Stacking Regressor Model for $PM_{2.5}$ Concentration Prediction Based on Spatiotemporal Data

Mitra Unik[1], Imas Sukaesih Sitanggang[2], Lailan Syaufina[3], I Nengah Surati Jaya[4]
Department of Computer Science, IPB University, Bogor, Indonesia[1, 2]
Department of Informatics Engineering, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia[2]
Department of Silviculture, IPB University, Bogor, Indonesia[3]
Department of Forest Management, IPB University, Bogor, Indonesia[4]

***Abstract*—This study presents the development of a predictive model for $PM_{2.5}$ concentrations resulting from forest and peatland fires in Riau Province, utilizing the stacking regressor technique within an ensemble learning framework. The model integrates spatiotemporal data from remote sensing and ground-based sensors at a resolution of 1 km x 1 km, demonstrating its effectiveness in capturing the intricate patterns of $PM_{2.5}$ concentrations. By combining Random Forest, Gradient Boosting Machine (GBM), and XGBoost, with RidgeCV as a meta-learner, the model attained optimal performance, achieving $R^2 = 0.851$, MAE = 0.045 µg/m³, and MSE = 0.003 µg/m³. The incorporation of temporal feature engineering techniques, including lag and rolling window methods, significantly enhanced prediction accuracy, enabling the model to effectively capture seasonal variations and temporal dynamics. Key variables, such as air temperature, evapotranspiration, and Aerosol Optical Depth (AOD), were found to exhibit strong correlations with $PM_{2.5}$ concentrations. The findings from this research contribute to the formulation of data-driven policies for air quality management and pollution mitigation, with the potential for broader application in regions encountering similar environmental challenges.**

***Keywords*—Ensemble learning; $PM_{2.5}$ prediction; remote sensing; stacking regressor; spatio-temporal data**

I. Introduction

$PM_{2.5}$ (Particulate Matter $\leqslant$ 2.5 micrometres per cubic metre), which mainly comes from biomass burning such as forest and land fires, vehicle emissions, and coal combustion, causes various serious health impacts [1]. The measurement of $PM_{2.5}$ due to forest fires faces challenges such as the episodic nature of fires, limited monitoring stations, and limited data availability [2]. Measurement approaches include ground stations with high accuracy but limited coverage, as well as satellite remote sensing that has wide and continuous coverage [3]. Satellite technology is effective in detecting fires, exposure to air pollution, and concentrations of aerosol particles including $PM_{2.5}$ [4].

This research analyzes the performance of various machine learning algorithms, namely Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Neural Network (NN), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN), with the evaluation metrics of Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). To improve prediction accuracy, feature engineering is applied through the creation of lag and rolling window features. Lag features are based on the concept that historical values of a variable, such as $PM_{2.5}$ concentrations, can influence current or future values, especially in time series data [5], [6]. Variables such as aerosol concentration, relative humidity, ground surface temperature, and air temperature are lagged to capture temporal influences. In addition, rolling window statistics, such as mean, median, and standard deviation, are calculated to capture long-term trends and seasonal patterns, helping the model understand the dynamics of $PM_{2.5}$ changes influenced by seasonal factors or other external events [7]. Riau Province - Indonesia was chosen as the research location because it has the largest peatland in Sumatra Island, which is 3.89 million hectares out of a total of 5.85 million hectares. This condition makes Riau Province an appropriate location for study the impact of forest and peatland fires on $PM_{2.5}$ concentrations [8], [9]. The aim of this research is to develop a machine learning ensemble model with optimised regressor stacking, and to integrate temporal dynamics and trend patterns to predict $PM_{2.5}$ concentrations using 1 km x 1 km spatial and daily temporal remote sensing and ground sensor data, thereby supporting environmental management and public health policy.

II. Related Work

Research relevant to this study includes various $PM_{2.5}$ prediction models that integrate remote sensing-based predictor data, meteorological parameters and land use. Simple regression models such as Linear Regression (LR) and Multiple Linear Regression (MLR) are often used due to their simplicity, but they fail to capture non-linear relationships in high-dimensional datasets [10], [11], [12]. In contrast, machine learning techniques such as Random Forest (RF), Gradient Boosting (GB), and XGBoost have shown better ability in handling complex data and producing more accurate predictions [13], [14].

Further performance improvements are achieved through ensemble learning methods, such as Bagging, Boosting, and Stacking, which combine multiple models to reduce their individual weaknesses and improve prediction reliability [15], [16]. For example, research by Chen [17] showed that the stacking regressor model with meta-learner was able to achieve a coefficient of determination (R^2) of 0.85 and a Root Mean Squared Error (RMSE) of 17.3 μg/m³, which was superior to the single model. In addition, model combinations such as RF, GB, and Linear Mixed Regression (LMR) by Matsuki [18] and findings Li [19] demonstrating the importance of spatial resolution in improving the accuracy of $PM_{2.5}$ predictions.

Recent studies have also shown the successful application of ensemble models in predicting $PM_{2.5}$ concentrations in various regions, such as China [18], South Asia [4], United States of America [20], and Italy [14]. Stacking regressor, in particular, is becoming a highly relevant method due to its ability to integrate predictions from base models such as RF, GB, and XGBoost using a meta-learner, which optimises the combination of predictions to produce more accurate final results [21]. This approach has shown its effectiveness in capturing complex and non-linear data patterns, which are often unreachable by conventional regression models.

