Machine Learning-Based Climate Prediction in Indonesia: A Baseline Experiment

Faisal Rahutomo[1], Bambang Harjito[2]
Electrical Engineering Department, Sebelas Maret University, Surakarta, Indonesia 57126[1]
Informatics Department, Sebelas Maret University, Surakarta, Indonesia 57126[2]

Abstract—**This study presents the results of a series of machine learning experiments conducted on Indonesian climate data collected between 2010 and 2020. The findings offer a comparative foundation for future research. Weather prediction remains a significant challenge due to the complex interplay of various climatic factors. Weather stations typically record data at hourly or daily intervals, resulting in large volumes of historical weather information. When appropriately processed, this extensive dataset offers valuable opportunities for predictive modeling. The study explores two primary approaches to leveraging big data for weather forecasting. The first employs a machine learning classification technique to predict categorical weather conditions based on existing feature values. The second utilizes time series forecasting to predict continuous weather parameters using historical data. Multiple classification and forecasting algorithms were evaluated and compared. Notably, the year-on-year forecasting approach outperformed several modern techniques, including deep learning, in terms of predictive accuracy. Despite the application of deep learning, classification models achieved a maximum accuracy of only 0.811. Forecasting methods generally produced a mean absolute percentage error (MAPE) of 3–4%. However, year-on-year forecasting—identified through exploratory data visualization—reduced the prediction error to below 1.6%. Another key contribution of this research is the emphasis on the critical role of data visualization prior to algorithmic modeling. The findings highlight the importance of human intervention in the early stages of data analysis, particularly for visual exploration and feature assessment. Classification models were found to underperform due to overly generalized feature representations. In contrast, forecasting techniques, supported by informed human-guided preprocessing, yielded more reliable and accurate results.**

Keywords*—*Indonesia; climate data; experiment baseline; machine learning; prediction

I. Introduction

Weather is a natural phenomenon that is very difficult to predict accurately. Many parameters affect weather changes, such as temperature, humidity, air pressure, and wind speed [1]. These factors are dynamic and challenging to predict because they are filled with uncertainty and are very complex. However, weather prediction is an important aspect that significantly impacts various sectors, such as agriculture, transportation, and disaster management [2], [3]. The need for reliable predictions becomes even more critical as our world becomes increasingly dependent on the weather [4]. Accurate information about future weather can help the public and industry make the right decisions [5].

Weather prediction is done by using historical data and atmospheric variables to make estimates of future weather conditions. The weather prediction approaches are diverse, ranging from simple statistical methods to advanced technologies such as artificial intelligence (AI) [6]. Classification and regression-based machine learning modeling approaches can be used to scientifically analyze extreme weather phenomena [7]. Currently, national weather centers around the world use a numerical weather prediction model that forecasts weather conditions based on the state of the atmosphere. This model requires high computational performance and takes many hours [8]. In the case of rainfall prediction, accurate rainfall forecasting remains a challenge due in part to the non-linear nature of rainfall. Prediction is done using time series forecasting, which attempts to uncover hidden patterns in the data and, using known values, predict future data with a reasonable degree of accuracy [9]. Stochastic models such as seasonal autoregressive integrated and moving average with exogenous variable (SARIMAX) are widely used for rainfall forecasting [10]. One branch of AI is machine learning, which utilizes large amounts of data to recognize complex and nonlinear relationships, making it highly effective for generating predictions [11], [12]. Machine learning models can be trained to create more accurate weather predictions using historical data such as temperature, humidity, and air pressure [13]. Machine learning methods and deep learning techniques have proven their success in various fields and have the potential to improve numerical weather prediction models to produce faster and more accurate predictions [14].

Along with technology development, machine learning-based weather prediction methods continue to develop, ranging from traditional to complex methods [15]. Machine learning methods such as artificial neural network (ANN), gradient boosting, and Random Forest (RF) can be trained on historical data to detect patterns. For daily and weekly rainfall prediction, categorical boosting (CatBoost), extreme gradient boosting (XGBoost), and RF proved to be the best performers with high accuracy and good pattern capture ability [16].

Advanced technologies allow for the discovery of hidden patterns that might otherwise go unnoticed, making optimal use of available data [11]. In the context of improving global climate model projections using statistical downscaling, deep learning offers practical, scalable alternatives [17]. Nevertheless, each method comes with trade-offs in terms of accuracy, efficiency, and data requirements.

For high-stakes applications such as natural disaster prediction, extremely high accuracy is required, often under constraints of limited resources. Achieving this depends on the availability of both comprehensive climate data and sufficient computational capacity [18]. Effective forecasting is foundational to building robust predictive systems [19]. This makes it essential for researchers to carefully select suitable methods for each specific use case. Fortunately, access to meteorological data is no longer a significant barrier, with numerous publicly available datasets—both free and paid—offered by various platforms. In Indonesia, the meteorology, climatology, and geophysics agency (BMKG) provides weather data that can be accessed with varying levels of permission. However, access alone is insufficient; meaningful insights can only be derived through appropriate data processing techniques. The choice of method directly influences the quality of the output.

This paper presents a preliminary investigation into weather prediction using machine learning, serving as a baseline for future studies. The dataset includes weather observations from 192 stations across Indonesia, covering the period from 2010 to 2020. To the best of the authors' knowledge, no previous studies have applied machine learning to this specific dataset. While several Indonesian studies post-2020 have explored weather prediction, they either utilize different data or are not published in international languages. For example, [20] focused on Jakarta's historical weather trends; [21] used gated recurrent units (GRU)and XGBoost with data from the Juanda Meteorology Station in Sidoarjo; [22] applied support vector machine (SVM) to Perak's maritime data; and [23] used Extreme Gradient Boosting on data from Semarang. Other works, such as [24] and [25], focused on rainfall forecasting in Banten and clustering analysis of weather patterns in Denpasar, respectively. Studies like [26] and [27] also applied various algorithms including SVM, autoregressive integrated moving average (ARIMA), and long short-term memory (LSTM) to local weather data.

Several additional studies make use of BMKG data with different objectives. For instance, [28] emphasized dataset quality control without employing machine learning, while [29] and [30] applied machine learning to Himawari-8 satellite data. [31] analyzed weather sentiment using text data from X (formerly Twitter), and [32] applied machine learning techniques to earthquake detection. Earlier work, such as [33], predated the dataset used in this study. Other notable studies include [34], which combined satellite, radar, and rain gauge data for ensemble-based rainfall estimation, though these datasets are not publicly available.

This paper extends the authors' prior research [35], [36], [37], [38], [39], [40]. This research focuses on exploring a range of machine learning approaches—from traditional to modern—to optimize data utilization. These approaches are grouped into three categories: traditional, conventional, and modern. By applying different methods to the same dataset, the study demonstrates how methodological choices can significantly influence outcomes. Early tests using several machine learning models yielded promising but not outstanding results, prompting further investigation. Subsequent analysis of time series data revealed recurring annual patterns through data visualization. This observation led to the implementation of a year-on-year forecasting strategy. Data aggregation was also conducted on weekly, monthly, and quarterly scales to enhance pattern recognition and model performance.

The remainder of the paper is organized as follows: Section II describes the research methodology; Section III presents the results and discussion, including the data visualization process that guided the forecasting technique; and Section IV concludes with key findings and suggestions for future work.

A. Original Contribution

This research makes several original contributions, including:

- Exploring the utilization of weather station data provided by the Indonesian Meteorology and Geophysics Agency.
- Testing the application of this data to predict weather using a classification approach, which incorporates three methods: traditional, conventional, and modern.
- Evaluating the effectiveness of the data for forecasting by implementing three different approaches: statistical methods, machine learning, and deep learning.
- Highlighting the importance of data visualization prior to further utilization.

This contribution has not been addressed in previous studies, particularly due to the specificity of the data available.

II. Method

This study comprised several key stages, including data collection, preprocessing, model training, and performance evaluation. The dataset used in this research was sourced from the BMKG, an open-access public data provider. The dataset included essential meteorological parameters such as temperature, humidity, rainfall, wind direction, and solar intensity.

Data preprocessing involved handling missing values using both backward fill and forward fill techniques, followed by normalization to ensure consistency and improve model performance. An overview of the research workflow is illustrated in Fig. 1.

Following preprocessing, the cleaned dataset was used to train three machine learning models, each representing a different generation of predictive approaches. These include: k-nearest neighbors (KNN), representing traditional machine learning, SVM, representing conventional techniques, and LSTM, representing modern deep learning methods.

After training, each model was evaluated based on its predictive performance using four key metrics: accuracy, precision, recall, and F1-score. These metrics were used to identify the most effective model among KNN, SVM, and LSTM. Additionally, variations in the dataset were tested to examine their influence on model performance and generalizability.

