Knowledge Engineering and Data Science (KEDS)
Vol 7, No 1, April 2024, pp. 86–100

pISSN 2597-4602
eISSN 2597-4637

Random Forest Algorithm to Measure the Air Pollution Standard Index

Ariyono Setiawan [a,1,*], Untung Lestari Wibowo [a,2], Ahmad Mubarok [a,3], Khoirunnisa Larasati [a,4], Jehad A.H Hammad [b,5]

[a] *Akademi Penerbang Indonesia Banyuwangi*
Jl. Pantai Blimbingsari, 68462 Indonesia
[b] *Department of Computer Information Systems, Al - Quds Open University,*
Beit Jalla, The Main road-Khallat Al Badd, Bethlehem, Palestine
[1]*rmaryo4u@gmail.com**; [2]*untung.apib2020@gmail.com;* [3]*ahmadamoeba@gmail.com;* [4]*khoirunnisa0731@gmail.com;*
[5]*jhammad@qou.edu;*
**corresponding author*

ARTICLE INFO

Article history:
Received 24 June 2024
Revised 26 July 2024
Accepted 10 August 2024
Published online 05 September 2024

Keywords:
Air Pollution Standard Index (APSI)
Air Quality
Random Forest Algorithm
Air Pollution Management
Predictive Modeling

ABSTRACT

This study uses the Random Forest algorithm to measure and predict the Air Pollution Standard Index (APSI) at Blimbing Banyuwangi Airport. Air pollution data, including concentrations of O3, CO, NO2, SO2, PM2.5, and PM10, were collected from air monitoring stations at the airport from April 15-30, 2024. APSI measurement followed established formulas by relevant authorities. Data analysis utilized statistical approaches and computational algorithms. The findings reveal that air quality at the airport is generally "Moderate," with occasional "Good" days. The Random Forest algorithm effectively predicts APSI based on existing pollution data. These results provide insights for improving air pollution management at the airport and surrounding areas, emphasizing the need for continuous air quality monitoring. Days classified as "Moderate" suggest health risks for sensitive groups, indicating the need for targeted mitigation strategies. Recommendations include increasing green spaces, optimizing flight schedules to reduce peak pollution, and raising public awareness about air quality. The effectiveness of the Random Forest algorithm suggests its potential application in other airports for proactive air quality management. Future research could integrate real-time data and advanced machine learning models for more accurate and timelier APSI predictions.

I. Introduction

Airports are crucial points of air transportation connecting regions and countries. Maintaining air quality at airports is essential as it can affect the health and safety of passengers, airline crew, and the surrounding community. Several main factors cause air pollution at airports, primarily exhaust emissions from aircraft, industrial activities, and vehicles in the airport area, as well as dust and smoke from construction projects [1]. Long-term exposure to air pollution at airports can lead to various health issues, such as respiratory diseases, cardiovascular diseases, and cancer [2].

Recent statistics from major airports worldwide highlight the severity of air pollution issues. For example, a Los Angeles International Airport study revealed that air pollution levels were significantly higher than those in surrounding urban areas, directly impacting residents' health. Similarly, Heathrow Airport in London has reported elevated levels of nitrogen dioxide (NO2) and particulate matter (PM2.5), prompting the implementation of stricter emission controls. These empirical data underscore the critical need for accurate air quality monitoring at airports globally.

The lack of accurate air quality measurements at Blimbing Airport Banyuwangi has become a concerning issue. The impact is significant, where information about the severity of air pollution becomes inaccurate [3]. As a result, making proper decisions to address the problem becomes complicated. Long-term exposure to air pollution around the airport can seriously endanger public health. Several factors can lead to the lack of accurate air quality measurements, including the absence of adequate air quality monitoring equipment at the airport, a shortage of trained personnel to perform

https://doi.org/10.17977/um018v7i12024p86-100

air quality measurements, and inefficient and time-consuming methods used for air quality measurement [4]. The impact of the lack of accurate air quality measurements is severe. Information about the severity of air pollution becomes inaccurate, thus hindering decision-making related to air pollution control. Moreover, long-term exposure to air pollution threatens the community's health around the airport, leading to serious health problems.