III. METHOD

A. Location, Period and Research Data

This research was conducted in Riau Province, Indonesia, during the period 1 March 2022 to 31 March 2024. Geographically, Riau Province is located between 01°05'00‘’ N to 02°25'00‘’ N and 100°00'00‘’ E to 105°05'00‘’ East. Riau is the part of Sumatra Island that has the largest area of peatland, with 3.89 million hectares out of a total of 5.85 million hectares. The province frequently experiences forest and peatland ecosystem fires, which have the potential to cause haze disasters with transnational impacts. In this study, the prediction of $PM_{2.5}$ concentrations due to forest and peatland ecosystem fires uses meteorological, environmental and geospatial data. Data were obtained from the air quality sensor of the Meteorology, Climatology and Geophysics Agency (BMKG) at Sultan Syarif Kasim II Airport Pekanbaru (101.45° East, 0.46° LU) as well as through satellite remote sensing. Data collection was conducted with daily temporal and spatial resolution, using a 30,000-metre buffer from the ground sensor, and a spatial buffer every 1,000-metres within the 30,000-metre range according to the Area of Interest (AOI), as shown in Fig. 1.

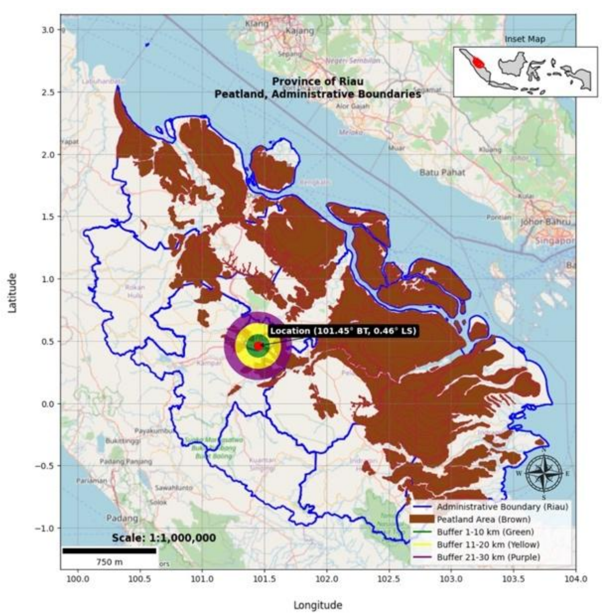


Fig. 1. Map of the study area and AOI.

B. Research Stage

The research utilizes machine learning algorithms as base models to enhance prediction accuracy through stacking, detailing the procedure, stacking architecture, and performance evaluation, as shown in Fig. 2.

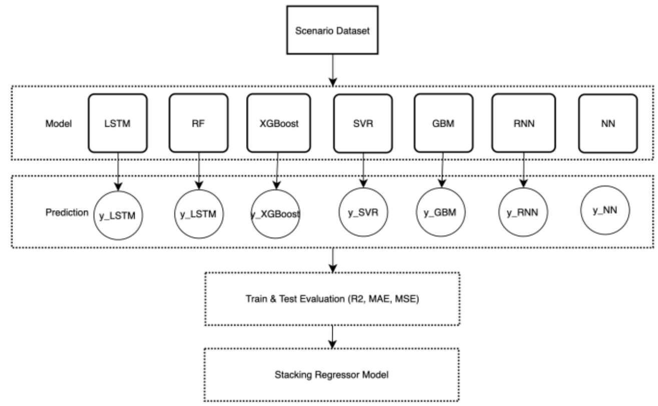


Fig. 2. General stages of modelling using the base model algorithm.

In the initial stage seven different machine learning algorithms as base models to get predictions from each model, namely LSTM, RF, XGBoost, SVR, GBM, NN, and RNN. Each base model generates predictions for the test data, which are referred to as (y_LSTM, y_RF, y_XGBoost, y_SVR, y_GBM, y_RNN, and y_NN). These predictions are generated from the training process performed on the training data. Each base model is evaluated using several evaluation metrics such as R^2, MSE, and MAE. This evaluation aims to measure how well each base model performs against the test data. The model with the best performance on these evaluation metrics is used as the basis for the next stage, which is the development of the ensemble learning model - Attention stacking regressor Model. Furthermore, the research process involves several main stages in applying the stacking regressor method to predict $PM_{2.5}$ concentrations. These stages include dataset preparation, feature engineering, dataset sharing, basic model development, stacking regressor-meta learner modelling, model evaluation and result interpretation as visualised in Fig. 3.

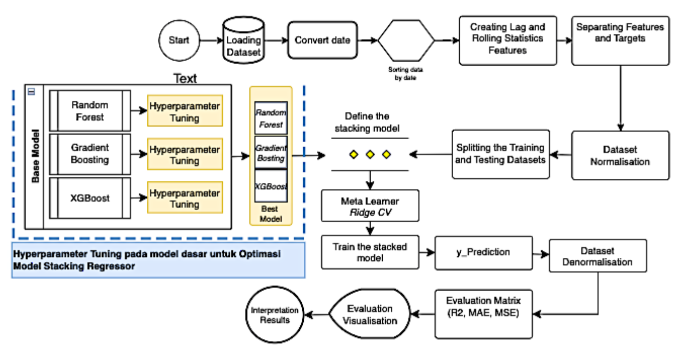


Fig. 3. Research stages of ensemble learning model - stacking regressor.

IV. RESULTS AND DISCUSSION

A. Research Dataset

In general, the predictors used as features of the prediction model for $PM_{2.5}$ concentrations resulting from forest and peatland fires, taken from ground and remote measurement sensor stations are as shown in below.