A. Dataset Description

The data used in this study were sourced from two publicly available datasets on Kaggle. The primary dataset was "Climate Data Daily IDN" (IND), available at [41], contained daily weather observations from 192 weather stations across Indonesia, spanning a 10-year period from 2010 to 2020. The

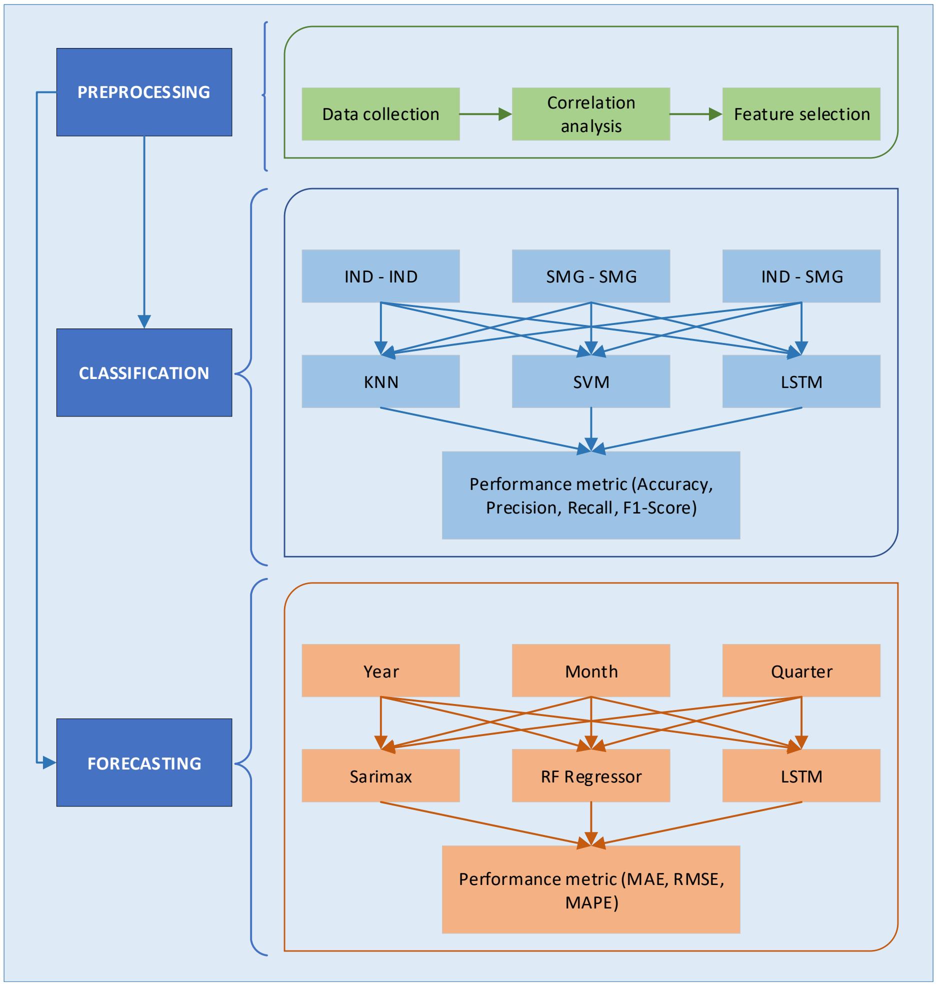


Fig. 1. Research diagram.

TABLE I. FEATURE DESCRIPTION

Feature	Description
Tn	min temperature (°C)
Tx	max temperature (°C)
Tavg	avg temperature (°C)
RH_avg	avg humidity (%)
RR	rainfall (mm)
ss	duration of sunshine (hour)
ff_x	max wind speed (m/s)
ddd_x	wind direction at maximum speed (°)
ff_avg	avg wind speed (m/s)
ddd_car	most wind direction (°)
station_id	station id which record the data. Detail of the station can be found in station_detail.csv
RainToday	rain status

second dataset was "Semarang daily climate data 2020-2023" (SMG), accessible at [42], provided daily weather station data from Semarang, Central Java, covering the years 2020 to 2023.

As summarized in Table I, the combined datasets consist of 11 features and a total of 590,261 observations—589,266 from weather stations across Indonesia between 2010 and 2020, and 1,355 from the Semarang station between 2020 and 2023. Fig. 2 illustrates the dataset characteristics. To facilitate classification-based weather prediction, a new binary feature, RainToday, was derived by converting the numerical rainfall measurement into categorical values ("Yes" if rainfall occurred, "No" otherwise). Similarly, the RainTomorrow label was generated by shifting the RainToday feature forward by one day, representing the target variable for next-day rain prediction.

B. Data Preprocessing

Data preprocessing was done to clean, organize, and prepare raw data before being used to build machine learning models on our data. This stage included removing irrelevant data features such as date and station ID [16]. This step was carried out in several stages, including filling in missing values, outlier removal, normalization, and quantile transformation. Performing data cleaning of unmeasured data by filling the empty data. A normalization process was required to achieve

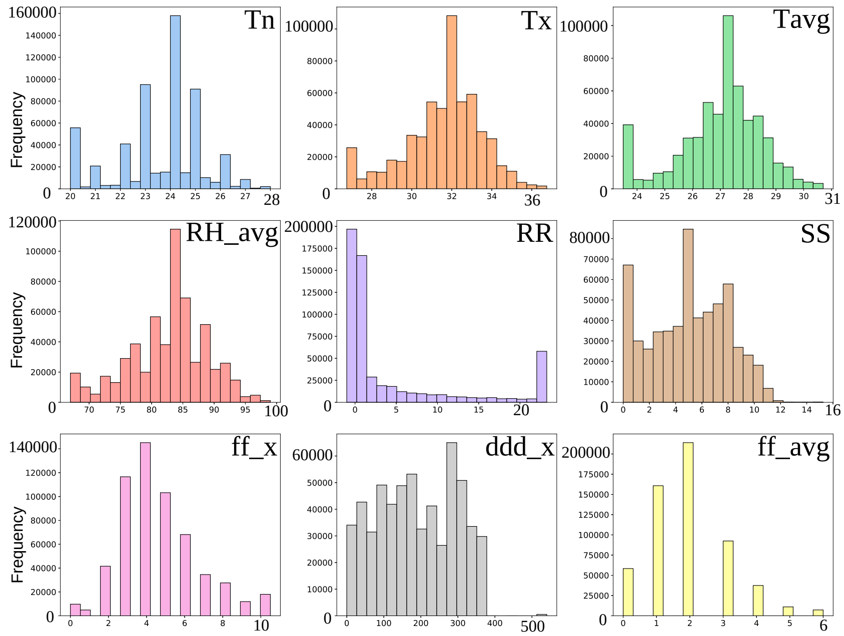


Fig. 2. Characteristics of the data collected by BMKG over a 10-year period at 192 weather stations in Indonesia.

equal weight in each feature. In the research, this step aimed to ensure that the difference in scale between features did not affect the model's performance. For example, if a feature has a much more extensive range of values than others, specific machine learning algorithms, such as KNN or SVM, may give the feature a disproportionate weight. In addition, quantile transformation was performed to obtain normally distributed data to remove the skewness of the data to become uniform, reducing the effect of outliers and improving model performance [43].

C. Correlation Analysis

Correlation analysis is a method used to determine the relationship between two variables. This correlation value shows the strength of the relationship between the two and how the two variables influence each other. This step is one way that can be used in addition to the chi-square approach to reduce the size of the feature dimension in machine learning training. In this research, this step was implemented by only training the model with closely related features.

D. Feature Selection

Feature selection is a processing technique used to reduce the dimension of features. This technique is done by selecting the features that have the most influence on the target to be achieved. The goal is to reduce the training dimension so that it can improve the model's capabilities and speed up the training process. Some methods, such as correlation analysis and chi-square, were used to find the most influential features.

E. Data Train and Test

This study's data underwent partitioning into training and testing sets through the k-fold cross-validation technique where k equals 10. The selection of this method was driven by its ability to deliver dependable model performance assessments while minimizing potential biases during data partitioning [43]. The method partitioned data into multiple subsets (folds), enabling each data point to serve as training and testing data. Employing this method mitigates potential bias from utilizing a single data division. The evaluation results from each fold underwent averaging to achieve a more stable and reliable overall model performance than what single evaluations provide.

K-fold cross-validation enables model testing across multiple data subsets, which helps determine if the model overfits or generalizes effectively to new data. This method examines multiple data combinations to decrease the typical variance seen in evaluation results when training and testing rely on a single data subset. This method allows for a detailed assessment of model performance across diverse datasets to confirm its reliability.

