation, has increased the frequency and intensity of extreme weather events [16]. Global warming, increasing concentrations of greenhouse gases, and changing rainfall patterns are some of the factors contributing to this climate change.

B. *Long Short Term Memory (LSTM)*

LSTM is another type of module of Recurrent Neural Network [17]. LSTM was created by Hochreiter and Schmidhuber [18]. Unlike a single hidden layer in an RNN, LSTM stores information in the control unit outside the normal flow of the RNN [19]. LSTM also has a chain structure like the RNN structure, the difference lies in the structure of the repetition module [20]. LSTM is suitable for problems that have long-term dependencies [21]. Considering long-term information is the inherent behavior of LSTM [22]. In the LSTM structure, there is a memory block which contains three gates (memory gate, forgetting gate, and output gate). On the other hand, LSTM also has a memory unit, which functions to control the transfer of information to the next stage [23].

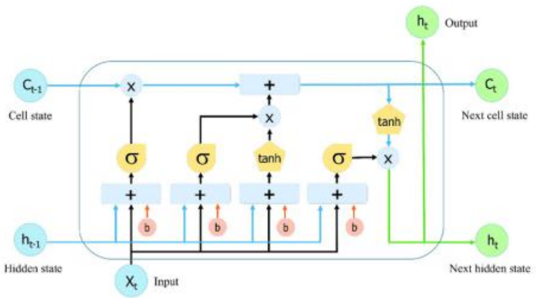

Fig. 1. LSTM Architecture

Here's how the LSTM architecture works from Figure 2 which is explained by the equation below [24]:

- Forget Gate
  The forget gate determines how much information was stored at a previous time or information that was removed.
  $$f_t = \sigma\left(w_f \times [h_{t-1}, x_t] + b_f\right) \quad (1)$$
  where:
  w_f: The weight of the foregate gate matrix
  x_t: Input data
  h_{t-1}: Output of the previous memory block
  b_f: Bias from foregate gate
  σ : Sigmoid function
- Input Gates
  The gate input determines how much information is obtained from $x_t$what is stored in the cell state c_t:
  $$i_t = \sigma(w_f \times [h_{t-1}, x_t] + b_i) \quad (2)$$
- Output Gate
  The gate output has a relationship to the output result h_t.
  $$\tilde{c}_t = tanh(w_c \times [h_{t-1}, x_t] + b_c) \quad (3)$$
  $$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (4)$$
  $$h_t = o_t \times \tanh(c_t) \quad (5)$$
  From the equation above, it can be seen that the LSTM input not only produces output h_{t-1} from the hidden layer neurons in the last stage but also contains the memory unit values in the LSTM unit. LSTM can effectively avoid gradient loss, can remember long-term historical information, and is suitable for long-term time series data more effectively.

C. *Weather Prediction Using LSTM*

In the context of weather prediction, LSTM can utilize historical weather data to identify patterns and trends that can be used to predict future weather conditions [25]. By considering variables such as temperature, humidity, atmospheric pressure, and wind speed, LSTM can produce more accurate prediction models compared to traditional statistical methods [26]. LSTM's advantage in handling sequential data makes it a very useful tool in predicting extreme weather events that are influenced by many interrelated factors [27].

D. *Physical Theory in Weather Prediction*

Weather predictions are strongly influenced by the physical laws that govern atmospheric dynamics. Some relevant physical theories include:

- Ideal Gas Law: States that the pressure (P), volume (V), and temperature (T) of a gas are related to each other, which is very important in understanding atmospheric dynamics. The equation is [28]:
  $$PV = nRT \quad (6)$$
  Where(P) is the pressure , (V)is the volume, (n)is the number of moles of gas, (R)is the ideal gas constant, and (T)is the temperature in Kelvin.

- Laws of Thermodynamics: These principles govern the transfer of energy and phase changes of water in the atmosphere, such as condensation and evaporation, which influence cloud formation and precipitation. An example of an equation is [29]:

$$Q = mc\Delta T \quad (7)$$

where (Q) is the heat added or released, (m) is the mass, (c) is the heat capacity, and (ΔT) is a change in temperature.

- Newton's Laws of Motion: Used to model the movement of air in the atmosphere, including winds and vertical air currents that can influence the weather. The basic equation is [30]:

$$F = ma \quad (8)$$

where (F) is the force, (m) is the mass, and (a) is the acceleration.

E. Evaluation Criteria

To evaluate the performance of the model with the method used, the author uses mean absolute error (MAE) and root mean square error (RMSE) with the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i' - y_i)^2} \quad (9)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i' - y_i| \quad (10)$$

III. Research Method

This research included in the quantitative research category uses the Long Short-Term Memory (LSTM) algorithm. LSTM is a type of artificial neural network architecture, which is specifically designed to handle time sequence data processing.