B. *$PM_{2.5}$ Concentration in the Study Period*

During the study, $PM_{2.5}$ concentrations were analysed through two graphs (Fig. 4): $PM_{2.5}$ Level Distribution and $PM_{2.5}$ Trend over Time. Fig. 4(a) shows that most of the $PM_{2.5}$ concentrations were in the range of 15-30 µg/m³, falling into the Good to Moderate category, with concentrations above 55 µg/m³ rarely occurring, signalling generally safe air quality. Fig. 4(b) shows the daily trend of $PM_{2.5}$ from 2022 to 2024, where 74.83% of days are in the Moderate category, 22.16% in the Good category, and 3.01% in the Unhealthy category for the Sensitive Group. Overall, the graph shows that while most days have safe to moderate air quality, there are certain periods where $PM_{2.5}$ concentrations increase to potentially dangerous levels, especially for vulnerable groups. This emphasises the importance of continuous air quality monitoring to anticipate health risks, particularly during periods of increased pollution.

A pattern of fluctuations in $PM_{2.5}$ concentrations was seen throughout the year, with a significant peak occurring at the end of 2023, which was most likely related to forest and peatland fires in Riau, covering more than 2,000 hectares in October 2023 [22].

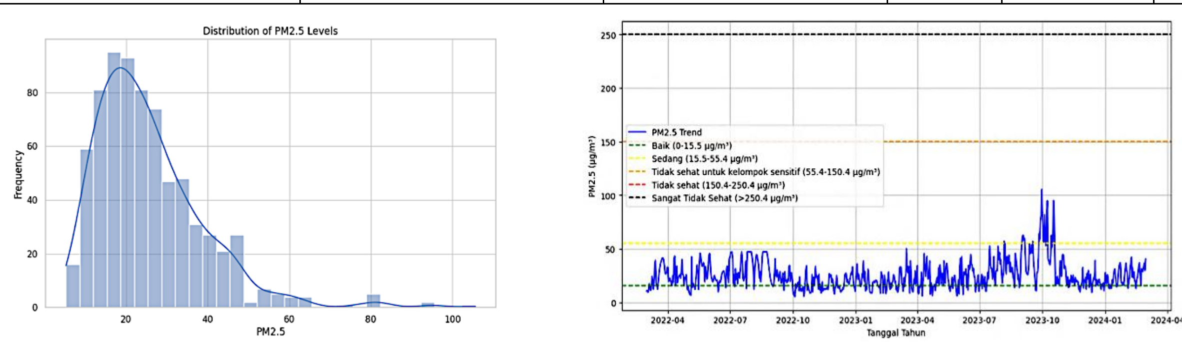
C. *Feature Correlation with $PM_{2.5}$ Concentration*

Feature correlation analysis aims to identify the variables that have the strongest relationship with $PM_{2.5}$ concentrations in the dataset. Results in Fig. 5 shows the correlation heatmap for all features in the dataset against the $PM_{2.5}$ target. Based on the results of the correlation analysis of $PM_{2.5}$ in Fig. 6, the red colour represents a strong positive correlation, while the blue colour shows a significant negative correlation.

The feature with the greatest influence is TEMP (air temperature), which has a significant positive correlation, indicating that an increase in temperature tends to increase $PM_{2.5}$ concentrations. In addition, ET (Evapotranspiration) features at certain radii, such as ET30, ET27, and ET28, also show strong positive correlations, signalling that areas with high evapotranspiration rates tend to have greater $PM_{2.5}$ concentrations. AOD (Aerosol Optical Depth), especially at large radii such as *max_AOD*, also showed a significant relationship with $PM_{2.5}$, reinforcing the link between aerosol particles in the atmosphere and $PM_{2.5}$ concentrations. These features were identified as the most relevant and influential variables in the air quality prediction model based on their strong relationship with $PM_{2.5}$.

TABLE I. COMMON PREDICTORS USED IN THE STUDY

Predictor	Description	Source	Unit	Temporal Resolution	Spatial Resolution
$PM_{2.5}$ ground	Particulate Matter ≤ 2.5 µg/m³	BMKG	µg/m³	Daily	30 Km
TEMP	Relative Temperature	BMKG	°C	Daily	30 Km
PRS	Air pressure	BMKG	hPa	Daily	30 Km
PRE	Rainfall	BMKG	mm	Daily	30 Km
RHU	Relative Humidity	BMKG	%	Daily	30 Km
SSD	Sunlight hours	BMKG	Hours	Daily	30 Km
WIN	Wind speed	BMKG	m/s	Daily	30 Km
Min/Max_NDVI_bufer 1 to NDVI_30	NDVI	MODIS/061/MYD13A1	Unitless	16 Days	1 Km
Min/Max _AOD_bufer 1 to NDVI_30	Aerosol Optical Depth	MODIS (Terra & Aqua MAIAC MCD19A2.061)	Unitless	Daily	1 Km
Min/Max _ET_bufer 1 to NDVI_30	Evapotranspirasi	MODIS/061/MOD16A2	kg/m²	8 Days	1 Km
Min/Max _LSTDay_bufer 1 to NDVI_30	Daytime surface temperature	MODIS/061/MOD11A1	°C	Daily	1 Km
Min/Max _LSTNight_bufer 1 to NDVI_30	Nighttime surface temperature	MODIS/061/MOD11A1	°C	Daily	1 Km



(a) (b)

Fig. 4. PM2.5 concentration in the study period (a) Level distribution (b) Concentration trends.

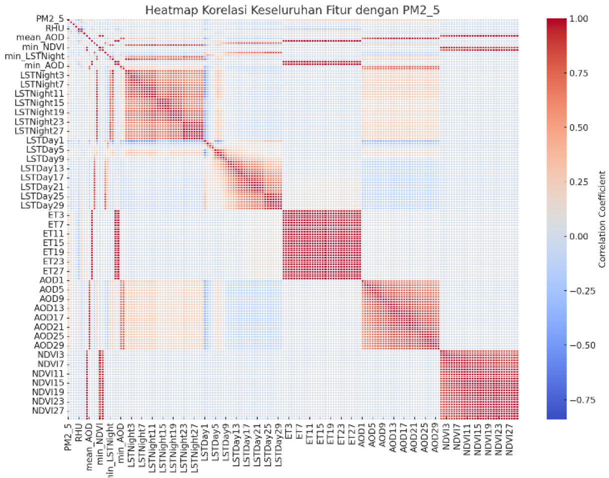


Fig. 5. Heatmap of feature correlation with $PM_{2.5}$ concentration.

