ISSN 2355-3286

An Adaptive Stacking Approach for Monthly Rainfall Prediction with Hybrid Feature Selection

Ahmad Zulfa[1,2], Ahmad Saikhu [1], Hilmil Pradana[1], Irvan Budiawan[3,4]
[1] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] The Agency for Meteorology, Climatology and Geophysics (Badan Meteorologi, Klimatologi, dan Geofisika, BMKG), Indonesia
[3] Department of Electrical Engineering, Faculty of Engineering, Universitas Jenderal Achmad Yani, Cimahi, Indonesia
[4]Center for Instrumentation Technology and Automation, Institut Teknologi Bandung, Bandung, Indonesia
[1,2]ahmad.zulfa@bmkg.go.id, [1]saikhu@if.its.ac.id, [1]hilmi@its.ac.id, [3,4]budiawan.irvan@gmail.com

Accepted April 30, 2025
Approved June 24, 2025

***Abstract*— Accurate rainfall prediction is critical for effective water resource management, agriculture, and climate risk mitigation. However, the inherent non-linearity and variability of rainfall patterns present significant modeling challenges. This study proposes an Adaptive Stacking Ensemble framework for monthly rainfall prediction, enhanced by a Hybrid Feature Selection strategy. The feature selection integrates three techniques—Correlation Analysis, Feature Importance (Random Forest), and Recursive Feature Elimination (RFE)—using a voting mechanism to ensure robust and consistent feature selection. The ensemble framework employs a diverse set of machine learning algorithms, including Random Forest, K-Nearest Neighbors, XGBoost, AdaBoost, Decision Tree, and Linear Regression, as base learners. Meta-learners are selected adaptively based on empirical performance, with the three top-performing models—Linear Regression, AdaBoost, and XGBoost—evaluated individually and collectively through a voting-based stacking approach. This flexible strategy ensures the model captures both linear and nonlinear dependencies in the data. Experimental results show that while standalone Linear Regression achieved the highest individual accuracy (R^2 = 0.931), the best ensemble performance was attained using the voting-based stacking model, which combined the top meta-learners and achieved an R^2 of 0.917, SMAPE of 13.33%, MAE of 0.287, and RMSE of 0.339. These findings confirm the effectiveness of adaptively integrating multiple strong learners in enhancing model generalization and prediction reliability for climatological applications.**

Index Terms—Elimination; Feature; Importance; Learning; Rainfall; Stacking.

I. Introduction

The Riau Islands Province is a maritime region with a coastline stretching 2,367.6 km and a total area of 251,810 km², of which only 4% consists of land, while the remaining 96% is made up of water [1]. The climate type of Tanjungpinang City is classified as an equatorial tropical climate with relatively high rainfall throughout the year. During the period from 1991 to 2024, the maximum recorded rainfall occurred in January 2021, reaching 926.9 mm over an average of 22 rainy days in that month. Despite the significant rainfall, there is no distinct separation between the rainy and dry seasons. Tanjungpinang City is categorized as a Non-Seasonal Zone (Non-ZOM), meaning it does not have a clear seasonal boundary between wet and dry periods, and its climatic analysis does not follow the strict monsoon patterns typically observed in other regions of Indonesia.

Changes in the global climate system have a significant influence on local weather patterns and climate variability. These impacts can be observed through shifts in rainfall distribution, temperature extremes, and the increasing frequency of extreme weather events in various regions [2].

Weather and climate have fundamental differences. Weather refers to various processes occurring in the atmosphere at a specific time and location, reflecting the immediate state of the atmosphere and its short-term changes within a particular area [3].

The characteristics of rainfall in a region are important to understand in order to determine water availability and to identify potential issues and disasters related to water resources. Information about rainfall characteristics, including the identification of wet months, moist months, and dry months, is highly valuable for regional management. Through this understanding, the utilization of rainfall can be optimized while minimizing any potential negative impacts that may arise [4].

The unit used for measuring rainfall parameters is millimeters. According to the Meteorology, Climatology, and Geophysics Agency of Indonesia (Badan Meteorologi, Klimatologi, dan Geofisika — BMKG), rainfall can be understood as measuring the height of accumulated rainwater. If the rainfall is recorded as 1 millimeter, it means that on a surface area of one square meter, the water would reach a height roughly equivalent to the thickness of a fingernail, or about the same as a medium-sized bottled water (approximately one liter). In other words, if rainfall is measured at a location with a water depth of 1 millimeter, and the water is evenly distributed over a flat surface without evaporation or runoff, it would represent that amount. Runoff itself refers to the water that flows over the ground surface as a result of rainfall or other water sources that do not infiltrate into the soil.

Rainfall classification in this study refers to the standards set by the Meteorology, Climatology, and Geophysics Agency (BMKG), which categorizes rainfall intensity as follows: (1) cloudy if rainfall is 0 mm/day; (2) light rain between 0.5–20 mm/day; (3) moderate rain between 20–50 mm/day; (4) heavy rain between 50–100 mm/day; (5) very heavy rain between 100–150 mm/day; and (6) extreme rain if rainfall exceeds 150 mm/day. However, for the purposes of this study, the classification is simplified into two main categories: rain and no rain. The "rain" category includes any rainfall greater than 0 mm/day, while the "no rain" category applies when the rainfall is exactly 0 mm/day [5].