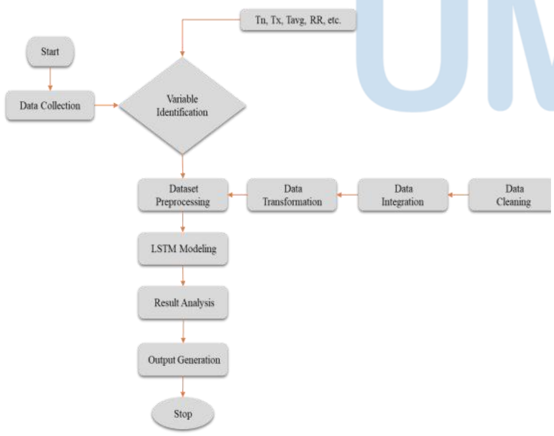

Fig. 2. Research Method

The stages in the research process are as follows:

- Data collection: At this stage, data will be collected before proceeding to the next process. Download data from 1 January 2023 to 31 May 2024 in Bogor City obtained from the BMKG website (http://dataonline.bmkg.go.id/dashboard_user).
- Variable identification: This stage involves identifying the variables to be analyzed in the research. Variables can be numeric or non-numeric data.
- Dataset pre-processing: This stage involves cleaning the data that has been collected by removing incomplete or irrelevant data. This includes deleting incomplete data, deleting duplicate data, and deleting data that is not relevant to the research topic.
- Data transformation: This stage involves converting the data into a format that can be used in analysis. This includes converting the data into a format that can be used in an LSTM model.
- Data integration: This stage involves combining data from various sources into a single dataset. This includes combining data from various sources, such as surveys, experiments, and observations.
- Data cleaning: This stage involves removing incomplete or irrelevant data from the dataset. This includes deleting incomplete data, deleting duplicate data, and deleting data that is not relevant to the research topic.
- LSTM Modeling: This stage involves using an LSTM model to analyze the data. The LSTM model is a type of model that can be used to analyze data that has a time structure.
- Result Analysis: This stage involves analyzing the results from the LSTM model to extract insights and conclusions. This includes using statistical analysis techniques to extract insights and conclusions from research results.
- Output Generation: This stage involves producing a report or presentation that describes the results of the research. This report or presentation can be used to share research results with others.

A. Data source

Weather data is taken from the BMKG website, Java Jarat Province climatology station Citeko Meteorological Station, Latitude -6.70000, Longitude 106.85000, Elevation 920. The following is the website address (https://dataonline.bmkg.go.id/home).

B. Population and Sample

The data population includes all daily weather data records, from 1 January 2023 to 31 May 2024, which reaches 1640 data. Data samples are Tn (minimum temperature), Tx (maximum temperature), Tavg (average temperature), RH_avg (average humidity), RR (rainfall), ss (duration of sunlight), ff_x (maximum wind speed), ddd_x (wind direction at maximum speed), ff_avg (average wind speed), ddd_car (most wind direction)

C. Data Retrieval Techniques

Online data collection, taken from the official website of the BMKG (Meteorology, Climatology and

Geophysics Agency) which can be used legally, by downloading the data periodically once a month from 1 January 2023 to 31 May 2024.

D. *Data Retrieval Techniques*

Weather data is taken from the BMKG website, Java Jarat Province climatology station Citeko Meteorological Station, Latitude -6.70000, Longitude 106.85000, Elevation 920. The following is the website address (https://dataonline.bmkg.go.id/home). The following data is an overview and explanation of the data analysis techniques that will be carried out in this research:

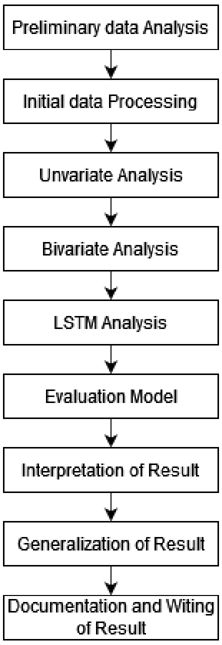

Fig. 3. Data Analysis Techniques

The following are the data analysis techniques for this research:

- Data analysis: to understand the characteristics of weather data.
- Data processing: ensuring data is in good condition.
- Univariate analysis: understanding the distribution between variables, including patterns and trends in the data.
- Bivariate analysis: understanding the correlation between variables.
- LSTM Analysis: train a model with an LSTM algorithm.
- Evaluation Model: evaluating includes the use of metrics such as MSE.
- Interpretation of results: predicting future weather using the LSTM model.
- Generalization of results: reviewing the results of sample weather data to the population as a whole.
- Documentation and writing of results: making conclusions and documenting the results that have been obtained.