D. Evaluate the Performance of the base Model Algorithm

Table II shows the model performance evaluation results, for $PM_{2.5}$ concentration prediction. The XGBoost model performed best on the training data with R^2 of 1.00, MAE of 0.07 µg/m³, and MSE of 0.01 (µg/m³)², indicating an almost perfect fit. However, on the test data, the performance decreased with an R^2 of 0.40, MAE of 7.18 µg/m³, and MSE of 109.65 (µg/m³)². The Random Forest model also showed good performance on training (R^2 0.92, MAE 2.81 µg/m³, MSE 13.38 (µg/m³)²) but decreased on testing (R^2 0.36, MAE 7.16 µg/m³, MSE 116.71 (µg/m³)²). The Gradient Boosting Machine, and Neural Network models had moderate performance with training R^2 of 0.84 and 0.81, and testing R^2 of 0.41 and 0.42, respectively. Meanwhile, the Support Vector Regression, LSTM and RNN models showed lower performance, with training R^2 ranging from 0.17 to 0.38 and testing R^2 between 0.14 and 0.27.

TABLE II. PERFORMANCE EVALUATION OF TRAINING AND TESTING MODELS

Model	Dataset Training Performance			Dataset Testing Performance		
	R^2	MAE	MSE	R^2	MAE	MSE
Random Forest	0.92	2.81	13.38	0.36	7.16	116.71
XGBoost	1.00	0.07	0.01	0.40	7.18	109.65
Support Vector Reg.	0.17	8.82	147.42	0.14	8.35	157.03
GBM	0.84	4.12	28.05	0.41	6.99	108.56
Neural Network	0.81	4.38	34.22	0.42	7.67	106.85
LSTM	0.38	7.81	109.17	0.27	8.36	133.65
RNN	0.33	8.07	118.26	0.23	8.39	141.53

E. Improving $PM_{2.5}$ Predictions by Capturing Temporal Dynamics and Trend Patterns

This research applies feature engineering techniques by creating lag and rolling window features that allow the model to capture dynamics and temporal trend patterns in time series data.

1) Lag creation: The lag feature is based on the concept that the historical value of a variable may affect the current or future value, especially in time series data. In the context of air pollution, $PM_{2.5}$ concentration on a particular day can be influenced by meteorological conditions, especially AOD [16], [17], [18], [19], [20], [21] and the environment on previous days. Therefore, the variables that were considered to have significant influence and lag features were created include: Representation of aerosol concentration in the atmosphere, which is correlated with $PM_{2.5}$ particles, Relative humidity of the air, which affects the formation and dispersion of pollutant particles, Ground surface temperature during the day, which can affect chemical and physical activity in the atmosphere, and Air temperature, an important factor in atmospheric processes.

The lag feature in time series data is calculated using a shift function, which represents the value of a variable in the previous time period. Conceptually, the lagt-n value describes the value of a variable at a given time that has been shifted by n time steps backwards, with n representing the number of lag periods taken into account. In Python programming, the lag feature is created by shifting the data 4 time steps back using the **.**shift() method.

The 4-day lag was selected based on exploration to find the optimal value. The dynamic characteristics of $PM_{2.5}$ that can persist and be influenced by atmospheric processes make this lag important in the model, allowing the utilisation of historical information to improve the accuracy of predicting concentration changes.

2) Statistics rolling window: Variables for which rolling statistics are calculated, such as max_AOD, mean_AOD, min_AOD, RHU, max_LSTDay, and TEMP, use specific time windows to apply statistical functions. The rolling mean provides a measure of the general trend by calculating the average of the values within that window, helping to understand the overall data pattern. The rolling mean calculation follows Eq. (1).

$$Rolling\ Mean = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

Where:

n : Total number of values in the window.

x_1: Individual values in the window.

$\sum_{i=1}^{n} x_i$: Sum of all values in the window.

Rolling median, If the number of data is even, the median is calculated as the average of the two middle values. If it is odd, the median is the centre value itself. The median is more resistant to outliers, so it gives a better idea of the centre of the data when there are extreme values. The even rolling median is calculated with Eq. (2) and the odd rolling median is calculated with Eq. (3).

$$Rolling\ Median_{even} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \tag{2}$$

$$Rolling\ Median_{odd} = x_{\frac{n+1}{2}} \tag{3}$$

Where:

n : Total number of values in the window.

Rolling Standard Deviation (Std) measures the spread of data; the larger the standard deviation value, the greater the variation in the data. This is important for understanding how stable or volatile $PM_{2.5}$ concentrations are. Rolling Standard Deviation (Std) is calculated with Eq. (4).

$$Std = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - Mean)^2} \quad (4)$$

Where:

n : Total number of values in the window.

x_i: Individual values in the window.

$Mean$: Average of the values in the window.

$(x_i - Mean)^2$: The squared difference between each value and the mean, which measures the deviation of each value from its centre.

Determination of the best rolling window size in modelling $PM_{2.5}$ concentrations was done by utilising the XGBoost Regressor model. The tested rolling windows varied from size 3 to 20. For each rolling window size, the XGBoost model was trained and evaluated to obtain R^2. The model was trained using normalised data to ensure the data was in a comparable range. The results of the rolling window evaluation can be seen in Fig. 6.