This study aims to develop a rainfall prediction model for Tanjungpinang City using an adaptive stacking ensemble learning approach combined with feature selection through the Voting Feature Selector method, where features selected by at least 2 out of 3 methods will be included in the base model. The model will then be evaluated using R^2, MAE, and RMSE. This model is expected to be a reliable solution for providing weather prediction information, particularly the rainy season forecast, and supporting better decision-making for various sectors dependent on weather conditions in the Tanjungpinang area.

Typically, stacking ensemble learning uses a combination of static models. With the adaptive stacking ensemble learning approach, the meta-learner is automatically selected based on the best performance of the base learners or initial results. With the adaptive stacking ensemble learning approach, the model is expected to deliver the best results as it adjusts to the previous training data

II. METHODOLOGY

The dataset used in this study is secondary data obtained from the III Class Meteorological Station Raja Haji Fisabilillah Tanjungpinang with official permission from the relevant authorities. In addition, global climate index data such as the Southern Oscillation Index (SOI) and the Indian Ocean Dipole (IOD) were sourced from trusted institutions, namely the National Oceanic and Atmospheric Administration (NOAA).

The dataset consists of 408 samples (34 years × 12 months) containing various climatological variables relevant to the prediction of monthly rainy seasons. The variables in this dataset are categorized as follows: Target (Dependent Variable): Rainfall (CH, in mm) → Represents the amount of monthly rainfall (curah hujan) to be predicted.

Predictor (Independent Variables): Air Pressure (hPa): Consists of two variations: P Mean Sea level pressure correction (MSL), P0 Land surface pressure correction (Stasiun). Air Temperature (°C): Consists of six variations: T07 (Temperature at 07:00), T13 (Temperature at 13:00), T18 (Temperature at 18:00), T (Temperature average), Tx (Temperature maximum), Tn (Temperature minimum). Relative Humidity (%): Consists of four variations: RH07 (Humidity at 07:00), RH13 (Humidity at 13:00), RH18 (Humidity at 18:00), RH (Humidity average). SSS (Solar Radiation): Monthly solar radiation intensity influences weather patterns. Global Climate Indices:

SOI (Southern Oscillation Index) → Measures the air pressure difference between Tahiti and Darwin, influencing the El Niño and La Niña phenomena.

IOD (Indian Ocean Dipole) → An index describing the sea surface temperature difference between the western and eastern parts of the Indian Ocean, which can affect rainfall patterns in Indonesia.

A. *Exploratory Data Analysis*

This process helps us get a closer look at the dataset's contents, such as the characteristics of each variable, their distribution, and whether there are potential issues such as missing data, outliers, or other anomalies. During this exploration phase, several tasks are performed, including:

1. Viewing the data structure
2. Checking data types
3. Reviewing value distribution
4. Observing time trends
5. Identifying missing or invalid values
6. Analyzing relationships between variables

TABLE I. METADATA

No	Fitur (Unit)	[Min, Max]	[Mean, Stdev]
Tekanan Udara (millibar)			
1	P (MSL)	1008.50, 1013.70	1010.93, 0.86
2	P0 (Stasiun)	1006.21, 1011.40	1008.60, 0.85
Temperature Udara (°C)			
3	T07	21.20, 30.80	24.59, 0.83
4	T13	26.10, 38.80	29.55, 0,99
5	T18	24.70, 29.40	27.73, 0.62
6	T	25.00, 28.10	26.60, 0.62
7	Tx	30.50, 34.80	32.79, 0.69
8	Tn	18.20, 24.30	22.29, 0.93
Relative Humidity (%)			
9	RH07	88.00, 98.00	94.81, 1.70
10	RH13	57.00, 97.00	73.26 ,4.95

11	RH18	71.00, 96.00	81.66, 3.46
12	RH	77.00, 96.00	86.15, 2.68
Other			
13	SSS	2.30, 84.40	42.25, 18.11
14	CH	0.00, 926.90	275.21, 151.95
15	SOI	-28.60, 27.10	-0.44, 10.62
16	IOD	-1.18, 1.81	-0.01, 0.47

Table I presents the metadata of all features used in this study, including measurement units, minimum and maximum values, as well as the mean and standard deviation (stdev). These features cover air pressure, air temperature at different observation times (07:00, 13:00, and 18:00), relative humidity, and additional parameters such as Sea Surface Salinity (SSS), rainfall (CH), the Southern Oscillation Index (SOI), and the Indian Ocean Dipole (IOD). This statistical summary provides an initial overview of the distribution and variability of the input data used for monthly rainfall prediction modeling.

B. Feature Selection

In building an accurate and efficient prediction model, selecting the right features is a crucial step. Too many features can make the model complex and slow, while too few can reduce the quality of predictions. Therefore, in this study, the Hybrid Feature Selection Framework approach is used, which is a combined method of three feature selection techniques.