IV. Results and Discussion

A. *Data retrieval*

Data was taken via the official BMKG website which can be used legally, along with the website address (https://dataonline.bmkg.go.id/home) by downloading once a month.

B. *Data preprocessing*

The first step in this process is to load the cleaned data. This dataset contains important weather variables such as average temperature (Tavg) in degrees Celsius (°C), average relative humidity (RH_avg) in percentage (%), daily rainfall (RR) in millimeters (mm), sunshine duration (ss) in hours (hrs), and daily maximum wind speed (ff_x) in meters per second (m/s). A sample excerpt of the data is shown in Table 1, which includes daily records from January 1, 2023, to May 31, 2024.

TABLE I. CLEAN DATA READING CODE

No	Date	Tavg	RH_avg	RR	ss	ff_x
1	01-01-2023	19.6	93	31.40	0.0	3.0
2	01-02-2023	20.2	96	1.60	0.0	4.0
3	01-03-2023	19.5	96	1.00	6.0	4.0
4	01-04-2023	20.3	90	10.80	0.0	4.0
5	01-05-2023	21.3	81	9.30	0.4	4.0
…	…	…	…	…	…	…
513	05-25-2024	23.0	84	9.90	2.0	4.0
514	05-28-2024	23.1	84	6.75	6.1	4.0
515	05-30-2024	23.0	82	0.00	0.5	4.0
516	05-31-2024	23.1	84	0.00	7.4	4.0

In the table, the numeric data are initially stored as floating-point numbers, for example, temperature values like 19.6°C and rainfall amounts such as 31.40 mm. To facilitate easier classification of weather conditions, these floating-point values are converted to integer values by rounding to the nearest whole number. For instance, a temperature of 19.6°C is rounded to 20°C, and rainfall of 31.40 mm becomes 31 mm. This conversion simplifies subsequent analysis and speeds up the classification process.

Following this conversion, weather conditions can be categorized based on a comparison between rainfall (RR) and sunshine duration (ss). If the rainfall amount is greater than the sunshine duration, the day is classified as a rainy day. If both values are equal, the day is considered cloudy. If the rainfall is less than the sunshine duration, the day is classified as sunny.

Moreover, to identify extreme weather events, the average values for each weather variable over the observation period are calculated. These averages serve as a baseline to determine upper and lower threshold limits that define extreme conditions, such as extremely high or low temperatures and unusually heavy rainfall. These thresholds are useful for further analysis and decision-making. Data after data type change:

TABLE II. CODE TO CHANGE FLOAT DATA TO INTEGER

No	Date	Tavg	RH_avg	RR	ss	ff_x
1	01-01-2023	19	93	31	0	3
2	02-02-2023	20	96	1	0	4
3	03-03-2023	19	96	1	0	4
4	04-04-2023	20	90	10	0	4
5	05-05-2023	21	81	9	0	4

Table 2 shows the data after converting floating-point values to integers for easier processing. The table includes daily records of temperature (Tavg in °C), relative humidity (RH_avg in %), rainfall (RR in mm), solar intensity (ss in hours), and wind speed (ff_x in m/s).

Weather conditions are classified into three categories: rainy, cloudy, and clear. A day is considered rainy if the rainfall (RR) exceeds the solar intensity (ss), cloudy if both values are equal, and clear if the rainfall is less than the solar intensity. The table displays sample data after this conversion, which facilitates the classification process. Weather data with conditions:

TABLE III. RESULTS OF WEATHER CONDITIONS

No	Date	Tavg	RH_avg	RR	ss	ff_x	Wx cond.
1	01-01-2023	19	93	31	0	3	Rain
2	01-02-2023	20	96	1	0	4	Rain
3	01-03-2023	19	96	4	0	4	Rain
4	01-04-2023	20	90	10	0	4	Rain
5	01-05-2023	21	81	9	0	4	Rain
…	…	…	…	…	…	…	…
512	05-25-2024	23	84	9	2	4	Rain
513	05-28-2024	23	84	6	6	4	Cloudy
514	05-29-2024	23	83	4	7	3	Light
515	05-30-2024	23	82	0	6	4	Light
516	05-31-2024	23	84	0	7	4	Light

Table 3 presents a detailed overview of daily weather conditions recorded over the study period from January 1, 2023, to May 31, 2024. Each row contains data for a specific date, including the average temperature (Tavg in °C), average relative humidity (RH_avg in %), rainfall (RR in mm), solar radiation or sunshine duration (ss in hours), and maximum wind speed (ff_x in m/s). These meteorological parameters are used to classify daily weather conditions into categorical labels: "Rain," "Cloudy," or "Light".