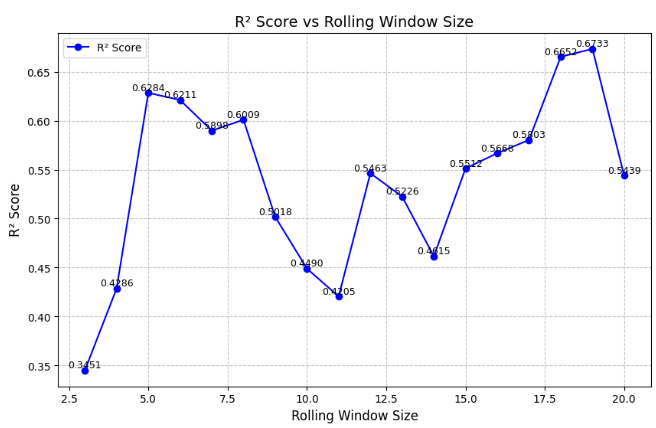


Fig. 6. Rolling window size evaluation results.

The analysis shows that the optimal rolling window size for $PM_{2.5}$ prediction is 19 days with an R^2 Score of 0.6733. Rolling window sizes that are too small or too large tend to produce suboptimal performance, with the second peak at 5 days (R^2 = 0.6284) and the lowest performance at 10 days (R^2 = 0.4490). A larger rolling window is able to capture more historical information, thus improving the model's ability to predict $PM_{2.5}$ dynamics. However, the application of the rolling and lag features led to the appearance of NaN values at the beginning of the data (e.g., the first 18 rows for a 19-day rolling window), which were removed using data.dropna() after the addition of the features. A summary comparison of the datasets before and after feature addition can be seen in Table III. The effect of data transformation with lag and rolling window features on data representation is shown in Fig. 7. The original variables (e.g., mean_AOD, RHU, max_LSTDay, min_LSTDay, and TEMP) shown in the left graph (blue) do not reflect the temporal dynamics clearly. In contrast, the transformed variables with a lag period of 4 and a rolling window on the right graph (red) show a clearer and more dynamic historical pattern. Features such as mean_AOD_lag4 capture the influence of previous conditions on current values, thus improving the model's ability to understand temporal relationships. This transformation significantly improves the model's ability to capture complex patterns, which in turn is expected to improve the accuracy of $PM_{2.5}$ predictions.

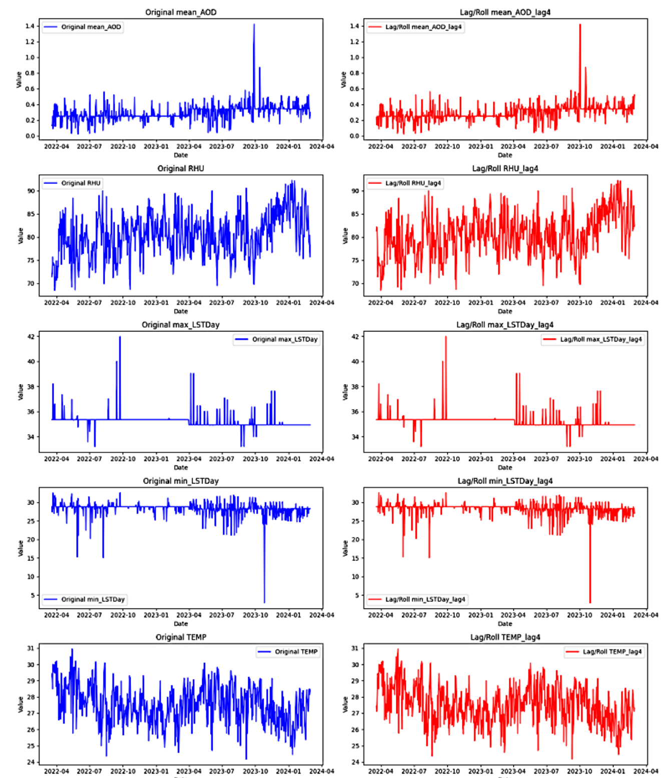


Fig. 7. Comparison of datasets before and after adding lag and rolling window features.

TABLE III. SUMMARY OF DATASETS BEFORE AND AFTER LEG AND ROLLING WINDOWS PROCESSING

Criteria	Before Lag & Rolling Feature Addition	After Addition of Lag & Rolling Feature
Number of Rows	731	713
Number of Columns	173	175
Average mean_AOD	0.2915	0.2934
Average RHU	80.43	80.45
Average max_LSTDay	35.23	35.23
Average min_LSTDay	28.22	28.23
Mean TEMP	27.54	27.53
Standard Deviation of mean_AOD	0.1076	0.1076
RHU Standard Deviation	4.59	4.63
Standard Deviation of max_LSTDay	0.53	0.32
min_LSTDay Standard Deviation	1.89	1.47
TEMP Standard Deviation	1.18	1.18

F. Performance of $PM_{2.5}$ Prediction Model with and without Temporal Features

Before the temporal features were applied, the top three basic models-Random Forest, Gradient Boosting Machine, and XGBoost-had relatively low R^2 values of 0.36, 0.41, and 0.40, and high MAE and MSE. However, after the temporal features were included, the performance of the models improved significantly. Random Forest recorded an R^2 of 0.761, Gradient Boosting Machine achieved an R^2 of 0.767, and XGBoost recorded the highest R^2 of 0.798, with lower MAE and MSE. This shows that the application of temporal features can substantially improve the accuracy of $PM_{2.5}$ prediction. Table 4 presents the performance evaluation of $PM_{2.5}$ prediction models before and after the addition of temporal features (lag and rolling window).