With the combination of these three approaches, the selected features are believed to have a significant contribution to the accuracy of monthly rainfall predictions. The Voting Feature Selector means that features will be selected if they are included in at least two of the three methods mentioned above. This approach strikes a balance between the robustness of statistical analysis and the power of machine learning models, making the selected features more validated from various.

C. Base Learner

Three algorithms were selected as base learners in the stacking scheme due to their ability to handle various types of data and their complementary characteristics:

Random Forest (RF) – An ensemble method based on trees that is resistant to overfitting and effective in identifying feature interactions. One effective way to enhance prediction accuracy is by using a random forest ensemble model, which combines the results of multiple decision trees. This approach allows the model to learn from different perspectives, leading to more stable and reliable predictions [6], [7].

K-Nearest Neighbors (KNN) – A non-parametric model that works based on the proximity of feature values, suitable for detecting recurring local patterns. The KNN algorithm is a reliable approach that makes predictions by averaging the values of nearby data points. It determines what counts as “nearby” based on the distance between each observation and the input being evaluated [8].

Decision trees (DT) – are commonly used in operations research, especially in decision analysis, to help find the best way to reach a goal. This model looks like a tree, where each branch represents a question that helps classify the data, and the leaves show the final result or category of the data [8].

XGBoost (XGB)– A boosting algorithm that is very popular due to its high accuracy and computational efficiency, excelling in handling tabular data. The XGBoost algorithm is a highly effective machine learning method that's gained popularity for time series forecasting. It works by combining multiple regression trees to make predictions and is especially good at handling seasonal and nonlinear patterns in data [9].

Linear Regression (LR) – is one of the simplest types of algorithms. Its main goal is to reduce the gap between predicted values and actual data. This algorithm is designed to produce numerical (quantitative) results, and it's often used to make predictions about future outcomes based on existing data [10].

Adaptive Boosting (ADB) – also known as AdaBoost, is an iterative algorithm first introduced by Freund and Schapire in 1997. It is an ensemble learning method that aims to build a strong predictive model by combining several simple models, known as "weak classifiers."

The core idea of AdaBoost is to run a simple learning algorithm repeatedly, while gradually shifting the focus toward training data that is harder to predict. This is done by adjusting the weights (or probabilities) of the data in each iteration [11].

D. Adaptive Stacking Ensemble Learner

After all base learners are trained and evaluated using validation metrics, the next step involves constructing the final predictive model through a stacking ensemble framework. Stacking ensemble learning is a powerful technique that integrates multiple models in a layered architecture to enhance predictive performance. It leverages the strengths of diverse algorithms and mitigates individual weaknesses, thereby reducing error in both classification and regression tasks [12].

Unlike conventional stacking methods that predetermine the meta-learner, this study initially proposed an adaptive stacking strategy. The main idea was to promote the best-performing base learner—based on key metrics such as R^2, RMSE, and MAE—as the meta-learner. However, during empirical evaluation, it was discovered that the model with the highest individual performance—Linear Regression ($R^2 = 0.931$)—did not necessarily yield the best results when applied as the final estimator in the ensemble framework.

To refine the strategy, several high-performing base learners were evaluated as potential meta-learners. In this extended approach, the top three models based on individual performance—Linear Regression, AdaBoost, and XGBoost—were each tested as meta-learners. This allowed the stacking mechanism to be adaptive not only in selecting a single best learner but in exploring how different strong learners perform in combining outputs from the base layer (Random Forest, KNN, Decision Tree, and others).

As an additional enhancement, a voting-based ensemble was constructed using the outputs from the three stacking models, each with one of the top-ranked meta-learners. By combining their predictions through a voting mechanism, the ensemble aimed to further reduce variance and capitalize on the strengths of each configuration. The results from this voting-based stacking demonstrated improved generalization and accuracy, outperforming individual stacking variants.

This strategy reaffirms the foundational principle of adaptive stacking—not to select the meta-learner rigidly based on isolated validation scores, but to assess ensemble effectiveness empirically. By incorporating multiple high-performing candidates as meta-learners and integrating their outputs through voting, the proposed approach remains flexible, data-driven, and capable of capturing complex inter-model interactions. Ultimately, this leads to a more resilient and accurate predictive system, particularly valuable for tasks such as monthly rainfall prediction where data variability and non-linearity are prevalent.