For example, on January 1, 2023, the recorded values were a temperature of 19°C, humidity of 93%, rainfall of 31 mm, no sunlight (0 hours), and wind speed of 3 m/s, leading to a "Rain" classification. Toward the end of the dataset, we observe decreasing rainfall and increasing sunshine values. For instance, on May 30, 2024, the data shows 23°C temperature, 82% humidity, 0 mm rainfall, 6 hours of sunshine, and 4 m/s wind speed, categorized as "Light" weather.

This categorization is a precursor to the development of predictive models, as it enables the identification of weather patterns and assists in detecting extreme events. The qualitative weather condition in the "Wx cond." column is derived from logical thresholds primarily influenced by the balance between rainfall and sunlight thus forming a foundational step for further quantitative analysis and machine learning classification.

TABLE IV. AVERAGE VALUE RESULTS

No	Date	Tavg	RH_avg	RR	ss	ff_x	Wx cond.	Avg
1	01-01-2023	19	93	31	0	3	Rain	29.40
2	01-02-2023	20	96	1	0	4	Rain	24.36
3	01-03-2023	19	96	4	0	4	Rain	24.84
4	01-04-2023	20	90	10	0	4	Rain	25.04
5	01-05-2023	21	81	9	0	4	Rain	23.20
…	…	…	…	…	…	…	…	
512	05-25-2024	23	84	9	2	4	Rain	24.40
513	05-28-2024	23	84	6	6	4	Cloudy	24.79
514	05-29-2024	23	83	4	7	3	Light	24.26
515	05-30-2024	23	82	0	6	4	Light	23.06
516	05-31-2024	23	84	0	7	4	Light	23.70

Table 4 presents the daily average values of key weather parameters, including average temperature (Tavg, in °C), relative humidity (RH_avg, in %), rainfall (RR, in mm), sunlight duration (ss, in hours), and maximum wind speed (ff_x, in m/s), along with the qualitative weather condition (Wx cond.). The final column, "Avg," represents a composite index calculated from the combined values of the aforementioned weather parameters. This index is used as a quantitative reference to evaluate and classify daily weather conditions.

For example, on January 1st, 2023, the average temperature was 19°C, relative humidity was 93%, rainfall reached 31 mm, sunlight duration was 0 hours, and the maximum wind speed was 3 m/s. The weather condition was classified as "Rain," with a calculated average index value of 29.40. In contrast, on January 2nd, 2023, despite the temperature increasing to 20°C, the rainfall dropped to just 1 mm, resulting in a lower average value of 24.36.

Across the full dataset of 516 days, the highest "Avg" value recorded was 45.0, indicating severe weather conditions, likely driven by intense rainfall or an extreme combination of other factors. The lowest "Avg" value was 18.84, reflecting calm and possibly clear weather. The overall mean of the "Avg" values is approximately 24.69, which is rounded to 25 and used as a standard threshold to differentiate between normal and extreme weather conditions in West Java.

This average weather index plays a crucial role in training and evaluating machine learning models, such as LSTM and RandomForestClassifier, by providing a consistent numerical basis for labeling weather as either "normal" or "extreme" based on real-time parameter combinations.

Following the detailed breakdown of daily weather averages in Table 4, an overall statistical summary is provided in Table 5 to establish a clearer reference for normal and extreme weather conditions based on quantitative thresholds.

TABLE V. AVERAGE VALUE RESULTS

Highest (Average) Score: 45.0
Lowest (Average) Score: 18.84
Value (Average) of average value: 24.688820116054156

Table 5 presents the summary statistics of the average daily weather scores. The highest average score recorded is 45.0 units, the lowest is 18.84 units, and the overall mean average is approximately 24.69 units. Based on this, the typical daily weather value for West Java is around 24.69 units.

This average value serves as a key threshold for classifying daily weather conditions into either "normal" or "extreme." Days with an average score significantly above this threshold (e.g., above 26 or 29 units) are more likely to be associated with extreme weather events, particularly heavy rainfall. Conversely, days with scores below the threshold typically reflect normal or calm weather conditions. This statistical benchmark is used as a reference input for machine learning models such as LSTM and Random Forest to improve accuracy in predicting and labelling weather conditions.