The methodologies applied in training and testing for classification models diverge from those used in forecasting models. The forecasting model employed time-based split and backward chaining methods to ensure data division while preserving chronological order. The study applied a temporal division method where data from 2010 to 2018 serves for training while data from 2019 to 2020 is used for testing. The backward chaining approach involved testing newer data after training with older data, and then the training dataset expanded backward while maintaining the same testing period.

F. Classification and Forecasting Model

This study implemented several algorithms: KNN, SVM, LSTM, SARIMAX, and RF regressor. Classification experiments included KNN, SVM, and LSTM, while forecasting experiments tested SARIMAX, RF regressor, and LSTM.

KNN is a traditional machine learning algorithm that can be used in classification and regression cases. This algorithm compares the new data group to the existing dataset. The difference is measured using distance metrics such as Euclidean, Manhattan, or Minkowski. The grouped data is measured by its distance from several nearest neighbors. Most classes from several k are used as the grouping result. The k value determined at the beginning greatly influences the results that will be issued. A smaller k value can cause the model to be more sensitive to differences in existing data. At the same time, when the value is too large, the model will have difficulty recognizing existing patterns.

SVM is a classification algorithm that finds the best hyperplane to separate categories in a dataset [44]. A hyperplane is a line that maximizes the difference or distance that divides between classes. Data that has little difference with the hyperplane is called a support vector, where this value plays a vital role in determining the position of the hyperplane. Unlike other machine learning approaches, SVM uses kernel functions to convert nonlinear problems into linear problems and reduce the complexity of mapping. SVM reduces the complexity of mapping and transforms nonlinear issues into linear problems using kernel functions, in contrast to other machine learning techniques. The advantage of SVM in classifying is its ability to handle high-dimensional datasets and is more resistant to overfitting conditions.

LSTM is an improvisation of the recurrent neural network (RNN) designed to learn patterns in sequential data, such as time series, text, or signals. LSTM was developed to overcome the vanishing gradient problem in standard RNN models and capture the data's long-term effects [45]. LSTM is helpful in various applications such as time series prediction, text analysis, and speech recognition [46]. However, getting optimal results requires much data and extended training.

SARIMAX is a statistical model used to predict datasets with a time sequence. This model further develops ARIMA model, which shows the influence of seasonal patterns and external variables (exogenous variables). SARIMAX combines auto-regressive components that can capture the relationship between current data and past data. Integrated, which eliminates the tendency of data changes. And moving average, which estimates the relationship between observation results and errors that occur in values some time ago [47]. The seasonal component owned by SARIMAX considers recurring patterns that occur at specific intervals to suit seasonal data such as weather. In addition, external variables allow the model to combine the influence of external factors to be more accurate in predicting complex scenarios.

RF regressor is an ensemble learning-based machine learning algorithm that combines multiple decision trees to improve its ability to predict values [48]. In handling the regression task, RF predicts a continuous target variable by averaging the output of each decision tree. Each tree is trained using random data to reduce overfitting and increase the generality of predictions to new datasets. This algorithm's superiority in learning complex interactions between variables makes it powerful in handling prediction tasks.

G. Model Evaluation Metrics

To compare the capabilities of each model, the model's capabilities need to be measured using several standard metrics. This study used two different models, classification and forecasting, which had different standards for measuring performance. The classification model aimed to determine data into specific categories, so the evaluation was based on accuracy and probability. This was done by measuring each model's accuracy, precision, recall, and F1-Score values. On the other hand, forecasting models aimed to predict values based on historical data, so their evaluation was done using error-based metrics [49]. This research used three metrics: mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The three metrics assessed the amount of deviation that occurs between the predicted value and the actual value.

III. Result and Discussion

A. Dataset Statistics

Fig. 2 shows the characteristics of the data collected by BMKG over 10 years at 192 weather stations in Indonesia. Rain today was based on rainfall values, assuming that more than 1 mm characterizes rainy conditions. The RainTomorrow feature was the target feature, was obtained by looking at the RainToday feature the next day. Some features had a slope in their data distribution. This slope affected the model's ability to predict a value—a skewed data distribution results in inconsistent value changes.

A feature can have a strong or weak effect on the target feature. In linear model prediction, the data must have normally distributed data to assume the value correctly. Each change point in the independent variable has a different value in non-normally distributed data. Sometimes, the independent variable can have a significant or negligible effect on the target variable. In addition, normalization of feature values is critical to ensure that each feature has equal weight. This is to avoid features that are too strong due to their values being too high, so the value scale for each feature must be uniform, which can be a scale of 0-1 or 0-100.

B. Data Preprocessing

1) Missing values: Missing values in the data affected the learning process and impair the model's accuracy. This needed to be addressed to ensure optimal results. This was addressed by using imputation techniques such as forward or backward and median values. This step helped fill in the blank values so the data becomes complete. Applying backward and forward techniques was based on various data covering weather conditions across Indonesia. Indonesia's vast territory includes water areas, mountains, highlands, lowlands, and various other natural conditions that impact the diversity of climatic conditions. The backward and forward techniques ensure that filling in missing values refers to the same station data rather than general weather conditions in Indonesia.

2) Data cleaning: Data cleaning was done by removing features that were less relevant to use. This stage dealt with handling missing values. The correct action could produce an efficient dataset for machine learning. The dataset was

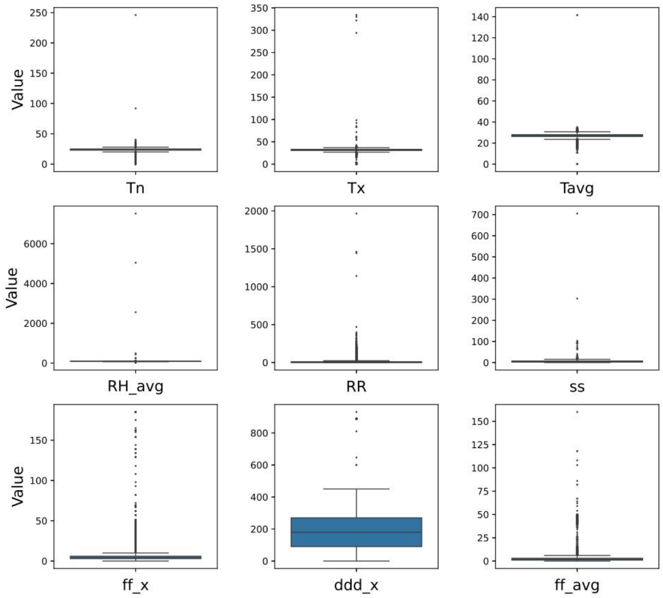


Fig. 3. Feature boxplot before removing outliers.

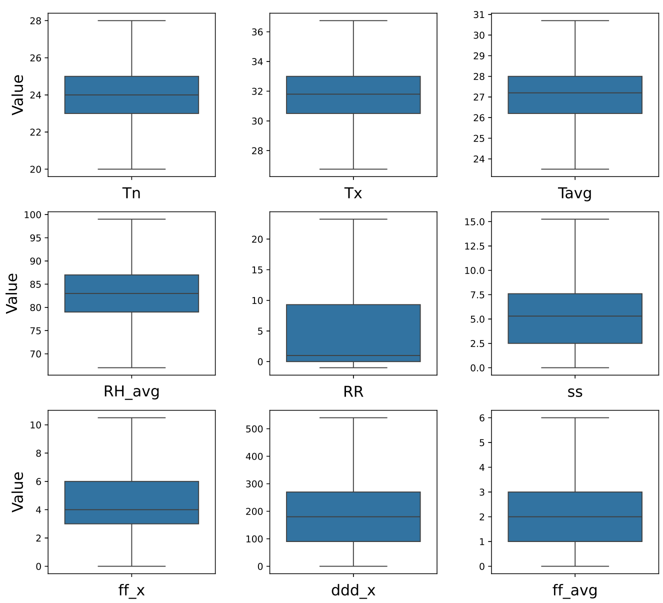


Fig. 4. Feature boxplot after removing outliers.

customized according to the model used. In forecasting models, time information was crucial. In addition, more effort needed to be made to ensure that there was no missing data in the time sequence to achieve optimal forecasting results. This process also handled outliers that appear in the dataset. Using the boxplot, we can see the values defined as outliers. This was addressed using the interquartile range approach. Fig. 3 and Fig. 4 show the condition before and after outlier removal. The data values after deletion were much more reasonable than before. This ensures the dataset was used in a wide range of conditions.