TABLE IV. EVALUATION OF $PM_{2.5}$ PREDICTION MODEL PERFORMANCE BEFORE AND AFTER INCORPORATING TEMPORAL FEATURES ON THE TEST DATASET

Model	Before			After		
	R^2	MAE	MSE	R^2	MAE	MSE
RF	0.36	7.16	116.71	0,761	0,058	0,005
GBM	0.41	6.99	108.56	0,767	0,058	0,005
XGBoost	0.40	7.18	109.65	0,798	0,053	0,005

G. Development of Ensemble Learning Model - Stacking Regressor

An ensemble learning model is applied using the Stacking Regressor approach to predict $PM_{2.5}$ concentrations due to forest and land fires along with the use of lag and rolling window features. The stacking approach combines multiple machine learning models (base learners) to improve prediction accuracy by utilising three strengths of each base model (RF, GBM and XGBoost). The results from these base models are then fed into a meta-learner, which in this case is RidgeCV. RidgeCV was selected as the meta-learner in this study for several technical reasons. First, RidgeCV employs L2 regularization to prevent overfitting and enhance model stability by reducing excessive model complexity. Second, RidgeCV is effective in addressing multicollinearity among the predictions from base models. Third, it automatically performs cross-validation to select the optimal alpha parameter, ensuring an appropriate balance between bias and variance. Additionally, RidgeCV is computationally efficient compared to other meta-learners and is versatile in integrating predictions from various base models with different characteristics (e.g., Random Forest, which tends to be more robust with non-linear data, and XGBoost, which is more sensitive to structured data) [23].

We used the best alpha value (0.1) from the search results on a logarithmic scale from 10-6 to 106 to effectively combine the predictions from the base model. Once trained, the stacking regressor model using RidgeCV as a meta-learner gave excellent results. The evaluation results on the test dataset (see Table V) showed an R^2 value of 0.845, with an MAE of 0.044 $\mu g/m^3$ and MSE of 0.003 $(\mu g/m^3)^2$.

H. Hyperparameter Tuning for Base Model Optimisation and Stacking Regressor via Grid Search

Grid Search with Cross-Validation (GSCV) is a robust method for optimizing hyperparameters in deep learning models, where cross-validation plays a critical role in enhancing model accuracy by systematically using different subsets of the training data for both training and testing [24], [25]. This approach evaluates the performance of hyperparameters across all potential configurations, making it a thorough and exhaustive search technique [26]. In this study, hyperparameter tuning was conducted using Grid Search to enhance the performance of each base model based on neg_mean_squared_error, with five-fold cross-validation (cv=5) ensuring the stability of performance, and n_jobs=-1 utilized to fully leverage all available processors. The optimal parameters identified through Grid Search were subsequently employed for the base learners, as detailed in Table VI.

TABLE V. EVALUATION OF ENSEMBLE LEARNING MODEL - STACKING REGRESSOR

Model	R^2	MAE	MSE
Stacking Regressor	0,845	0,044	0,003

TABLE VI. INITIAL PARAMETER RESULTS AND HYPERPARAMETER TUNING RESULTS WITH GRID SEARCH FOR EACH MODEL

Model	Parameters	Initial Parameters values	Parameter value after tuning
GBM	n_estimators	100	200
	learning_rate	0.1	0.2
	max_depth	3	3
	min_samples_split	2	5
	min_samples_leaf	1	1
	random_state	42	42
XGBoost	n_estimators	200	100
	learning_rate	0.1	0.1
	max_depth	5	6
	subsample	0.8	1.0
	min_child_weight	1	1
	colsample_bytree	-	1.0
	objective	'reg'	'reg'
	random_state	42	42
RF	n_estimators	100	300
	max_depth	None (unlimited)	None (unlimited)
	min_samples_split	2	2
	min_samples_leaf	1	1
	random_state	42	42
RidgeCV (meta-learner)	alphas	np.logspace(-6, 6, 13)	np.logspace(-6, 6, 13)
	store_cv_values	True (opsional)	True (opsional)

After tuning, significant improvements were observed in the models (Table VII). For Gradient Boosting, the number of estimators (n_estimators) increased from 100 to 200, and the learning rate (learning_rate) from 0.1 to 0.2, enhancing learning detail at the risk of overfitting. In XGBoost, n_estimators decreased from 200 to 100, but max_depth and subsample increased, balancing tree depth and data sampling efficiency. For Random Forest, n_estimators increased from 100 to 300, improving model stability and accuracy by averaging more tree predictions.

I. Performance Evaluation of the Stacking Regressor Model

After optimisation, the three base models were combined using the stacked regressor model, where the predictions from each base model became the input for the meta-learner (RidgeCV). Table VII shows the evaluation of the stacked regressor model before and after hyperparameter tuning. Fig. 8 displays the scatter plot between the actual and predicted values for each tuned base model as well as the meta-learner. The points on the stacking regressor are closer to the reference line (y = x), indicating higher prediction accuracy compared to the base model.

TABLE VII. MODEL EVALUATION BEFORE AND AFTER HYPERPARAMETER TUNNING

Model	Before hyperparameter tuning	After hyperparameter tuning	Increase R^2	MAE decrease	MSE Decrease
RF	R^2 = 0,761, MAE = 0,058, MSE = 0,005	R^2 = 0,776, MAE = 0,057, MSE = 0,005	+0,015	-0,001	0,000
GBM	R^2 = 0,767, MAE = 0,058, MSE = 0,005	R^2 = 0,781, MAE = 0,055, MSE = 0,005	+0,014	-0,003	0,000
XGBoost	R^2 = 0,798, MAE = 0,053, MSE = 0,005	R^2 = 0,835, MAE = 0,048, MSE = 0,004	+0,037	-0,005	-0,001
Stacking Regressor	R^2 = 0,845, MAE = 0,044, MSE = 0,003	R^2 = 0,851, MAE = 0,045, MSE = 0,003	+0,006	+0,001	0,000