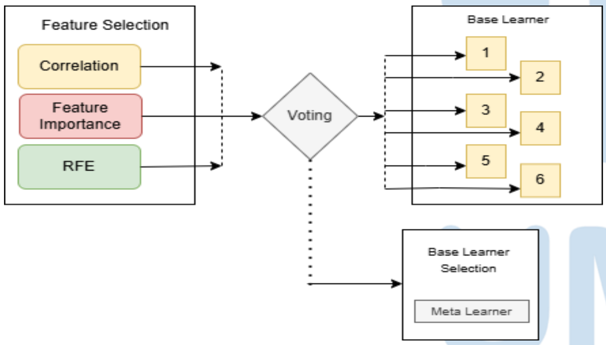

Fig. 1. Voting Feature Selector

"Fig 1" illustrates the architecture of the proposed adaptive stacking model incorporating hybrid feature selection and ensemble learning. Initially, three feature selection techniques—Correlation, Feature Importance, and Recursive Feature Elimination (RFE)—are applied independently. A voting mechanism is then employed to determine the optimal feature subset. The selected features are used to train multiple base learners, including Random Forest (RF), K-Nearest Neighbors (KNN), and XGBoost (XGB). Subsequently, a meta-learner is constructed by aggregating the outputs of the best-performing base learners to improve predictive accuracy and generalization.

Base models, or Level-0 models, are the first set of algorithms trained on the original data. They each make their own predictions, which are then used as input for the next step. The Level-1 model, known as the meta-model, learns how to blend those predictions from the base models in the most effective way to improve overall accuracy [13].

E. Tuning Hyperparameter

To make the tuning process more efficient and comprehensive, this study uses the Grid Search technique combined with Cross-Validation. With this approach, the system will try various combinations of hyperparameter values and evaluate the performance of each combination using training data that is randomly split into several folds. The final result of this process is the best combination of settings for each model that provides the most accurate and stable predictions.

Tuning is performed not only on the base models—XGBoost, AdaBoost, Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), and Linear Regression (LR)—but also on the meta-learner used in the adaptive stacking scheme. In this way, every layer of the modeling process is thoroughly optimized to operate harmoniously and effectively.

Cross-Validation Strategy – Given the relatively small size of the dataset—consisting of 408 monthly samples over a 34-year period—model validation becomes a crucial step to ensure the results are unbiased and generalize well to unseen data. To address this, cross-validation techniques were employed. During the hyperparameter tuning phase for the base models (XGBoost, AdaBoost, RF, KNN, DT, and LR), a 3-fold cross-validation scheme was applied. For the final training of the stacking ensemble, a 5-fold cross-validation was used to improve the model's robustness and evaluation stability. Although the folds were generated randomly, attention was given to maintaining balanced data distribution due to the temporal and seasonal nature of climatological data. This validation strategy helps reduce the risk of overfitting and provides a more reliable estimate of model performance under real-world conditions.

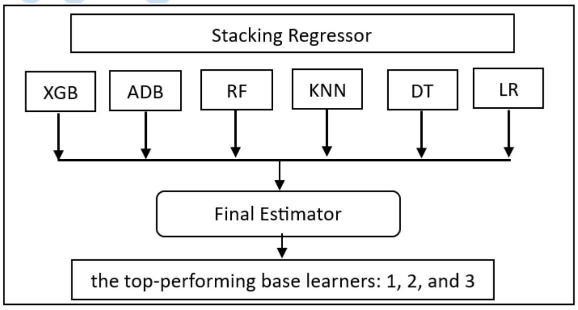

Fig. 2. Adaptive Stacking Ensemble

"Fig 2" The figure illustrates a brief overview of the stacking ensemble learning process. After completing the feature selection phase, the next step involves constructing a prediction model using the Adaptive Stacking Ensemble approach. In this setup, multiple base models—namely XGBoost, AdaBoost, Random Forest, K-Nearest Neighbors, Decision Tree, and

Linear Regression—are simultaneously trained to generate initial predictions.

Following the adaptive stacking principle, instead of selecting a single meta-learner, the final estimator is constructed using the top-performing base learners, as determined by their predictive performance across evaluation metrics. Specifically, the three best-performing models are promoted to serve as meta-learners, and their outputs are integrated in the final stage to improve robustness and accuracy. This strategy allows the ensemble to leverage both linear and nonlinear strengths across models, resulting in a more reliable and generalized prediction system.

III. Results and Discussion

This section presents the outcomes of the proposed monthly rainfall prediction model along with a comprehensive analysis of its performance. The results are discussed in relation to the accuracy of predictions, the effectiveness of the hybrid feature selection framework, and the comparative performance of different machine learning algorithms used. Furthermore, the impact of each selected feature on the model's predictive capability is evaluated, followed by a discussion on how the adaptive stacking approach enhances generalization across seasonal data. The findings are interpreted based on statistical metrics and are compared against baseline and conventional ensemble models to highlight the advantages of the proposed method.

TABLE II. FEATURE SELECTION

No.	Feature Selection	Selected Features
1	Correlation	RH13, RH, RH18, T13, RH07, T18, T, SSS, Tx
2	Feature Importance	RH13, T13, SSS, RH, Tx
3	RFE	P0, T07, T, Tn, RH13
4	Voting Feature Selector	RH13, RH, T13, T, SSS, Tx

Table II presents the results of feature selection obtained from three different methods: Correlation, Feature Importance, and Recursive Feature Elimination (RFE). Correlation Analysis selects features such as RH13, RH, RH18, T13, RH07, T18, T, SSS, and Tx, which show a strong relationship with the target. Meanwhile, the Feature Importance method (using decision tree models) identifies RH13, T13, SSS, RH, and Tx as important features. RFE, which gradually eliminates less important features, selects P0, T07, T, Tn, and RH13.