3) Normalization: Fig. 5 shows the difference between before and after preprocessing. Each weather feature had its range of values. For example, between temperature and humidity, in this condition, humidity had a much larger value than temperature; this needed to be avoided to ensure each feature had equal weight. Large values had a significant impact on the resulting model. To avoid features that were too strong because their values were too high, the value scale of each feature was normalized to a 0-1 scale.

4) Transformation: Some machine learning algorithms can perform better when the numeric input variable in the regression case has a standard probability distribution, such as Gaussian (normal) or uniform distribution. This approach is necessary to produce optimal output, especially in the case of classification. The transformation used was a quantile transformation that can transform numerical data into a normal distribution. For this reason, a transformation was performed to generalize the training dataset and solve it accurately using the linear approach. Fig. 5 shows that the transformation stage changes the distribution of previously widened data to the sides or has skewness to be concentrated in the center and form a bell.

C. Correlation Analysis

Correlation analysis illustrates the relationship between good features; this step provides an overview of features that contribute to the target variable, as in Fig. 6 and Fig. 7. Fewer data features provide an advantage in streamlining training time so that the process can run faster. The range of correlation values is between -1 and 1; 1 indicates the feature shows a powerful relationship, while a value of -1 illustrates a very weak relationship. Based on the evaluation using matrix correlations, the features that were most related to the target feature of tomorrow's rain conditions were today's rain conditions (RainToday), today's average humidity (RH_Avg), today's rainfall (RR) and the direction of the wind (ddd_x). Features with low correlation values can be considered not to be used because they do not significantly impact the final result to reduce the dimensions of the features to shorten the training process.

D. Feature Selection

Correlation analysis is one of the approaches in feature selection. Another approach used is chi-square; the chi-square score shows the value of the degree of importance of the feature to the target feature as in Fig. 8. In this case, the three features with the highest importance were RainToday, RR, and RH_avg. The order of these features was similar to the correlation analysis results. Thus, the condition of rain tomorrow was strongly influenced by these three features. Based on correlation analysis and chi-square value, the number of input features was reduced by focusing only on the RainToday, RR, and RH_avg features in predicting tomorrow's conditions.

E. Classification Model Performance

The dataset, after undergoing various preprocessing techniques, was input into several machine learning models for training and evaluation. To ensure the validity and robustness of the evaluation process, a k-fold cross-validation technique was employed, dividing the data into 10 folds. This method

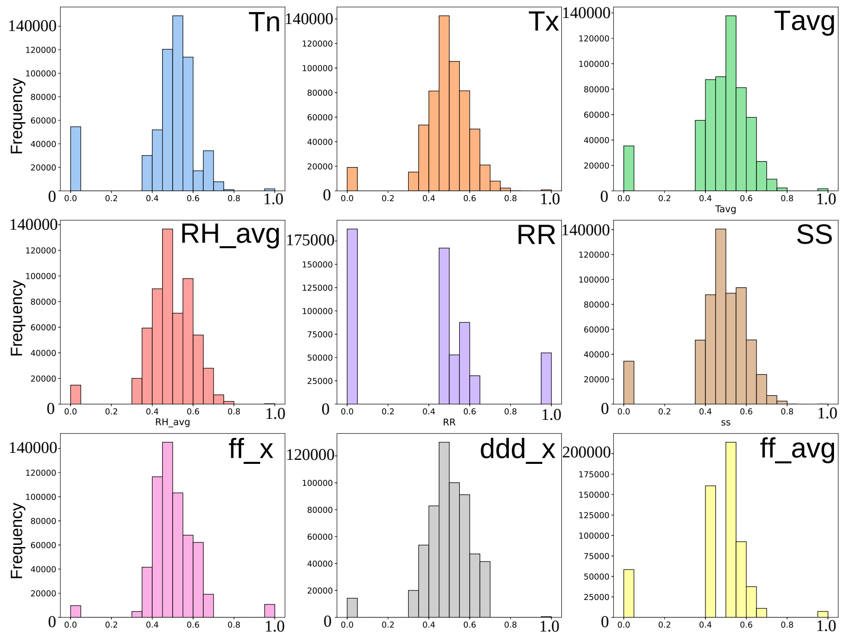


Fig. 5. Characteristics of the data collected by BMKG over a 10-year period at 192 weather stations in Indonesia after preprocessing.

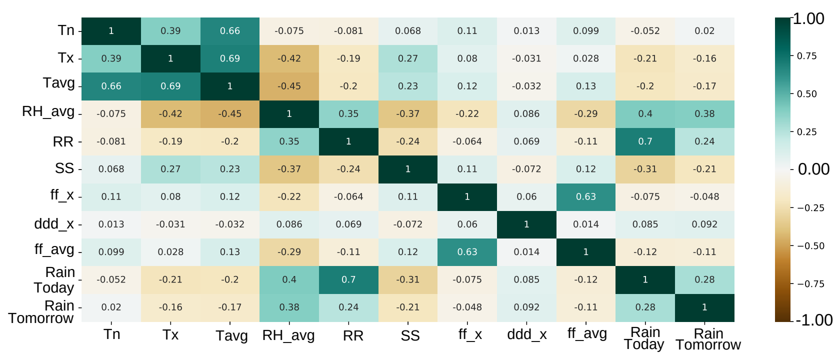


Fig. 6. Correlation heatmap before normalization and transformation.

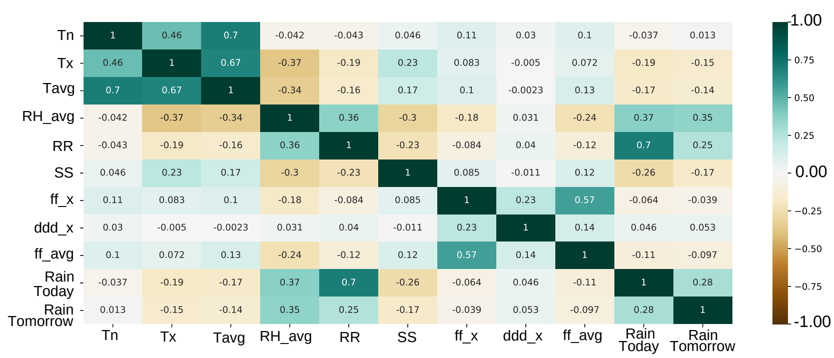


Fig. 7. Correlation heatmap after normalization and transformation.

was applied to both the IND and SMG datasets. However, in the merged condition (INDtoSMG)—where the IND dataset serves as the training set and the SMG dataset as the test set—k-fold cross-validation was not used.

Model performance was assessed using standard classification metrics: accuracy, precision, recall, and F1-score. The results, summarized in Table II, show that differences in model performance across dataset variations were not

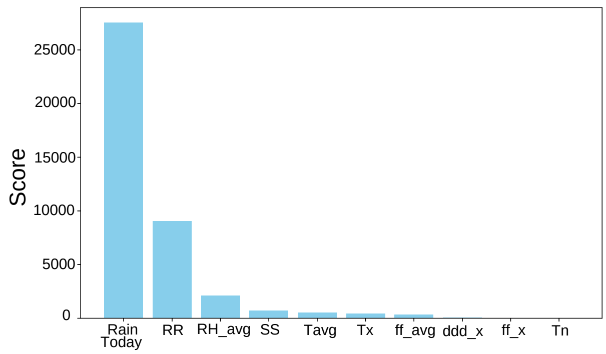


Fig. 8. Feature selection using chi-square scores.

substantial. This indicates that, in general, the classification capabilities of the models are comparable. However, a notable exception was observed in the KNN-INDtoSMG model, which achieved the highest accuracy score of 0.811. In this scenario, the model was trained on the IND dataset and tested on the SMG dataset, demonstrating that historical data can be effectively used to predict future outcomes.

TABLE II. Classification Model Performance

Method	Dataset	Accuracy	Precision	Recall	F1-Score
KNN	IND	0.661	0.597	0.546	0.570
	SMG	0.663	0.630	0.605	0.616
	INDtoSMG	0.811	0.790	0.768	0.779
SVM	IND	0.686	0.636	0.637	0.637
	SMG	0.672	0.655	0.570	0.608
	INDtoSMG	0.685	0.636	0.635	0.635
LSTM	IND	0.692	0.661	0.588	0.622
	SMG	0.657	0.642	0.545	0.585

The KNN operates by identifying the K most similar data points in the training set and predicting the target class based on the majority class of these neighbors. Choosing an odd value of K is crucial to avoid tie situations and ensure a deterministic outcome. The strong performance of KNN suggests that the algorithm successfully captured data patterns relevant to prediction, particularly in the INDtoSMG transfer.