C. Long Short Term Memory Model Result

Before training the Long Short-Term Memory (LSTM) model, it is essential to prepare clean and well-labeled data that accurately reflects daily weather conditions. This involves transforming raw meteorological variables into structured features and categorizing them based on threshold values to distinguish between *normal* and *extreme* weather. These classifications are critical for supervised learning models such as LSTM, which rely on labeled sequences to learn temporal patterns.

TABLE VI. CODE FOR READING CLEAN DATA 26

No	Date	Tavg	RH_avg	RR	ss	ff_x	Wx cond.	Avg	Label
1	01-01-2023	19	93	31	0	3	Rain	29.40	extreme
2	01-02-2023	20	96	1	0	4	Rain	24.36	normal
3	01-03-2023	19	96	4	0	4	Rain	24.84	normal
4	01-04-2023	20	90	10	0	4	Rain	25.04	normal
5	01-05-2023	21	81	9	0	4	Rain	23.20	normal
…	…	…	…	…	…	…	…	…	…
512	05-25-2024	23	84	9	2	4	Rain	24.40	normal
513	05-28-2024	23	84	6	6	4	Cloudy	24.79	normal
514	05-29-2024	23	83	4	7	3	Light	24.26	normal
515	05-30-2024	23	82	0	6	4	Light	23.06	normal
516	05-31-2024	23	84	0	7	4	Light	23.70	normal

For instance, on January 1, 2023, the high rainfall (31 mm), high humidity (93%), and no sunlight led to a composite score of 29.40, which exceeds the threshold and is labeled as *extreme*. In contrast, on January 2, 2023, the rainfall drops to 1 mm and the score is 24.36, thus labelled as *normal*.

This labelling serves as the basis for training and evaluating the LSTM model. Prior to LSTM, a RandomForestClassifier was applied to verify label consistency, achieving 100% accuracy in classifying the normal versus extreme labels based on the input variables. The LSTM model was then trained on this dataset over 50 epochs using a threshold score of ~25 units to forecast weather labels for the next 7 days, learning temporal dependencies and fluctuation patterns in the weather variables.

D. Analysis and prediction

Based on the value from the process, it is 24.68 rounded to 25 as a benchmark for normal daily weather values in the West Java region. Below is a graph of weather conditions. Threshold value 26. Obtained in the graph below, the red dot pattern appears quite a lot. There were 115 conditions above the threshold value of 26

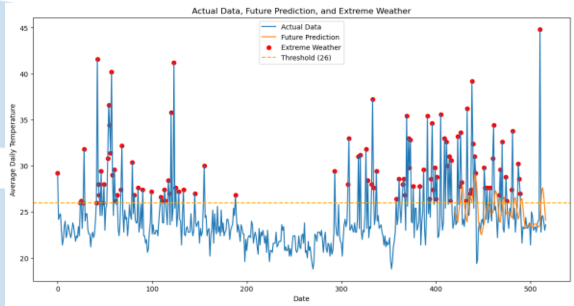

Fig. 4. Threshold Graph 26

Predict weather conditions for the next 7 days using a threshold value of 26, obtained conditions for the next 7 days as shown in the image below.

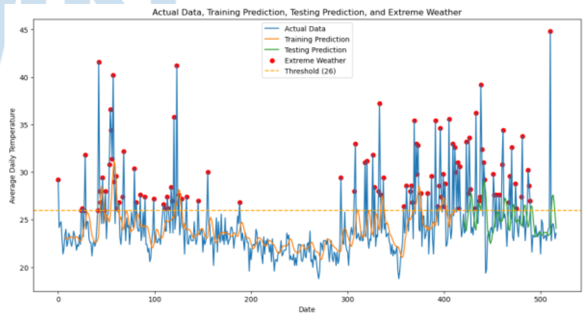

Fig. 5. Threshold 26 Prediction Graph

Next, Table 7 presents the weather prediction results using a threshold value of 26. The time steps show predicted average weather scores along with corresponding weather conditions such as “Bright” (clear) and “Rain” (extreme).

TABLE VII. PREDICTION RESULTS WITH THRESHOLD 26

Time	Step	Prediction	Wx Cond.
520	25.463339	Normal	Bright
521	24.702570	Normal	Bright
522	24.522461	Normal	Bright

Time	Step	Prediction	Wx Cond.
523	24.209877	Normal	Bright
524	27.699635	Extreme	Rain
525	28.909733	Extreme	Rain
526	27.102564	Extreme	Rain

For the next 6 days, the weather forecast is normal, bright conditions and rain on the 7th day with extreme weather. The threshold value can be input as desired with the standard normal weather value in West Java being 24.68 which is rounded to 25. The blue graph is the actual data value, the orange graph is the training data graph, the green graph is the test data graph, and the red graph is the test data graph. at the very end is the actual prediction graph.