To achieve more stable and objective results, we apply the Voting Feature Selector by combining the results of these three methods. Finally, the features selected through voting are RH13, RH, T13, T, SSS, and Tx. This process ensures that the features used in the model are the most consistent and relevant according to various feature selection approaches.

After building the model with the Adaptive Stacking Ensemble approach, performance evaluations are conducted on each of the base learners, namely Random Forest (RF), K-Nearest Neighbors (KNN), and XGBoost (XGB), as well as on the stacking model itself. This evaluation aims to assess how well each model predicts the target. The three metrics used to evaluate model performance are the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) [14], [15]. The evaluation results from each model are summarized in the following Table III.

TABLE III. EVALUATION OF BASE LEARNERS AND ADAPTIVE STACKING

No	Base Learner	Evaluation			
		R^2	*SMAPE*	*MAE*	*RMSE*
1	XGB	0.907	14.34	0.305	0.360
2	ADB	0.909	13.68	0.293	0.356
3	RF	0.899	14.56	0.314	0.375
4	KNN	0.749	20.93	0.478	0.591
5	DT	0.811	18.10	0.419	0.513
6	LR	**0.931**	**11.81**	**0.259**	**0.309**

Table III presents the performance comparison of six base learner models—XGBoost (XGB), AdaBoost (ADB), Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), and Linear Regression (LR) evaluated using four metrics: coefficient of determination (R^2), symmetric mean absolute percentage error (SMAPE), mean absolute error (MAE), and root mean square error (RMSE).

Among all models, Linear Regression (LR) demonstrates the most outstanding performance, achieving the highest R^2 value of 0.931, which reflects its strong ability to explain the variance in monthly rainfall. Furthermore, LR recorded the lowest error across all other metrics—SMAPE of 11.81, MAE of 0.259, and RMSE of 0.309—highlighting not only its high predictive accuracy but also its effectiveness in modeling the predominantly linear relationships found in the dataset.

Following LR, AdaBoost and XGBoost also achieved strong results with R^2 values of 0.909 and 0.907, respectively. Both models outperformed RF in terms of generalization accuracy and yielded relatively low error rates. Their ensemble structure allows them to capture more complex patterns, particularly where non-linearity is present. Random Forest, although slightly behind (R^2 = 0.899), still demonstrated stable and consistent performance, reinforcing its reliability as a base ensemble model.

In contrast, K-Nearest Neighbors (KNN) and Decision Tree (DT) performed less favorably. KNN showed the weakest overall performance with an R^2 of

0.749 and the highest errors (SMAPE = 20.93, MAE = 0.478, RMSE = 0.591), indicating a tendency toward overfitting and limited generalization. Similarly, DT reported suboptimal results (R^2 = 0.811, RMSE = 0.513), suggesting difficulty in capturing the variability of the rainfall data when used independently.

In summary, while ensemble methods such as AdaBoost, XGBoost, and RF offer robust performance and better flexibility in capturing nonlinear patterns, Linear Regression remains the best-performing base learner in this study. Its consistently high accuracy across all evaluation metrics reinforces the presence of strong linear components in the rainfall dataset. These results also underscore the importance of carefully selecting base learners in ensemble strategies, depending on the complexity and structure of the target variable.

TABLE IV. EVALUATION OF STACKING AND VOTING STACKING

No	Adaptive Stacking Model	Evaluation			
		R^2	*SMAPE*	*MAE*	*RMSE*
1	XGB+ADB+RF+KNN+DT (Meta Learner LR)	0.916	13.53	0.288	0.341
2	XGB+RF+KNN+DT +LR (Meta Learner ADB)	0.914	13.24	0.288	0.345
3	ADB+RF+KNN+DT +LR (Meta Learner XGB)	0.913	13.43	0.293	0.347
4	Voting Stacking (Meta Learner LR+ADB+XGB)	0.917	13.33	0.287	0.339

Table IV presents a comparative analysis of four adaptive stacking ensemble configurations, each constructed from various combinations of base learners and meta-learners. The models are evaluated using four key performance metrics: the coefficient of determination (R^2), symmetric mean absolute percentage error (SMAPE), mean absolute error (MAE), and root mean square error (RMSE).

Among all configurations, Model 4, which employs a hybrid meta-learner composed of Linear Regression (LR), AdaBoost (ADB), and XGBoost (XGB) through a soft voting mechanism, delivers the best overall performance. It achieves the highest R^2 score (0.917), the lowest RMSE (0.339), and the lowest MAE (0.287), indicating strong predictive accuracy and generalization capability. The combination of diverse meta-learners enables this approach to effectively capture both linear and non-linear patterns in the predictions generated by the base learners.

Model 1, which uses Linear Regression as a single meta-learner stacked over five base learners (XGB, ADB, RF, KNN, and DT), also demonstrates competitive performance with an R^2 of 0.916 and RMSE of 0.341. Although slightly less accurate than the voting-based ensemble, it outperforms both Model 2 and Model 3, which utilize AdaBoost and XGBoost respectively as meta-learners.