The success of KNN in the INDtoSMG scenario might also be attributed to the inclusion of Semarang (SMG) weather data within the broader IND dataset. Although the datasets were collected in different years, the underlying weather patterns in the Semarang region may have remained sufficiently consistent, allowing the model to generalize well.

In principle, modern methods—including those based on ANN and deep learning—are expected to outperform traditional approaches due to ongoing advancements. These methods have demonstrated superior accuracy across numerous applications. However, the observed insignificant performance difference between conventional and modern models in this study raises important considerations.

Model performance is influenced not only by the architecture but also by the quality of the dataset and the features used. In many cases, feature selection plays a decisive role in achieving high accuracy. Generic features that do not reflect specific weather phenomena can lead to suboptimal predictions. Moreover, employing models that are misaligned with the problem type may further degrade performance. For instance, the LSTM model, typically suitable for time-series forecasting, was included in classification-based comparisons, potentially contributing to its relatively poor performance.

Weather prediction is inherently complex due to the highly dynamic and non-linear nature of atmospheric systems. The three key features used in this study—RainToday, RR (rainfall), and RH_avg—are all related to moisture content, a crucial indicator in rain prediction. However, these surface-level variables are outcomes of more complex environmental interactions. For example, humidity is influenced by temperature, which in turn depends on solar radiation, itself affected by factors such as cloud cover and time of day. These interdependencies are difficult to capture from surface-level observations alone.

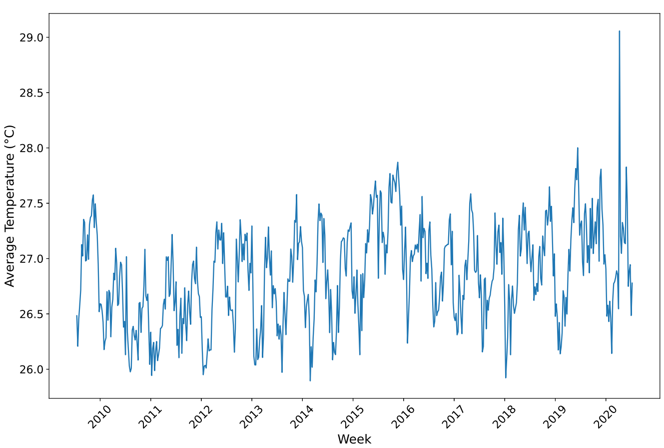
Another limitation is the temporal resolution of the dataset. The data used in this study was collected on a daily basis, meaning that only daily summaries of weather conditions are available. Such granularity may overlook important short-term fluctuations—e.g., shifts in humidity or temperature—that could occur within a single day and are crucial for accurate rain prediction. These intra-day variations are likely to be masked in daily aggregates.

Improving prediction accuracy may require either the acquisition of higher-resolution (e.g., hourly) data or the application of advanced feature engineering techniques. For instance, incorporating temperature or humidity at specific times of day (e.g., 6 a.m. or 5 p.m.) could provide a more nuanced representation of the daily weather cycle. However, implementing such strategies is constrained by limited access to fine-grained climate data.

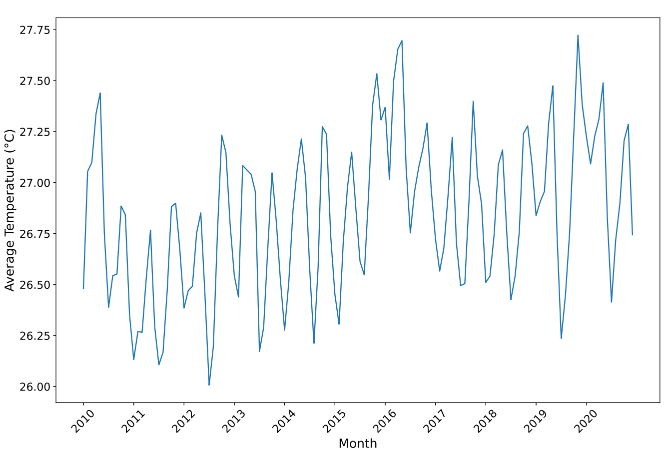
This is seen from the results obtained; no matter how good the model is built or how sophisticated the approach is, the features used remain one of the crucial factors that determine the ability of machine learning. Representation learning is an essential first step in finding machine learning solutions to some instances. In utilizing data, the proper representation, the right technique, and the correct model tuning are mandatory to achieve optimal solutions.

F. Visualization of Repeating Patterns

Utilization of weather datasets using a classification approach produced poor results based on the accuracy value. The reason is that the weather features used are too general. More specific weather features with a smaller scale are needed for better results. For this reason, the data was utilized using other approaches that may obtain better results. This section discusses the exploration of machine learning models in forecasting weather parameters. The exploration was conducted on various time scales, ranging from daily, weekly, monthly, quarterly, to yearly, to understand how the model adapts to changes in patterns over different periods. In this experiment, the dataset used had a daily resolution, so it was necessary to aggregate the data to a broader time range to suit the needs of the analysis. The aggregation process was carried out by calculating each weather parameter's average value or sum at various predetermined time scales. Fig. 9(a) shows a graph of the average temperature from the results of this aggregation, which shows a recurring pattern that occurs every year. This pattern shows a tendency that arises or a seasonal trend so that it can be used to improve forecasting accuracy. The model can capture changes that occur more effectively through this understanding to produce more accurate forecasts. This tendency was also seen on a monthly time scale, as in Fig. 9(b) The graph shows that temperature tends to increase at the beginning and end of the year. At the same time, it tended to decrease in the middle of the year. This pattern shows that weather forecasting is possible if the model can learn trends that occur over time, which is the goal of machine learning. With this foundation, the authors applied machine learning to study the patterns and build models to predict future temperatures. This exploration used various approaches, from statistical models such as SARIMAX to deep learning-based models such as LSTM.



(a)



(b)

Fig. 9. Average temperature dynamics for 2010-2020 on a weekly (a) and monthly (b) scale.

G. Forecasting Model Performance

To evaluate the performance of each model, this research used MAE, RMSE, and MAPE evaluation metrics to measure the error rate between the predicted and actual values. In addition, we investigated the effect of the amount of historical data on model accuracy by varying the dataset during the training process. This variation was done by gradually adding historical data, starting from 2018 to 2010, as shown in Fig. 10. Through this approach, we aimed to understand to what extent the amount of historical data can improve the model's predictive ability and reduce the error rate in temperature forecasting.

Fig. 10(a) shows the results of temperature forecasting using the SARIMAX model using 2018-2019 data in the training process to predict 2020 data. The blue line was the data used in the training process, the red line was the predicted data, and the green line was the actual data. Based on the pattern of training data values, it can be expected at a glance that the maximum and minimum values will widen following the training data pattern. It turns out that the predicted data does not match the actual data, but the model can capture the patterns.

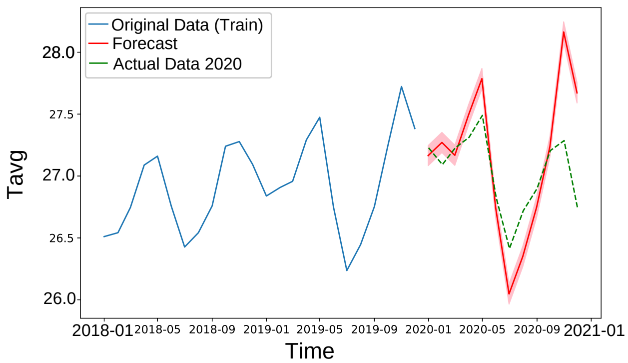
Fig. 10(b) shows the results of temperature forecasting using the SARIMAX model based on the 2010-2018 dataset. With more data in the training process, the model must recognize more common patterns in each period. However, this can also increase the risk of overfitting or underfitting if the model is not well-regulated. Based on the results shown, the model can identify trends and patterns, although there is still a slight deviation in the prediction of temperature values. Further parameter optimization can help improve the model's accuracy and stability in capturing the temperature change dynamics.

The results of the dataset variation in nine experiments are shown in Table III. From these results, it can be seen that the amount of data used affected the performance of the model. Adding historical data improves forecasting accuracy by allowing the model to recognize long-term patterns better. However, in some cases, the model experienced underfitting, which was characterized by higher error values due to its inability to capture complex patterns in the data. Therefore, choosing the optimal amount of data becomes crucial in achieving a balance between generalization and prediction accuracy.