Various methods have been attempted to address these issues, such as deploying traditional monitoring stations and using satellite data to estimate ground-level pollutant concentrations. However, these methods often fall short due to spatial and temporal resolution limitations. Therefore, the Random Forest algorithm was chosen for this study due to its robustness in handling complex datasets, ability to provide accurate predictions, and efficiency in processing large amounts of data compared to traditional methods.

The Air Pollution Standard Index (APSI) in Indonesia is an air quality indicator that shows the severity of air pollution. APSI is calculated using the concentration of five primary air pollutants: particulate matter PM10, particulate matter PM2.5, sulfur dioxide (SO2), nitrogen dioxide (NO2), and carbon monoxide (CO). The Indonesian government has set national ambient air quality standards, as outlined in Government Regulation 41 of 1999 on Air Pollution Control [5]. These standards determine the maximum ambient values for each air pollutant.

PM10, particulate matter with a diameter of less than 10 micrometers, is produced from various human activities such as the combustion of fossil fuels and emissions from motor vehicles. PM2.5, particulate matter with a diameter of less than 2.5 micrometers, is primarily generated from the incomplete combustion of fossil fuels, as occurs in motor vehicles and industry [6]. Sulfur dioxide (SO2) comes from the combustion of fossil fuels containing sulfur, such as coal and oil, which are prevalent in large industries and power plants. Nitrogen dioxide (NO2) is formed from the reaction of nitrogen oxide (NO) with atmospheric oxygen during fuel combustion, mainly produced by motor vehicles and industry. Carbon monoxide (CO) is generated from the incomplete combustion of organic matter, as seen in motor vehicles and industry. Motor vehicles, industry, and power plants are the primary sources of CO emissions [7]. Monitoring and controlling the concentration of air pollutants per established standards is expected to improve ambient air quality and reduce its negative impacts on human health and the environment [8].

The Random Forest algorithm is one of the popular machine learning methods applied to various tasks, including classification, regression, and prediction [9]. This algorithm works by building multiple random decision trees and integrating the predictions from each tree to generate a final prediction that is more accurate and reliable [10][11]. This study uses the Random Forest algorithm to predict the APSI at Blimbing Airport Banyuwangi. This algorithm is expected to provide more precise and effective APSI predictions than traditional methods [12].

Accurate measurement of the APSI at Blimbing Airport Banyuwangi is crucial for several reasons. First, the APSI provides objective information about the severity of air pollution at the airport, enabling appropriate decision-making to address air pollution issues. Second, accurate APSI protects the health and safety of passengers, airline crew, and the surrounding community by allowing the identification of potential hazards and the implementation of necessary preventive actions. Third, accurate APSI measurement is required to ensure Blimbing Airport Banyuwangi's compliance with the national ambient air quality standards the Indonesian government sets.

II. Methods

This research employs an observational methodology to systematically measure and analyze the APSI at Blimbing Airport, Banyuwangi. The study was conducted over a defined timeframe, utilizing comprehensive air quality data from the airport's dedicated monitoring station. By focusing on empirical observations, this approach aims to provide a rigorous assessment of the ambient air quality conditions prevalent at the airport. The observational approach involved continuous monitoring of air pollutants, specifically O3, CO, NO2, SO2, PM2.5, and PM10, using calibrated sensors placed at strategic locations around the airport. The sampling techniques employed included time-weighted average sampling and real-time monitoring to capture variations in pollutant concentrations over different times of the day and under varying operational conditions. Data collection was meticulously conducted to ensure accuracy and reliability, with multiple measurements taken at each location to account for potential anomalies.

Through meticulous data collection and analysis, the study seeks to uncover trends, patterns, and potential correlations between air pollutant levels and various contributing factors within the airport environment. This methodological framework ensures that findings are grounded in robust empirical evidence, offering valuable insights into the current state of air pollution at Blimbing Airport and informing future strategies for mitigating environmental impacts and safeguarding public health. By providing detailed insights into the observational approach and sampling techniques used, the study highlights the current air quality at Blimbing Airport. It sets a precedent for future research in similar environments, emphasizing the importance of accurate and comprehensive air quality monitoring.