From various experiments on the threshold values above, the higher the threshold value entered, the rarer the pattern that appears, then the predicted weather will be normal and clear. It can be seen that the extreme weather patterns that occur throughout 2023 to 2024, the extreme weather that occurs is extremely rainy weather. The following is a table of various experimental threshold values.

To evaluate the sensitivity and precision of the model in detecting extreme weather events, a series of experiments were conducted using various threshold values applied to the composite weather score (referred to as "Avg"). These thresholds serve as classification cut-off points, determining whether a particular day is categorized as experiencing *normal* or *extreme* weather. The goal was to observe how changing the threshold impacts the number of extreme weather events detected and how the model responds in predicting upcoming conditions.

TABLE VIII. EXPERIMENTAL RESULTS OF VARIOUS THRESHOLDS

No	Treshlod value	Emerging Extreme Weather Patterns	Predict conditions that will occur
1	26	115	normal conditions : 6, Rain under extreme conditions: 1
2	29	94	normal conditions : 7, Rain under extreme conditions: 7
3	32	52	normal conditions : 7, Rain under extreme conditions: 7
4	35	33	normal conditions : 7, Rain under extreme conditions: 7
5	38	29	normal conditions : 7, Rain under extreme conditions: 7
6	41	21	normal conditions : 7, Rain under extreme conditions: 7
7	44	15	normal conditions : 7, Rain under extreme conditions: 7
8	47	9	normal conditions : 7, Rain under extreme conditions: 7

Table 8 presents the detailed results of these experiments, where threshold values range from 26 to 47. As the threshold increases, the number of detected extreme weather patterns decreases significantly. At a threshold of 26, a total of 115 extreme weather events are identified. This indicates a highly sensitive setting where even moderately high weather scores are flagged as extreme. However, such sensitivity might increase false positives. As the threshold rises to 29, the number of detected extreme events drops to 94, and at 47, only 9 events are flagged, indicating a sharp decline in identified extremes representing a more conservative approach that reduces false alarms but may miss certain impactful events.

Across all threshold levels, the model consistently forecasts seven days of weather conditions, maintaining a stable predictive routine. The distinction lies in the classification: with lower thresholds, the forecast includes more days labelled as "rain under extreme conditions," whereas at higher thresholds, the number of such labels decreases. For example, at a threshold of 29, the model predicts seven days of rain classified as extreme, while still recognizing some days as normal. This consistency in forecasting behaviour despite varying thresholds indicates that the model is adaptive and responsive to new classification criteria.

These findings highlight the importance of careful threshold selection in weather modelling. Lower thresholds improve sensitivity and are useful for early warning systems, especially in areas vulnerable to frequent rainfall or fluctuating weather. In contrast, higher thresholds favour specificity, minimizing false alarms and better serving scenarios where only the most severe conditions should be flagged. Therefore, tuning the threshold value is essential to align the model with specific operational goals whether to emphasize precautionary warnings or prioritize alert accuracy.

E. Interpretation of Result

The results obtained by this research, from data from January 1 2023 to June 31 2024, by inputting the desired threshold value, can provide output information that matches the input value. The normal weather value obtained from this research is 24.68, rounded up to 25. The higher the threshold value, the more extreme the weather and the less frequently it occurs. The method used is Long Short-Term Memory, to predict and read weather patterns. Determine the occurrence of extreme or normal weather by calculating the average value in each column, because each attribute has a mutually supporting relationship. Like the picture below.

TABLE IX. NORMAL WEATHER

No	Date	Tavg	RH_avg	RR	ss	ff_x	Wx cond.	Avg	Label
6	01-07-2023	22	79	0	6	3	Light	22.0	extreme
7	01-08-2023	21	88	0	5	3	Light	23.4	normal
9	01-10-2023	22	83	1	2	3	Light	22.2	normal
10	01-11-2023	21	83	0	5	4	Light	22.6	normal
11	01-12-2023	22	85	3	4	4	Light	23.6	normal
…	…	…	…	…	…	…	…	…	…

No	Date	Tavg	RH_avg	RR	ss	ff_x	Wx cond.	Avg	Label
511	05-26-2024	22	90	0	0	2	Cloudy	22.8	normal
513	05-28-2024	23	84	6	6	4	Clear	24.6	normal
514	05-29-2024	23	83	4	7	3	Light	24.0	normal
515	05-30-2024	23	82	0	6	4	Light	23.0	normal
516	05-31-2024	23	84	0	7	4	Light	23.6	normal

On 01 – 07 – 2023 the Tavg (average temperature) value is 22°C, RH_avg (humidity) has a value of 79% indicating very humid, RR (rainfall) has a value of 0 mm, SS (sunlight intensity) has a value of 6 hours, and ff_x (maximum wind speed) has a value of 3 m/s. The average obtained is 22.0, so the condition is normal. Rainy weather conditions are caused because the RR (rainfall) value is higher than the ss (sunlight intensity) value according to a journal and vice versa. Also supported by the RH_avg (humidity) value, which is 79% very humid, and other related variables. Then following are the results of extreme weather conditions.