Interestingly, when Linear Regression is applied independently (i.e., as a standalone model without stacking), it achieves even better results—R^2 of 0.931, SMAPE of 11.81, MAE of 0.259, and RMSE of 0.309—outperforming all stacking models. This finding suggests that Linear Regression is highly effective when applied directly to the original feature space of the dataset.

However, when used as a meta-learner in the stacking ensemble, its performance slightly degrades. This discrepancy may be attributed to the nature of the input it receives at the meta-level: rather than raw features, it processes predictions from the base learners, which may contain correlated errors and non-linear interactions. As a purely linear model, Linear Regression may struggle to effectively integrate such complex prediction spaces. In contrast, when applied directly to the raw data, it can fully leverage linear dependencies and underlying statistical distributions.

Therefore, the following conclusions can be drawn:

- If the objective is to achieve the highest standalone accuracy, then Linear Regression as an individual model is the most effective choice.

- If the goal is to obtain balanced and robust performance within an ensemble framework, the voting stacking approach (Model 4) is the most advantageous, due to its ability to combine the strengths of multiple meta-learners.

- Stacking ensembles with a single meta-learner may offer good results but tend to be less flexible and adaptive compared to hybrid strategies that employ voting across multiple meta-models.

These results reinforce the importance of adaptivity in ensemble learning strategies, particularly in data-driven domains like rainfall prediction, where capturing diverse patterns is essential for generalization.

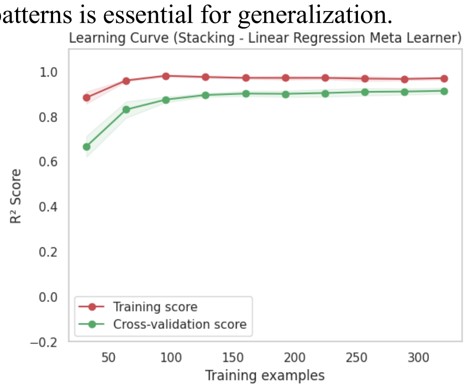

Fig. 3. Learning Curve Stacking (Linear Regression - R^2)

The "Fig. 3," illustrates the learning curve for the stacking ensemble model using Linear Regression as the meta-learner. The training and cross-validation R^2 scores demonstrate a stable and converging pattern as the number of training examples increases. The training score remains consistently high (close to 0.97), indicating low bias, while the cross-validation score gradually improves and stabilizes around 0.91, suggesting good generalization. The narrow gap between both curves confirms that the model does not suffer from significant overfitting or underfitting, reflecting a well-balanced bias-variance tradeoff.

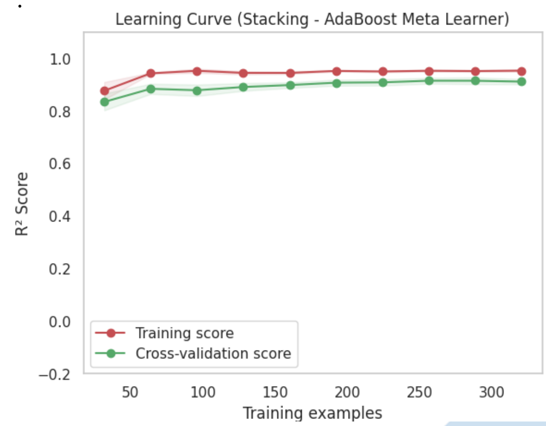

Fig. 4. Learning Curve Stacking (AdaBoost - R^2)

The "Fig. 4," shows the learning curve of the stacking ensemble model with AdaBoost as the meta-learner. The training R^2 score remains consistently high (above 0.93), while the cross-validation score steadily improves and plateaus around 0.91 as the number of training examples increases. The narrow and stable gap between training and validation curves indicates low variance and good generalization ability. This pattern suggests that the model is well-fitted, with minimal risk of overfitting or underfitting, making AdaBoost an effective choice for meta-learning in this stacking configuration.

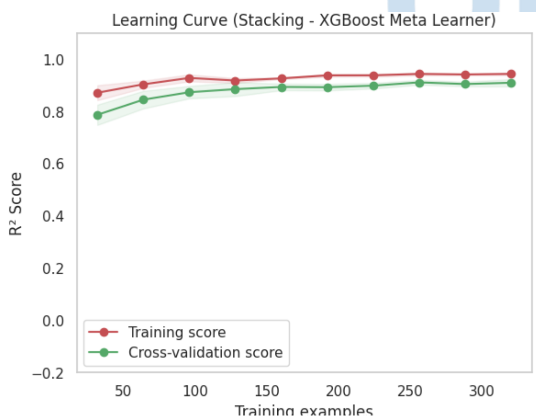

Fig. 5. Learning Curve Stacking (XGBoost - R^2)

The "Fig. 5," presents the learning curve of the stacking ensemble model utilizing XGBoost as the meta-learner. The training R^2 score remains consistently high (above 0.92), while the cross-validation score steadily improves and converges toward 0.90 as more training examples are added. Although a slight gap persists between the training and validation curves, it narrows progressively, indicating improved generalization with increased data. This trend suggests a stable learning process with moderate variance and reflects XGBoost's strong capability in capturing complex nonlinear relationships when used as a meta-learner within the stacking framework.