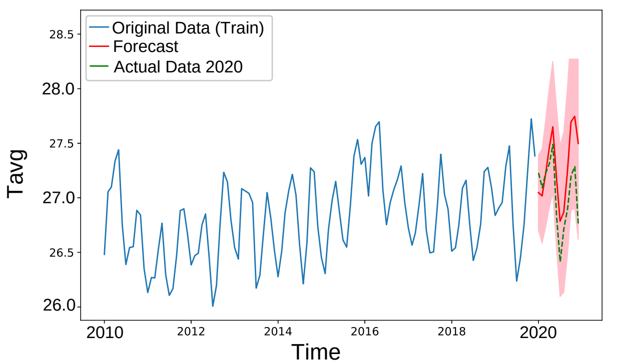
H. Forecasting Year-Over-Year Model Performance

Fig. 9 shows the temperature changes from 2010-2020 that form a pattern from year to year. Based on that, this research utilized the existing data for a broader period and compare it on a Year on Year (YoY) basis. The daily data was aggregated into weekly, monthly, quarterly, and annual data. The data was then fed into the SARIMAX, RF regressor, and LSTM models. This approach provides much better prediction results, as shown in Table III. Using annual data from 2010-2018, the 2019-2020 weather data could be accurately predicted with an error value of less than 5%. To further confirm these results, we forecasted monthly and quarterly data at 192 stations in Indonesia.

Fig. 11 shows the visualization of the experimental results. The results showed that the evaluation values were stable

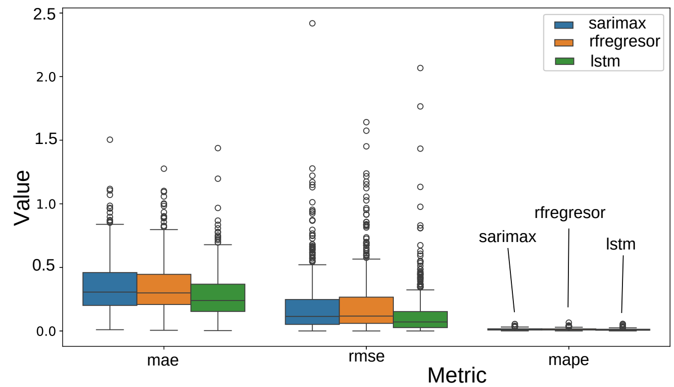


(a)

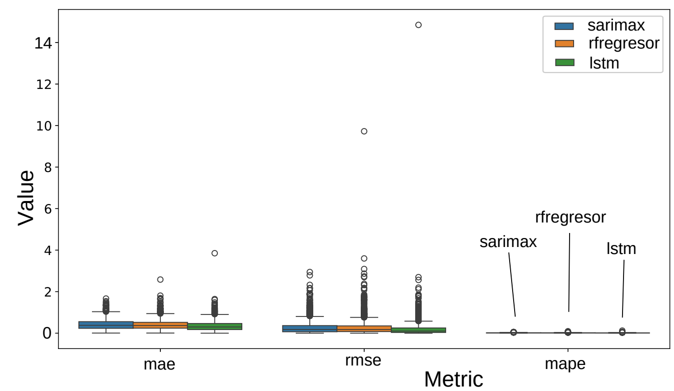


(b)

Fig. 10. Temperature forecasts using the SARIMAX model with different amounts of historical data.



(a)



(b)

Fig. 11. Forecasting performance of the SARIMAX, RF regressor and LSTM models using quarterly (a) and monthly (b) data.

across various data variations, so the model performance was stable. Like the classification model, the three forecasting models performances did not significantly difference, while the LSTM model achieved the best performance based on the MAPE parameter.

The evaluation results indicate that the SARIMAX model can recognize temperature change patterns from historical data quite well. The model successfully made predictions with an error rate of 4% with an average difference to the actual value of 1.14 degrees Celsius, as seen in Table III. These results indicated that SARIMAX can predict long-term temperature reasonably well. Although errors still occur, this can be reduced with further optimization.

Based on performance measurements, the evaluation shows that the RF model could recognize the pattern better than SARIMAX. The model had an error rate of 3% with an average difference of 0.8 degrees Celsius. These results indicated that the RF model can predict temperature more accurately because of its advantage in recognizing non-linear relationships. Further optimization includes selecting more relevant features and setting parameters that may improve the capabilities of this model.

The evaluation results of the LSTM model show that the model could capture temperature patterns and produced predictions with an error of 3% with an average difference of 0.87 degrees Celsius. This indicates that LSTM can utilize long-term information better than other models because its architecture is designed to handle time series data with long-term dependencies.

In addition, researchers varied the amount of historical data used in the training process, which was carried out gradually from one year earlier. The aim was to determine the effect of the amount of historical data on prediction accuracy. The results showed that model performance increased with the increase in the historical data used in the training process.

However, although adding historical data increases accuracy, several aspects must be considered. LSTM models require more extended training than statistical and decision tree-based models such as SARIMAX and RF. In addition, choosing the optimal amount of historical data is crucial, as too much data can cause the model to become too complex and prone to overfitting.

Using daily data to predict daily weather over the next year resulted in a significant error rate, where the expected pattern deviated significantly from the actual data. This indicates that daily data can directly affect the model's ability to capture weather patterns, especially since high daily fluctuations can make recognizing long-term trends with wider data variations challenging.

Based on the analysis of the weather data in time sequence using visualization, it is seen that the weather parameter values show a sinusoidal pattern that repeats within a year, reflecting the existence of seasonal trends that can be utilized for forecasting. This finding inspires an alternative approach to predicting using the YoY method, where weather data for the same period in previous years is used as a reference to predict values for the following year in the same week, month, or quarter.

To apply this approach, daily data was first aggregated into broader weekly, monthly, quarterly, and yearly time scales. This experiment focused on monthly, quarterly, and annual data, where weather data from the same month or quarter from previous years were used as training data. This approach aimed to improve the stability of the model by reducing the noise from daily fluctuations and emphasizing more obvious seasonal patterns.

With better results, the experiment was conducted on one weather feature, TX, at one weather station in Indonesia. The YoY mechanism was applied to 192 stations in Indonesia to validate the model performance further. With limited resources, the experiment was conducted on only one feature by forecasting monthly and quarterly using three machine-learning models. The results of the experiment are summarized in Fig. 11 and Table IV.

Based on Table IV, LSTM performs best overall, especially on monthly data with a MAPE value of 0.012744, demonstrating its ability to model complex patterns. LSTM also excels on annual and quarterly data with MAPE values of 0.006747 and 0.010244, respectively, making it the most reliable method for various time scales. SARIMAX, as a statistical model, excels in annual data with a MAPE of 0.008620, showing its superiority in capturing obvious seasonal patterns. However, there is a decrease in performance on monthly data because it is difficult to learn more complex patterns. Meanwhile, RF Regressor can show reasonably consistent results on each time scale, showing the best performance on quarterly data but not as good as LSTM.

LSTM is the most accurate method, especially on quarterly and monthly data. The SARIMAX model is more suitable for annual data with more apparent seasonal patterns. At the same time, RF Regressor can be used as a simpler alternative but is not as accurate as LSTM.

Based on the results obtained, the performance of each model can be seen even though it has almost the same performance. In terms of time scale, the error that occurs on a quarterly scale is much smaller than on a monthly scale. This is because the pattern of value changes on a quarterly scale tends to be more stable than the changes that occur on a monthly scale, which tends to be more volatile.

The results show that the YoY approach significantly impacts forecasting performance. The idea was obtained after the weather data visualization stage. Standard pre-processing processes such as missing value, transformation, and normalization are insufficient to direct the machine to learn. The right approach is needed to obtain optimal results. This is obtained through a good understanding of the data. Data visualization is essential in obtaining ideas to direct the machine to produce optimal solutions, as shown in Fig. 9.

TABLE III. FORECASTING MODEL PERFORMANCE

Method	MAE	RMSE	MAPE
SARIMAX	1.142	1.152	0.037
RF regressor	0.796	1.075	0.026
LSTM	0.861	0.874	0.028

TABLE IV. FORECASTING YEAR-ON-YEAR MODEL PERFORMANCE

Method	Series	MAE	RMSE	MAPE
SARIMAX	Year	0.229	0.088	0.009
	Quartal	0.345	0.187	0.013
	Month	0.422	0.286	0.016
RF regressor	Year	0.268	0.129	0.010
	Quartal	0.342	0.202	0.013
	Month	0.410	0.289	0.015
LSTM	Year	0.178	0.074	0.007
	Quartal	0.278	0.127	0.010
	Month	0.347	0.206	0.013

I. *Classification vs. Forecasting*

The choice between classification and forecasting depends on the available dataset and the purpose. Classification and forecasting approaches are often used to interpret future conditions, especially weather. Classification categorizes conditions like future weather, while forecasting provides continuous values like temperature or rainfall. Understanding the available data is an essential factor in determining the optimal solution. There is no fair way to compare classification and forecasting, only based on the desired goal. Based on the dataset available in this study, the forecasting approach is a more appropriate step in utilizing the available data. If forced to use the classification approach, the dataset may accurately predict the weather on a broader time scale, such as weekly, by including weather features on certain days.