TABLE X. EXTREME WEATHER

No	Date	Tavg	RH_avg	RR	ss	ff_x	Wx cond.	Avg	Label
0	01-01-2023	19	93	31	0	3	Rain	29.2	extreme
24	01-25-2023	19	98	6	4	3	Rain	26.0	extreme
25	01-26-2023	20	90	17	0	4	Rain	26.2	extreme
27	01-28-2023	20	96	14	0	0	Rain	26.0	extreme
28	01-29-2023	20	97	39	0	3	Rain	31.8	extreme
…	…	…	…	…	…	…	…	…	…
481	05-26-2024	22	90	55	0	2	Rain	33.8	extreme
487	05-02-2024	23	91	34	0	3	Rain	30.2	extreme
488	05-03-2024	23	88	25	4	3	Rain	28.6	extreme
489	05-04-2024	22	87	17	6	3	Rain	27.0	extreme
516	05-25-2024	21	93	107	0	3	Rain	44.8	extreme

On 01-01-2023, there were 19 extreme weather conditions recorded, exceeding the monthly average of 29 normal conditions. The average temperature (Tavg) was 25 °C, indicating normal thermal conditions. However, the daily rainfall (RR) reached 31 mm, which qualifies as heavy rainfall and could potentially trigger flooding in affected regions.

To predict such extreme weather events, the process involves data collection, preprocessing, normalization, and evaluation using well-established metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These metrics are widely adopted in meteorological time-series prediction due to their interpretability and sensitivity to error magnitudes.

TABLE XI. ERROR CALCULATION RESULTS

Root Mean Error (RMSE)	4.2423 **mm**
Mean Absolute Error **(MAE)**	2.7247 mm

The results indicate the highest accuracy with an RMSE of 4.24 mm and an MAE of 2.72 mm. These values demonstrate reasonably good predictions, as smaller MAE and RMSE values correspond to higher prediction accuracy.

V. CONCLUSION

The results of this research show that artificial intelligence with the Long Short Term Memory (LSTM) algorithm applied together with RandomForestClassifier on historical weather data can predict extreme weather with an accuracy of up to 100%. Analysis using the Mean Squared Error (MSE) model evaluation technique with 50 epochs and 8 trials at various threshold values (26, 29, 32, 35, 38, 41, 44, 47) shows a different pattern from the normal average of 24.68 in the West Java region. In predicting daily rainfall in Bandung City, data collection, preprocessing, normalization and evaluation were carried out using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). This research produces the highest accuracy value with an an RMSE of RMSE of 4.24 mm and an MAE of 2.72 mm. These results show quite good predictions, because the smaller the MAE and RMSE values, the better the prediction accuracy.

The author hopes that this research can be further developed by other researchers in the future. The suggestion given is to add several parameters or other methods that can improve the quality of research and compare which method is superior. Apart from that, increasing the amount of data used is expected to help obtain better prediction values.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to all parties who have supported the completion of this research. Special thanks are extended to the Meteorology, Climatology, and Geophysics Agency (BMKG) of Indonesia for providing access to relevant weather data used in this study. The author also wishes to acknowledge the guidance and encouragement provided by academic supervisors and mentors throughout the research process.

Appreciation is also extended to Universitas Indonesia for facilitating research resources and to fellow researchers and colleagues whose insights and constructive feedback have enriched the quality of this work. Lastly, heartfelt thanks go to the author's family and friends for their unwavering support, motivation, and understanding during the course of this study.

REFERENCES

[1] Russ, M. (2021). Knowledge management for sustainable development in the era of continuously accelerating technological revolutions: A framework and models. Sustainability, 13(6), 3353.

[2] Scharlemann, J. P., Brock, R. C., Balfour, N., Brown, C., Burgess, N. D., Guth, M. K., ... & Kapos, V. (2020). Towards understanding interactions between Sustainable Development Goals: The role of environment–human linkages. Sustainability science, 15, 1573-1584.

[3] Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. Journal of product innovation management, 37(3), 212-227.

[4] Strauss, S., & Orlove, B. (2021). Up in the air: the anthropology of weather and climate. In Weather, climate, culture (pp. 3-14). Routledge.