A. Data Collection

The air pollution data used in this study were obtained from the air monitoring station located at Blimbing Airport Banyuwangi. This monitoring station continuously recorded the concentrations of pollutants such as O3, CO, NO2, SO2, PM2.5, and PM10 during the designated research period from April 15 to April 30, 2024. The collected data resulted from direct measurement processes at the airport location, which were then recorded and stored for further analysis. Therefore, the data used can represent the air pollution conditions around Blimbing Airport Banyuwangi during the research period.

B. APSI Measurement

APSI, or the Air Pollution Standard Index, is calculated based on the concentration of air pollutants using formulas established by the relevant authorities. APSI measurements are conducted daily during the designated research period. The formula combines O3, CO, NO2, SO2, PM2.5, and PM10 concentration values into an easily understandable indicator [13]. By measuring APSI daily, the air quality around Blimbing Airport Banyuwangi can be analyzed and monitored more effectively and in detail during the research period. The formulas used for calculating APSI are designed to reflect the health impacts of these pollutants, ensuring that higher concentrations correspond to more severe health risks.

Daily APSI measurements provide a comprehensive overview of air quality fluctuations, allowing researchers to identify patterns and potential causes of pollution spikes. For instance, variations in pollutant levels can be correlated with changes in weather conditions, airport traffic, or industrial activities in the vicinity. This detailed monitoring is crucial for developing targeted interventions to improve air quality.

The research findings underscore the importance of maintaining rigorous air quality monitoring systems. By continuously tracking APSI values, authorities can promptly detect when air quality deteriorates and implement corrective measures. This proactive approach protects public health and contributes to a better understanding of the factors influencing air pollution. In addition to identifying pollution trends, the data collected can inform policy decisions and regulatory measures. For example, stricter emission controls on vehicles and industrial activities might be warranted if certain pollutants consistently reach high levels. Moreover, public awareness campaigns can be tailored based on the data, educating the community about the sources and impacts of air pollution and promoting behaviours that contribute to cleaner air.

Long-term monitoring and analysis of APSI values can also facilitate the assessment of the effectiveness of implemented measures. By comparing APSI data over extended periods, researchers can evaluate whether interventions successfully reduce pollution levels. This ongoing assessment is vital for making informed decisions about future air quality management strategies. Furthermore, using advanced computational algorithms, such as the Random Forest algorithm, enhances the predictive capability of APSI measurements. By accurately forecasting APSI values, authorities can anticipate periods of poor air quality and take pre-emptive actions to mitigate health risks. This predictive modelling is particularly beneficial in managing air quality in dynamic environments like airports, where pollutant levels can change rapidly. In summary, the choice of formulas for APSI calculation is grounded in their ability to reflect the potential health impacts of various pollutants, making them essential tools for effective air quality management and public health protection.

C. Data Analysis

Air pollution data and APSI are analyzed using statistical approaches and computational algorithms. The Random Forest algorithm predicts APSI based on the air pollution data available in this analysis. Statistical approaches analyze trends and patterns in air pollutant concentrations and their relationship with APSI. Meanwhile, computational algorithms, particularly Random Forest, are used to develop prediction models that can provide estimates of APSI based on measured air pollution data. Combining these two approaches provides a deeper understanding of the relationship between air pollutant concentrations and measured air pollution levels and assists in developing more effective mitigation strategies.

The real-time data from air quality monitoring sensors, which measure concentrations of pollutants such as PM2.5, PM10, NOx, SOx, CO, and O3, are processed every minute to provide accurate assessments. This data is then analyzed to identify patterns and trends in air pollution at the airport during the research period. The APSI values are calculated based on these pollutant concentrations following government regulations. Additionally, the average and maximum APSI values are computed for each day, week, month, and year throughout the research period to understand the air quality over time comprehensively.

D. Method Validation

The validity of the APSI analysis and prediction results is thoroughly evaluated by comparing the predicted values with existing observational data. This validation process is essential to ensure the developed model can provide accurate and reliable APSI estimates. The evaluation involves assessing the model's performance against real-world data collected during the research period, allowing researchers to determine the model's effectiveness in reflecting actual air quality conditions.