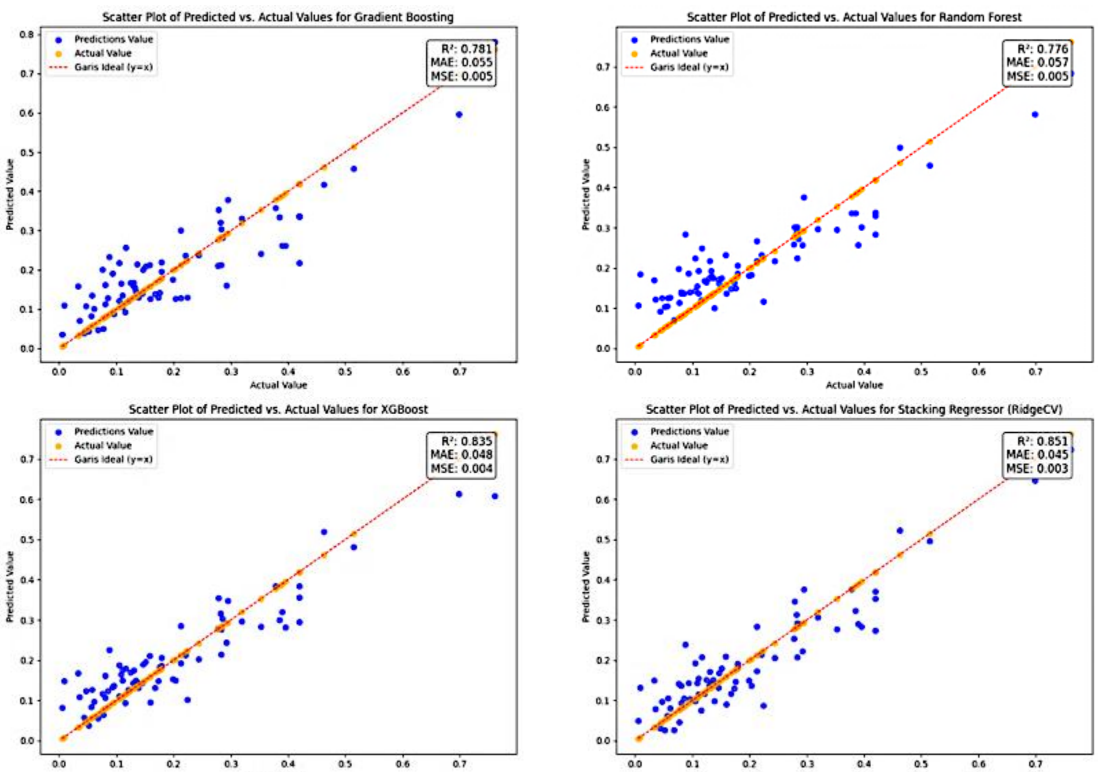


Fig. 8. Scatter plot of hyperparameter tuning performance of prediction model versus actual value.

V. CONCLUSION

This study successfully developed an effective prediction model for $PM_{2.5}$ concentrations caused by forest and peatland fires in Riau Province, employing an ensemble learning approach through the stacking regressor method. The model outperforms other methods, demonstrating superior prediction performance due to the integration of spatiotemporal data from remote sensing and ground sensors. By combining base models such as Random Forest, Gradient Boosting Machine (GBM), and XGBoost, optimized with RidgeCV as a meta-learner, the model achieved optimal performance with $R^2 = 0.851$, MAE = 0.045 µg/m³, and MSE = 0.003 µg/m³. The application of temporal feature engineering techniques, including lag and rolling window, significantly enhanced the model's accuracy, enabling a better understanding of seasonal patterns and temporal dynamics in $PM_{2.5}$ concentrations. Key variables such as air temperature, evapotranspiration, and Aerosol Optical Depth (AOD) were found to have strong correlations with $PM_{2.5}$

concentrations, highlighting the critical role of atmospheric conditions in influencing air pollution levels. This research makes a significant contribution to the development of data-driven air pollution mitigation policies and holds potential for global application in regions facing similar pollution challenges, supporting efforts for more responsive and evidence-based air quality policy planning and public health management.

ACKNOWLEDGMENT

We would like to thank BMKG Sultan Syarif Kasim II Pekanbaru-Riau for its valuable cooperation and contribution in collecting the data on which this research is based. We also express our appreciation to the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for providing financial support for the Doctoral Dissertation Research Grant, so that this research can be carried out properly.

REFERENCES

[1] P. Thangavel, D. Park, and Y. C. Lee, "Recent Insights into Particulate Matter (PM2.5)-Mediated Toxicity in Humans: An Overview," Jun. 01, 2022, MDPI. doi: 10.3390/ijerph19127511.

[2] Ditppu-KLHK, "Kondisi Kualitas Udara Di Beberapa Kota Besar Tahun 2019," Direktorat Pengendalian Pencemaran Udara- KLHK. Accessed: Feb. 06, 2022. [Online]. Available: https://ditppu.menlhk.go.id/portal/kontak-kami/?token=E7fKNFZqQzWdtteaDKXW

[3] N. Islam, S. K. Khan, A. Rehman, U. Aftab, and D. Syed, "Stock Prediction for ARGAAM Companies Dataset," KIET Journal of Computing and Information Sciences, vol. 6, no. 2, pp. 1–13, 2023.

[4] A. Mhawish et al., "Estimation of High-Resolution PM2.5over the Indo-Gangetic Plain by Fusion of Satellite Data, Meteorology, and Land Use Variables," Environ Sci Technol, vol. 54, no. 13, pp. 7891–7900, 2020, doi: 10.1021/acs.est.0c01769.