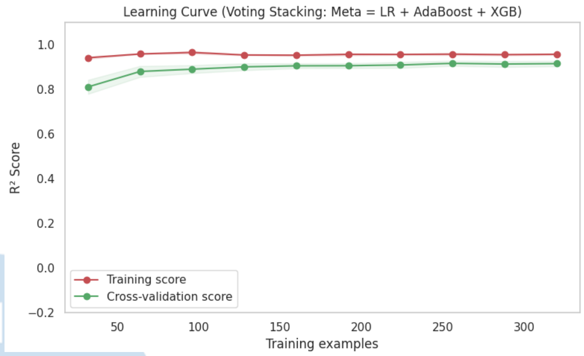

Fig. 6. Learning Curve Stacking (Voting LR+ADB+XGB - R^2)

The "Fig 6" illustrates the learning curve for the voting-based stacking ensemble, which integrates three meta-learners—Linear Regression (LR), AdaBoost (ADB), and XGBoost (XGB)—through a soft voting strategy. The training R^2 score remains consistently high (≈0.94), indicating the model's strong fit on the training data. Meanwhile, the cross-validation R^2 score steadily increases with more training examples, stabilizing around 0.917, which is the highest among all evaluated configurations.

The narrow and stable gap between the training and validation curves suggests a well-balanced bias-variance tradeoff, with no signs of overfitting or underfitting. This behavior highlights the robust generalization capability of the voting ensemble, benefiting from the complementary strengths of linear and nonlinear learners. The curve confirms that the model continues to learn effectively as more data is introduced and that it maintains high predictive stability across various training set sizes.

Overall, the learning curve validates the effectiveness of combining multiple strong meta-learners via soft voting, making this configuration the most reliable and accurate among the tested stacking strategies.

Stacking ensemble learning works in two layers: a group of base models and a meta-learner. First, the base models are trained using the original training data. Then, their predictions are gathered and used to train the meta-learner, which learns how to best combine those outputs for more accurate final results [16].

IV. Conclusion

This study has successfully demonstrated that the integration of a Hybrid Feature Selection Framework with an Adaptive Stacking Ensemble significantly enhances the accuracy and robustness of monthly rainfall prediction models. By combining Correlation Analysis, Feature Importance, and Recursive Feature Elimination (RFE) through a voting mechanism, the proposed feature selection approach effectively identifies the most relevant meteorological predictors while excluding less informative variables—such as the global climate indices SOI and IOD—which were not selected by any method. This led to a more parsimonious and computationally efficient model without sacrificing predictive performance.

Experimental results confirm that the proposed Adaptive Stacking approach outperforms individual learners and conventional ensemble methods. While standalone Linear Regression recorded the highest individual performance ($R^2 = 0.931$), it did not retain this advantage when used as a meta-learner in the stacking framework. Instead, the most effective configuration was achieved through voting-based meta-learning, combining Linear Regression, AdaBoost, and XGBoost, which produced the best overall ensemble performance with an R^2 of 0.917, MAE of 0.287, and RMSE of 0.339.

The learning curves of each stacking configuration further validated the model's generalization capability. The voting ensemble showed the most stable bias-variance tradeoff, benefiting from the diversity of its meta-learners. These findings emphasize that in adaptive ensemble learning, meta-learner selection should not be rigidly based on individual model scores but evaluated empirically within the ensemble context.

Overall, this research presents a robust, flexible, and data-driven predictive framework that can adapt to the nonlinear and dynamic nature of rainfall patterns. Its practical applicability holds strong potential for climate-sensitive sectors such as agriculture, hydrology, water resource management, and early warning systems for hydrometeorological hazards.

V. Suggestions

Based on the results and insights gained from this study, several directions are proposed for future research to further enhance the adaptability and predictive strength of the proposed framework. One potential improvement involves expanding the variety of meta-learners used in the stacking ensemble. While this study focused on top-performing learners such as Linear Regression, AdaBoost, and XGBoost, incorporating other advanced algorithms—such as support vector machines, deep learning models, or neural-based regressors—may improve performance under more complex or highly non-linear climate conditions.

Additionally, considering the spatial and temporal variability of rainfall, future research could explore region-specific adaptations or spatio-temporal extensions of the model to improve its generalization across different climatic zones. This would be particularly relevant for scaling the model to a national or regional level, where rainfall dynamics may vary significantly.

Enhancing the hyperparameter optimization process is also a promising avenue. The use of more sophisticated methods—such as Bayesian optimization or evolutionary algorithms—could yield better parameter configurations than traditional grid search, thus improving overall model efficiency and accuracy.