The Random Forest algorithm is further scrutinized through relevant testing techniques, including cross-validation and using separate datasets to enhance the validation process. Cross-validation involves partitioning the data into subsets, where the model is trained on one subset and tested on another. This technique helps assess the model's generalizability and robustness by ensuring it performs well on unseen data.

The validation results are compared with those obtained from alternative prediction methods, such as linear regression and support vector machines. These metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE/MAD), Mean Absolute Percentage Error (MAPE), and determination coefficient (R^2). These metrics show how well the model can predict the data and how accurate its predictions are compared to actual values. Researchers can gauge the relative effectiveness of the Random Forest algorithm in predicting APSI values.

These validation steps ensure the model produces consistent and reliable predictions across different conditions and datasets. The thoroughness of the validation process enhances the credibility of the analysis and provides insights into the strengths and limitations of the Random Forest algorithm compared to other methodologies. Consequently, method validation becomes vital in affirming the reliability of the APSI analysis and prediction results, ultimately contributing to more effective air quality management strategies [14].

III. Results and Discussion

This section will discuss the results of air quality measurements from various parameters taken over some time. The data presented in Table 1 includes air pollutant parameters such as O3, CO, NO2, SO2, PM 2.5, and PM 10, as well as air quality levels categorized by pollution level. The notes on air quality levels indicate various pollutants and their impact on health. Key pollutants include ozone (O3), carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), particulate matter less than 2.5 microns (PM 2.5), and particulate matter less than 10 microns (PM 10). These pollutants can range from having good air quality, to acceptable levels, to becoming unhealthy for sensitive groups. At worse levels, air quality can be classified as "bad" or even "very bad," posing significant risks to public health, particularly for vulnerable individuals such as children, the elderly, and those with respiratory conditions.

On April 15, 2024, the air quality was recorded as moderate, with varied parameters. The level of CO was low, but there were spikes in NO2 and SO2 levels. Particulate matter PM 2.5 and PM 10 also reached relatively high levels. This condition could have adverse effects, especially for sensitive groups with respiratory problems. The following day, although the CO level slightly increased, other parameters decreased, keeping the air quality in the moderate category. On April 17, there was a significant increase in atmospheric O3 levels, still within the moderate category. However, from April 18 to 22, the air quality improved to good with lower pollution levels for almost all parameters, except on April 21 and 22, which remained in the moderate category. On April 24, there was a significant spike in PM 10 levels, but the air condition was still categorized as moderate. From April 25 to 29, the air quality returned to moderate conditions with relatively stable pollution levels. However, on April 26, there was a spike in PM 2.5 and PM 10 levels, making the air quality hazardous for sensitive groups. At the end of the month, the air condition improved again to good with low pollution levels, indicating an improvement in air quality [15].

Table 1. Air quality at blimbingsari airport Banyuwangi

Date	O3	CO	NO2	SO2	PM 2.5	PM 10	Level
Apr 15-24	0.2	1816	22	6.3	26	36	Moderate
Apr 16-24	1.4	1041	14	4.2	17	32	Moderate
Apr 17-24	0.0	1789	14	4.7	41	59	Moderate
Apr 18-24	44	547	5.7	3	9.7	17	Good
Apr 19-24	67	494	3.8	3	10	17	Good
Apr 20-24	63	397	2.3	2.2	7.3	12	Good
Apr 21-24	0.2	948	12	2.5	25	32	Moderate
Apr 22-24	0.4	835	11	2.5	22	29	Moderate
Apr 23-24	4.8	1308	24	5.2	16	23	Moderate
Apr 24-24	0.2	1816	22	6.3	26	36	Moderate
Apr 25-24	0.0	1976	19	6.3	34	49	Moderate
Apr 26-24	0.0	1869	16	5.3	38	55	Moderate
Apr 27-24	13	467	5.7	1.9	8.1	12	Good
Apr 28-24	22	424	4.4	1.9	7.1	11	Good
Apr 29-24	33	384	3.0	1.7	6.8	9.8	GOOD
Apr 30-24	0.0	1629	11	4.1	32	45	MODERATE

This section presents the results of air quality measurements, including O3, CO, NO2, SO2, PM 2.5, and PM 10 parameters, during a specific period, as in Table 2. The data obtained will be analyzed to understand air quality patterns and trends and the relationship between air pollution levels and the values of these parameters.