[5] Z. Y. Chen et al., "Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China," Atmos Environ, vol. 202, pp. 180–189, 2019, doi: 10.1016/j.atmosenv.2019.01.027.

[6] J. Wei et al., "Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach," Remote Sens Environ, vol. 231, no. April, p. 111221, 2019, doi: 10.1016/j.rse.2019.111221.

[7] C. Li, N. C. Hsu, and S. C. Tsay, "A study on the potential applications of satellite data in air quality monitoring and forecasting," Atmos Environ, vol. 45, no. 22, pp. 3663–3675, 2011, doi: 10.1016/j.atmosenv.2011.04.032.

[8] PRIMS, "Data Kondisi Lahan Gambut dan Perkembangan Terkini Terkait Restorasi di Tujuh Provinsi Prioritas BRGM.," Badan Restorasi Gambut. Accessed: Feb. 01, 2022. [Online]. Available: https://prims.brg.go.id/peta

[9] S. Ritung, "Sosialisasi Peta Gambut BBSDLP 2019," in Perubahan Luasan Lahan Gambut Dari Hasil Pemutakhiran Pemetaan Lahan Gambut, Bogor, 2020. [Online]. Available: http://sawitwatch.or.id/wp-content/uploads/2020/12/TSVOL27_Gambut_BBSDLP_021220.pdf

[10] M. Sorek-Hamer, A. W. Strawa, R. B. Chatfield, R. Esswein, A. Cohen, and D. M. Broday, "Improved retrieval of PM2.5 from satellite data products using non-linear methods," Environmental Pollution, vol. 182, pp. 417–423, 2013, doi: https://doi.org/10.1016/j.envpol.2013.08.002.

[11] R. O. Saunders, J. D. W. Kahl, and J. K. Ghorai, "Improved estimation of PM2.5 using Lagrangian satellite-measured aerosol optical depth," Atmos Environ, vol. 91, pp. 146–153, 2014, doi: https://doi.org/10.1016/j.atmosenv.2014.03.060.

[12] J. Zhong et al., "Robust prediction of hourly PM2.5 from meteorological data using LightGBM," Natl Sci Rev, vol. 8, no. 10, 2021, doi: 10.1093/nsr/nwaa307.

[13] M. Unik, I. S. Sitanggang, L. Syaufina, and I. N. S. Jaya, "PM 2 . 5 Estimation using Machine Learning Models and Satellite Data : A Literature Review," vol. 14, no. 5, pp. 359–370, 2023.

[14] A. Shtein et al., "Estimating Daily PM2.5 and PM10 over Italy Using an Ensemble Model," Environ Sci Technol, vol. 54, no. 1, pp. 120–128, Jan. 2020, doi: 10.1021/acs.est.9b04279.

[15] L. K. Hansen and P. Salamon, "Neural network ensembles," IEEE Trans Pattern Anal Mach Intell, vol. 12, no. 10, pp. 993–1001, 1990.

[16] H. Tian, H. Kong, and C. Wong, "A Novel Stacking Ensemble Learning Approach for Predicting PM2.5 Levels in Dense Urban Environments Using Meteorological Variables: A Case Study in Macau," Applied Sciences, vol. 14, no. 12, p. 5062, Jun. 2024, doi: 10.3390/app14125062.

[17] J. Chen, J. Yin, L. Zang, T. Zhang, and M. Zhao, "Stacking machine learning model for estimating hourly PM2.5 in China based on Himawari 8 aerosol optical depth data," Science of the Total Environment, vol. 697, p. 134021, 2019, doi: 10.1016/j.scitotenv.2019.134021.

[18] K. Matsuki, V. Kuperman, and J. A. Van Dyke, "The Random Forests statistical technique: An examination of its value for the study of reading.," Sci Stud Read, vol. 20, no. 1, pp. 20–33, 2016, doi: 10.1080/10888438.2015.1107073.

[19] X. Li, Y. Li, Q. Ma, and S. Wang, "Random Forest Model for PM2.5 Concentration in China Using Himawari-8 Hourly AOD Product," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 1935–1938. doi: 10.1109/IGARSS47720.2021.9554364.

[20] G. Geng, X. Meng, K. He, and Y. Liu, "Random forest models for PM2.5 speciation concentrations using MISR fractional AODs," Environmental Research Letters, vol. 15, no. 3, p. 34056, 2020, doi: 10.1088/1748-9326/ab76df.

[21] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, 1992, doi: https://doi.org/10.1016/S0893-6080(05)80023-1.

[22] Frislidia, "Lebih dari 2.000 hektare lahan terbakar di Riau hingga 8 Oktober 2023," ANTARA, 2023. Accessed: Nov. 19, 2024. [Online]. Available: https://www.antaranews.com/berita/3765324/lebih-dari-2000-hektare-lahan-terbakar-di-riau-hingga-8-oktober-2023

[23] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2nd ed., vol. 2. Springer New York, NY, 2017. doi: https://doi.org/10.1007/978-0-387-84858-7.

[24] A. Chadha and B. Kaushik, "A Hybrid Deep Learning Model Using Grid Search and Cross-Validation for Effective Classification and Prediction of Suicidal Ideation from Social Network Data," New Gener Comput, vol. 40, no. 4, pp. 889–914, Dec. 2022, doi: 10.1007/s00354-022-00191-1.

[25] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction," Clean Eng Technol, vol. 15, Aug. 2023, doi: 10.1016/j.clet.2023.100664.

[26] G. M. Habtemariam, S. K. Mohapatra, and H. W. Seid, "Software reliability prediction using ensemble learning with random hyperparameter optimization," Review of Computer Engineering Research, vol. 11, no. 1, pp. 1–15, Jan. 2024, doi: 10.18488/76.v11i1.3597.