[5] Walsh, J. E., Ballinger, T. J., Euskirchen, E. S., Hanna, E., Mård, J., Overland, J. E., ... & Vihma, T. (2020). Extreme weather and climate events in northern areas: A review. Earth-Science Reviews, 209, 1033

[6] Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep learning-based weather prediction: a survey. Big Data Research, 23, 100178.

[7] Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems, 12(2), e2019MS001958.

[8] A. M. Siregar, (2020) “Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning,” Petir, vol. 13, no. 2, pp. 138–147 doi: 10.33322/petir.v13i2.998.

[9] A. Prakasa and F. D. Utami (2019) “BMKG The Integrated Weather Radar Information System of BMKG,” J. Telecomunication, Electron. Control Eng., vol. 1, no. 2, pp. 86–96

[10] Kurniadi, A., Weller, E., Salmond, J., & Aldrian, E. (2023). Future projections of extreme rainfall events in Indonesia.

[11] Walsh, J. E., Ballinger, T. J., Euskirchen, E. S., Hanna, E., Mård, J., Overland, J. E., ... & Vihma, T. (2020). Extreme weather and climate events in northern areas: A review. Earth-Science Reviews, 209, 103324.

[12] Fitzsimmons, J. P., Lu, R., Hong, Y., & Brilakis, I. (2022). Construction schedule risk analysis-a hybrid machine learning approach.

[13] Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water resources research, 56(1), e2019WR025326.

[14] AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., ... & Sadegh, M. (2020). Climate extremes and compound hazards in a warming world. Annual Review of Earth and Planetary Sciences, 48, 519-548.

[15] Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., ... & Vignotto, E. (2020). A typology of compound weather and climate events. Nature reviews earth & environment, 1(7), 333-347.

[16] Gulzar, A., Islam, T., Gulzar, R., & Hassan, T. (2021). Climate change and impacts of extreme events on human health: An overview. Indonesian Journal of Social and Environmental Issues (IJSEI), 2(1), 68-77.

[17] Hu, J., Wang, X., Zhang, Y., Zhang, D., Zhang, M., & Xue, J. (2020). Time series prediction method based on variant LSTM recurrent neural network. Neural Processing Letters, 52, 1485-1500.

[18] G. van Houdt, C. Mosquera, and G. Nápoles (2020)“A review on the long short-term memory model,” Artificial intelligence review: An international survey and tutorial journal, vol. 53, no. 8, pp. 5929– 5955, doi: 10.1007/s10462-020-09838-1.

[19] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

[20] Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., ... & Stadtler, S. (2021). Can deep learning beat numerical weather prediction?. Philosophical Transactions of the Royal Society A, 379(2194), 20200097.

[21] Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. (2020, November). Do RNN and LSTM have long memory?. In International Conference on Machine Learning (pp. 11365-11375). PMLR.

[22] Huang, X. (2021). Network Intrusion Detection Based on an Improved Long-Short-Term Memory Model in Combination with Multiple Spatiotemporal Structures. Wireless Communications and Mobile Computing, 2021(1), 6623554.

[23] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang (2019) “Shortterm residential load forecasting based on LSTM recurrent neural network,” IEEE Trans Smart Grid, vol. 10, no. 1, pp. 841–851, doi: 10.1109/TSG.2017.2753802.

[24] Ambaum, M. H. (2020). Thermal physics of the atmosphere (Vol. 1). Elsevier.

[25] Semeniuk, K., & Dastoor, A. (2020). Current state of atmospheric aerosol thermodynamics and mass transfer modeling: A review. Atmosphere, 11(2), 156.

[26] López, G., & Arboleya, P. (2022). Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the Andes Mountains, Ecuador. Renewable Energy, 183, 351-368.

[27] Suleman, M. A. R., & Shridevi, S. (2022). Short-term weather forecasting using spatial feature attention based LSTM model. IEEE Access, 10, 82456-82468.

[28] Rohli, R. V., Li, C., Rohli, R. V., & Li, C. (2021). The Seven Basic Equations in Weather Forecasting Models. Meteorology for Coastal Scientists, 171-185.

[29] Bohren, C., & Albrecht, B. (2023). Atmospheric thermodynamics. Oxford University Press.

[30] Sreeraj, M. S., Sruthi, S. L., & Sumod, S. G. (2024). Modeling in Fluid Mechanics: Atmospheric Modeling Using Navier Stokes Equation. In New Advances in Materials Technologies (pp. 59-86). Apple Academic Press.

[31] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, “Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting,” IEEE Access, vol. 8, pp. 71326–71338, 2020