IV. CONCLUSION

Weather forecasting is difficult due to the complexity of the interactions that occur. A machine learning approach with proper feature engineering and representation is essential to achieve maximum results. The results of classification and forecasting using three generations of machine learning show insignificant differences, so the initial stage of model development, such as feature engineering, is crucial. Model selection must also consider other factors, such as training time, computational efficiency, and ease of implementation. This aspect is essential in cases that require fast predictions. Understanding data is critical to support the proper and optimal use of data, which can be achieved through good dataset visualization. Classification approaches that perform poorly due to overly general features lead to the use of data for forecasting so that the data can have better benefits. In forecasting time series data such as weather predictions, the data collection or observation time scale needs to be considered. Forecasting over a broader period provides higher accuracy due to generalization while forecasting over a smaller period has the potential to reduce accuracy due to increased noise or data variation. Regular sequential forecasting produces errors of up to 3%. However, changing a broader time period approach to YoY reduces the error to less than 1.6%. Although the model algorithm is essential to produce good accuracy, basic things such as feature engineering, feature selection, and the methods used are also crucial in creating a good model. In addition, data visualization is needed to support a better understanding of the data so that it makes an optimal approach. Further development can focus on using more sophisticated techniques to recognize features and combine various models to improve performance.

Building on the findings of this study, several directions for future research emerge. First, while the current work employs relatively basic machine learning models such as KNN, SVM, and LSTM, future studies could investigate more advanced or hybrid approaches—such as combining deep learning with statistical or ensemble methods—to enhance predictive accuracy. Expanding the dataset in both spatial and temporal resolution, for instance through hourly measurements or integration of multi-source data from satellites, radar, and reanalysis products, could further improve model performance. Moreover, the present feature set focuses primarily on surface-level variables; incorporating atmospheric parameters, remote sensing indicators, and applying advanced feature engineering could better capture complex climatic interactions. The limited richness of the dataset may also contribute to the underperformance of deep learning models, suggesting that data augmentation techniques or transfer learning from global climate datasets could be beneficial. Hybrid algorithms such as LSTM + SARIMAX, CNN-LSTM, or RF + ARIMA should be tested to leverage both temporal dependencies and nonlinear relationships. The small performance gap between models points to possible bottlenecks in feature selection and preprocessing, warranting a more systematic analysis of feature importance. Given that the best classification accuracy achieved is approximately 0.81, further hyperparameter optimization—using Bayesian or evolutionary search methods—could help unlock additional performance gains. Additionally, the exploration of seasonal decomposition techniques, such as STL decomposition, prior to forecasting could improve signal extraction.

Acknowledgment

This work was supported in part by Sebelas Maret University under Grant No. 228/UN27.22/PT.01.03/2023. The experiments in this study were conducted using Python, with research assistant Rahmat Rohmani providing support in executing the program code.

References

[1] D. Didiharyono, G. Giarno, and S. Sapareng, "Changes in rainfall intensity, rising air temperature, wind speed, and its relationship with land use in makassar city," *JST (Jurnal Sains dan Teknologi)*, vol. 11, no. 2, p. 441–450, Aug. 2022. [Online]. Available: https://ejournal.undiksha.ac.id/index.php/JST/article/view/45604

[2] M. Anwar, S. Nugrohadi, V. Tantriyati, and V. Windarni, "Rain prediction using rule-based machine learning approach," *Advance Sustainable Science, Engineering and Technology*, vol. 2, no. 1, p. 0200104, 2020. [Online]. Available: http://journal.upgris.ac.id/index.php/asset/article/view/6019

[3] J. Helali, M. Nouri, M. Mohammadi Ghaleni, S. A. Hosseni, F. Safarpour, A. Shirdeli, P. Paymard, and Z. Kalantari, "Forecasting precipitation based on teleconnections using machine learning approaches across different precipitation regimes," *Environmental Earth Sciences*, vol. 82, no. 21, p. 495, Sep 2023. [Online]. Available: https://doi.org/10.1007/s12665-023-11191-9

[4] C. Lalitha and D. Ravindran, "Hybrid deep learning framework for weather forecast with rainfall prediction using weather bigdata analytics," *Multimedia Tools and Applications*, vol. 83, no. 25, pp. 66 413–66 432, Jul 2024. [Online]. Available: https://doi.org/10.1007/s11042-023-17801-9

[5] S. Kumari and P. Muthulakshmi, "Sarima model: An efficient machine learning technique for weather forecasting," *Procedia Comput. Sci.*, vol. 235, no. C, p. 656–670, Jan. 2024. [Online]. Available: https://doi.org/10.1016/j.procs.2024.04.064

[6] L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, and Z. Yang, "Machine learning methods in weather and climate applications: A survey," *Applied Sciences*, vol. 13, no. 21, p. 12019, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/21/12019

[7] M. Abdullah and S. Said, "Performance evaluation of machine learning regression models for rainfall prediction," *Iranian Journal of Science and Technology - Transactions of Civil Engineering*, dec 2024. [Online]. Available: https://link.springer.com/10.1007/s40996-024-01691-4

[8] L. Chen, X. Zhong, F. Zhang, Y. Cheng, Y. Xu, Y. Qi, and H. Li, "Fuxi: a cascade machine learning forecasting system for 15-day global weather forecast," *npj Climate and Atmospheric Science*, vol. 6, no. 1, p. 190, Nov 2023. [Online]. Available: https://doi.org/10.1038/s41612-023-00512-1

[9] J. Chen, H. Dai, S. Wang, and C. Liu, "Improving accuracy and efficiency in time series forecasting with an optimized transformer model," *Engineering Letters*, vol. 32, no. 1, p. 1–11, jan 2024.

[10] L. Parviz, K. Rasouli, and A. Torabi Haghighi, "Improving hybrid models for precipitation forecasting by combining nonlinear machine learning methods," *Water Resources Management*, vol. 37, no. 10, pp. 3833–3855, Aug 2023. [Online]. Available: https://doi.org/10.1007/s11269-023-03528-7

[11] T. Tang, T. Liu, and G. Gui, "Forecasting precipitation and temperature evolution patterns under climate change using a random forest approach with seasonal bias correction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 12 609–12 621, 2024. [Online]. Available: https://doi.org/10.1109/JSTARS.2024.3425639

[12] F. Rezaei Aderyani and S. Jamshid Mousavi, "Machine learning-based rainfall forecasting in real-time optimal operation of urban drainage systems," *Journal of Hydrology*, vol. 645, p. 132118, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169424015142

[13] P. S. Saketh, R. Rohit, and B. Suneetha, "Weather forecasting using machine learning," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, 2023, pp. 13–18. [Online]. Available: https://doi.org/10.1109/ICICT57646.2023.10134218

[14] C. Sengoz, S. Ramanna, S. Kehler, R. Goomer, and P. Pries, "Machine learning approaches to improve north american precipitation forecasts," *IEEE Access*, vol. 11, pp. 97 664–97 681, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3309054

[15] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine learning methods for weather forecasting: A survey," *Atmosphere*, vol. 16, no. 1, p. 82, 2025. [Online]. Available: https://www.mdpi.com/2073-4433/16/1/82

[16] V. Kumar, N. Kedam, O. Kisi, S. Alsulamy, K. M. Khedher, and M. A. Salem, "A comparative study of machine learning models for daily and weekly rainfall forecasting," *Water Resources Management*, vol. 39, no. 1, pp. 271–290, Jan 2025. [Online]. Available: https://doi.org/10.1007/s11269-024-03969-8

[17] K. Chattrairat, W. Wongseree, and A. Leelasantitham, "Comparisons of machine learning methods of statistical downscaling method: Case studies of daily climate anomalies in thailand," *Journal of Web Engineering*, vol. 20, no. 5, p. 1397–1424, Jul. 2021. [Online]. Available: https://journals.riverpublishers.com/index.php/JWE/article/view/4869

[18] M. Akbarian, B. Saghafian, and S. Golian, "Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over iran," *Journal of Hydrology*, vol. 620, p. 129480, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169423004225

[19] M. A. Abed, M. A. Imteaz, and A. N. Ahmed, "Utilising machine learning for pan evaporation prediction - a case study in western australia," in *International MultiConference of Engineers and Computer Scientists 2023 (IMECS 2023)*, vol. 2245, 2023, Conference paper, p. 14 – 18, cited by: 3. [Online]. Available: https://irepository.uniten.edu.my/handle/123456789/34493