Furthermore, integrating additional climate-related indicators, particularly those linked to ocean-atmosphere interactions, may help refine the model's ability to capture long-term and seasonal rainfall anomalies. Lastly, due to its robust and flexible nature, the proposed adaptive stacking framework holds significant promise for broader applications beyond rainfall prediction. It could be extended to areas such as drought monitoring, precision agriculture, flood risk management, and climate-related decision support systems, offering valuable tools for anticipating and mitigating the impacts of environmental variability.

References

[1] E. Oprasmani, T. Amelia, and E. Muhartati, “Membangun Masyarakat Peduli Lingkungan Pesisir Melalui Edukasi Kepada Masyarakat Kota Tanjungpinang Terkait Pelestarian Daerah Pesisir,” To Maega : Jurnal Pengabdian Masyarakat, vol. 3, no. 2, p. 66, Aug. 2020, doi: 10.35914/tomaega.v3i2.372.

[2] P. D. Riajaya, “Rainy season period and climate classification in sugarcane plantation regions in indonesia,” in IOP Conference Series: Earth and Environmental Science, Institute of Physics Publishing, Jan. 2020. doi: 10.1088/1755-1315/418/1/012058.

[3] R. Ruqoyah, Y. Ruhiat, and A. Saefullah, “Analisis Klasifikasi Tipe Iklim Dari Data Curah Hujan Menggunakan Metode Schmidt-Ferguson (Studi Kasus: Kabupaten Tangerang),” Serang, Jan. 2023. doi: https://doi.org/10.23960/jtaf.v11i1.327.

[4] Juliati, “ANALISIS KARAKTERISTIK CURAH HUJAN DENGAN MENGGUNAKAN KLASIFIKASI SCHMIDT-FERGUSSON DI KOTA MAKASSAR,” Jurnal Sains dan Pendidikan Fisika (JSPF), vol. 19, no. 2, pp. 229–235, Aug. 2023, doi: prefix10.35580.

[5] S. R. Aisy, M. K. Ramadhan, A. S. Salsabila, and R. Kurniawan, “Perbandingan Algoritma Klasifikasi Data Mining dalam Memprediksi Curah Hujan di Jawa Barat,” Seminar Nasional Sains Data, pp. 180–192, 2024, doi: https://doi.org/10.33005/senada.v4i1.179.

[6] S. D. Latif et al., “Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches,” Nov. 01, 2023, Elsevier B.V. doi: 10.1016/j.aej.2023.09.060.

[7] A. Raut, D. Theng, and S. Khandelwal, “Random Forest Regressor Model for Rainfall Prediction,” in 2023 International Conference on New Frontiers in Communication, Automation, Management and Security, ICCAMS 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICCAMS60113.2023.10526085.

[8] S. Biruntha, B. S. Sowmiya, R. Subashri, and M. Vasanth, “Rainfall Prediction using kNN and Decision Tree,” in Proceedings of the International Conference on Electronics and Renewable Systems, ICEARS 2022, Institute of Electrical and

Electronics Engineers Inc., 2022, pp. 1757–1763. doi: 10.1109/ICEARS53579.2022.9752220.

[9] P. Mishra et al., "Modeling and forecasting rainfall patterns in India: a time series analysis with XGBoost algorithm," Environ Earth Sci, vol. 83, no. 6, Mar. 2024, doi: 10.1007/s12665-024-11481-w.

[10] C. Vijayalakshmi, K. Sangeeth, R. Josphineleela, R. Shalini, K. Sangeetha, and D. Jenifer, "Rainfall Prediction using ARIMA and Linear Regression," in 2022 1st International Conference on Computer, Power and Communications, ICCPC 2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 366–370. doi: 10.1109/ICCPC55978.2022.10072125.

[11] A. Belghit, M. Lazri, F. Ouallouche, K. Labadi, and S. Ameur, "Optimization of One versus All-SVM using AdaBoost algorithm for rainfall classification and estimation from multispectral MSG data," Advances in Space Research, vol. 71, no. 1, pp. 946–963, Jan. 2023, doi: 10.1016/j.asr.2022.08.075.

[12] M. H. D. M. Ribeiro, R. G. da Silva, S. R. Moreno, V. C. Mariani, and L. dos S. Coelho, "Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting," International Journal of Electrical Power and Energy Systems, vol. 136, no. September 2021, 2022, doi: 10.1016/j.ijepes.2021.107712.

[13] P. G. Jaiswal et al., "A Stacking Ensemble Learning Model for Rainfall Prediction based on Indian Climate," 2023 6th International Conference on Information Systems and Computer Networks, ISCON 2023, pp. 1–6, 2023, doi: 10.1109/ISCON57294.2023.10112077.

[14] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Comput Sci, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[15] Roshani et al., "Analyzing trend and forecast of rainfall and temperature in Valmiki Tiger Reserve, India, using non-parametric test and random forest machine learning algorithm," Acta Geophysica, vol. 71, no. 1, pp. 531–552, Feb. 2023, doi: 10.1007/s11600-022-00978-2.

[16] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Comput Sci, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.