[20] D. Maheng, B. Bhattacharya, C. Zevenbergen, and A. Pathirana, "Changing urban temperature and rainfall patterns in jakarta: A comprehensive historical analysis," *Sustainability*, vol. 16, no. 1, p. 350, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/16/1/350

[21] H. Darmawan, M. Yuliana, and M. Z. S. Hadi, "Gru and xgboost performance with hyperparameter tuning using gridsearchcv and bayesian optimization on an iot-based weather prediction system," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 13, pp. 851–862, 6 2023. [Online]. Available: https://doi.org/10.18517/ijaseit.13.3.18377

[22] D. P. Hapsari, M. I. Utoyo, and S. W. Purnami, "A prediction of rainfall data based on support vector machine with stochastic gradient descent," in *Proceedings of the 5th International Workshop on Innovations in Information and Communication Science and Technology*. CEUR-WS, 3 2020. [Online]. Available: https://ceur-ws.org/Vol-2627/short10.pdf

[23] M. T. Anwar, E. Winarno, W. Hadikurniawati, and M. Novita, "Rainfall prediction using extreme gradient boosting," *Journal of Physics: Conference Series*, vol. 1869, p. 12078, 4 2021. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1869/1/012078

[24] Y. Ruhiat, "Forecasting rainfall and potential for repeated events to predict flood areas in banten province, indonesia," *Journal of Measurements in Engineering*, vol. 10, pp. 68–80, 6 2022. [Online]. Available: https://doi.org/10.21595/jme.2022.22363

[25] B. Prihasto, D. Darmansyah, D. P. Yuda, F. M. Alwafi, H. N. Ekawati, and Y. P. Sari, "Comparative analysis of k-means and k-medoids clustering methods on weather data of denpasar city," *Jurnal Pendidikan Multimedia (Edsence)*, vol. 5, pp. 91–114, 12 2023. [Online]. Available: https://doi.org/10.17509/edsence.v5i2.65925

[26] A. L. R. Putri, "Performance analysis of the support vector machine algorithm in predicting rain potential in dki jakarta," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 3, pp. 671–676, 6 2024. [Online]. Available: https://doi.org/10.59934/jaiea.v3i3.490

[27] A. A. Akbar, Y. Darmawan, A. Wibowo, and H. K. Rahmat, "Accuracy assessment of monthly rainfall predictions using seasonal arima and long short-term memory (lstm)," *Journal of Computer Science and Engineering (JCSE)*, vol. 5, pp. 99–114, 8 2024. [Online]. Available: https://icsejournal.com/index.php/JCSE/article/view/843

[28] M. Putra, M. S. Rosid, and D. Handoko, "High-resolution rainfall estimation using ensemble learning techniques and multisensor data integration," *Sensors*, vol. 24, no. 15, p. 5030, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/15/5030

[29] V. Monita, S. Raniprima, and N. Cahyadi, "Comparative analysis of daily and weekly heavy rain prediction using lstm and cloud data," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, pp. 833–842, 12 2024. [Online]. Available: https://journal.uad.ac.id/index.php/JITEKI/article/view/30374

[30] F. Simanjuntak, I. Jamaluddin, T.-H. Lin, H. A. W. Siahaan, and Y.-N. Chen, "Rainfall forecast using machine learning with high spatiotemporal satellite imagery every 10 minutes," *Remote Sensing*, vol. 14, no. 23, p. 5950, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/23/5950

[31] M. F. Sodiq and F. Amin, "Sentiment analysis of bmkg weather information service using k-nearest neighbor method," *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 4, pp. 822–835, 8 2024. [Online]. Available: https://doi.org/10.35870/ijsecs.v4i2.2881

[32] I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in indonesia using 30 years historical data," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, pp. 1331–1342, 6 2020. [Online]. Available: https://doi.org/10.12928/telkomnika.v18i3.14756

[33] A. G. Salman, B. Kanigoro, and Y. Heryadi, "Weather forecasting using deep learning techniques," in *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2015, pp. 281–285. [Online]. Available: https://doi.org/10.1109/ICACSIS.2015.7415154

[34] I. D. G. A. Putra, H. Nimiya, A. Sopaheluwakan, T. Kubota, H. S. Lee, R. P. Pradana, M. N. F. Alfata, R. B. Perdana, D. S. Permana, N. F. Riama, and D. Karnawati, "Development of typical meteorological years based on quality control of datasets in indonesia," *Renewable Energy*, vol. 221, p. 119699, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148123016142

[35] F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Test collection recycling for semantic text similarity," in *ACM International Conference Proceeding Series*, 2012. [Online]. Available: https://doi.org/10.1145/2428736.2428784

[36] C. Rahmad, F. Rahutomo, M. A. Gustalika, and I. F. Rahmah, "Identification of sugarcane maturity scale based on rgb, gabor feature extraction and support vector machine," *IOP Conference Series: Materials Science and Engineering*, vol. 434, 12 2018. [Online]. Available: https://doi.org/10.1088/1757-899X/434/1/012065

[37] F. Rahutomo, A. S. Mulyo, and P. Y. Saputra, "Automatic grammar checking system for indonesian," in *2018 International Conference on Applied Science and Technology (iCAST)*, 2018, pp. 308–313. [Online]. Available: https://doi.org/10.1109/iCAST1.2018.8751591

[38] A. A. Septarina, F. Rahutomo, and M. Sarosa, "Machine translation of indonesian: a review," *Communications in Science and Technology*, vol. 4, pp. 12–19, 7 2019. [Online]. Available: https://doi.org/10.21924/cst.4.1.2019.104

[39] M. F. S. Thirafi and F. Rahutomo, "Implementation of naïve bayes classifier algorithm to categorize indonesian song lyrics based on age," in *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2018, pp. 106–109. [Online]. Available: https://doi.org/10.1109/SIET.2018.8693201

[40] F. Rahutomo, A. S. Mulyo, and P. Y. Saputra, "Automatic grammar checking system for indonesian," in *2018 International Conference on Applied Science and Technology (iCAST)*, 2018, pp. 308–313. [Online]. Available: https://doi.org/10.1109/iCAST1.2018.8751591

[41] "Climate data daily idn," https://tinyurl.com/bmkgIDN2020.

[42] "Semarang daily climate data 2020-2023," https://tinyurl.com/bmkgSMG2023.

[43] M. M. Hassan, M. A. T. Rony, M. A. R. Khan, M. M. Hassan, F. Yasmin, A. Nag, T. H. Zarin, A. K. Bairagi, S. Alshathri, and W. El-Shafai, "Machine learning-based rainfall prediction: Unveiling insights and forecasting for improved preparedness," *IEEE Access*, vol. 11, pp. 132 196–132 222, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3333876

[44] N. Hayaty, H. Kurniawan, M. R. Rathomi, F. Chahyadi, and M. Bettiza, "Rainfall prediction with support vector machines: A case study in tanjungpinang city, indonesia," *BIO Web of Conferences*, vol. 70, p. 1003, nov 2023. [Online]. Available: https://www.bio-conferences.org/10.1051/bioconf/20237001003

[45] C. Yin, L. T. Binh, D. T. Anh, S. T. Mai, A. Le, V.-H. Nguyen, V.-C. Nguyen, N. X. Tinh, H. Tanaka, N. T. Viet, L. D. Nguyen, and T. Q. Duong, "Advanced machine learning techniques for predicting nha trang shorelines," *IEEE Access*, vol. 9, pp. 98 132–98 149, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3095339

[46] R. Teixeira, A. Cerveira, E. J. S. Pires, and J. Baptista, "Enhancing weather forecasting integrating lstm and ga," *Applied Sciences*, vol. 14, no. 13, p. 5769, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/13/5769

[47] F. R. Alharbi and D. Csala, "A seasonal autoregressive integrated moving average with exogenous factors (sarimax) forecasting model-based time series approach," *Inventions*, vol. 7, no. 4, p. 94, 2022. [Online]. Available: https://www.mdpi.com/2411-5134/7/4/94

[48] T. A. Gahwera, O. S. Eyobu, and M. Isaac, "Analysis of machine learning algorithms for prediction of short-term rainfall amounts using uganda's lake victoria basin weather dataset," *IEEE Access*, vol. 12, pp. 63 361–63 380, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3396695

[49] P. St-Aubin and B. Agard, "Precision and reliability of forecasts performance metrics," *Forecasting*, vol. 4, no. 4, pp. 882–903, 2022. [Online]. Available: https://www.mdpi.com/2571-9394/4/4/48