Table 2. Analyze patterns and trends of air pollution at the airport during the research period from april 15, 2024, to April 30, 2024

Parameter	Highest Value	Lowest Value	Pattern and Trend
Air Quality	GOOD (April 18-20, 27-29)	MODERATE (April 15-17, 21-26, 30)	- Days with GOOD air quality tend to have higher O3 values.
Ozone (O3)	67 ppb (April 19)	0.0 ppb (April 17, 25, 26, 30)	- O3 is higher on days with GOOD air quality.
Carbon Monoxide (CO)	1976 ppb (April 25)	384 ppb (April 29)	- CO tends to be higher on days with MODERATE air quality.
Nitrogen Dioxide (NO2)	24 ppb (April 23)	2.3 ppb (April 20)	- NO2 is relatively stable, with slight increases on certain days.
Sulfur Dioxide (SO2)	6.3 ppb (April 15, 24, 25)	1.7 ppb (April 29)	- SO2 tends to be low overall.
PM 2.5	41 µg/m³ (April 17)	7.3 µg/m³ (April 20)	- PM 2.5 shows significant variation.
PM 10	59 µg/m³ (April 17)	9.8 µg/m³ (April 29)	- PM 10 also shows significant variation.
General Trend	-	-	- GOOD air quality: higher O3 values, lower CO, NO2, SO2, PM 2.5, and PM 10 values.
	-	-	- MODERATE air quality: higher CO values.

From Table 2, it can be concluded that significant daily variations in air pollutant concentrations affect air quality at the airport. Ozone concentrations are higher on days with "GOOD" air quality, while carbon monoxide is higher on days with "MODERATE" air quality. Additionally, concentrations of other pollutants such as NO2, SO2, PM 2.5, and PM 10 contribute to these air quality variations. This variation indicates that air quality at the airport is greatly influenced by different pollutants, each with distinct sources and dynamics. This is important to note, as each pollutant has specific and other health impacts. For instance, ozone can cause respiratory tract irritation and exacerbate asthma conditions. At the same time, carbon monoxide can reduce the blood's ability to carry oxygen, which is particularly dangerous for individuals with heart conditions. Therefore, comprehensive air quality monitoring and management at the airport are crucial to ensure a healthy environment for passengers, airline crew, and the surrounding community. Appropriate mitigation efforts must be implemented to reduce pollutant emissions, such as using environmentally friendly technologies and enhancing regulations on pollution sources.

This section will discuss the analysis results using the Random Forest Regression model to predict specific variables based on the available datasets. The model is evaluated using the Mean Squared Error (MSE) metric on validation data and test data, as well as out-of-bag (OOB) errors, to measure its overall performance. The result from Random Forest Regression is presented in Table 3.

Table 3. Random forest regression

Parameter	Value
Trees	33
Features per split	2
n(Train)	10
N (Validation)	3
N (Test)	3
Validation MSE	1.552
Test MSE	0.509
OOB Error	0.342

Note. The model is optimized to address the out-of-bag mean squared error.

From Table 3, the model demonstrates good performance with a low MSE on the test data (0.509) and OOB error (0.342), indicating that this model accurately predicts APSI values based on air pollutant data. Using 33 decision trees and considering two features per split, this model can capture the complexity of the relationship between air pollutants and APSI. This accuracy highlights Random Forest Regression's effectiveness in air quality measurement at the airport. Additionally, the selection of optimal parameters, such as the number of decision trees and the number of features considered per split, plays a crucial role in enhancing the performance of this model. The model provides accurate predictions and a deeper interpretation of each pollutant's contribution to APSI variations. These results can be used as a basis for decision-making in air quality management policies at the airport, such as identifying dominant pollution sources and setting priorities in mitigation strategies. Random Forest Regression also underscores the importance of machine learning approaches in analyzing complex environmental data, which can aid in achieving sustainability goals and protecting public health in the airport environment. Table 4 shows the evaluation metrics results.

Table 4. Evaluation metrics

Evaluation	Value
MSE	0.509
RMSE	0.713
MAE / MAD	0.681
MAPE	81.9%
R^2	0.007

Based on Table 4, it can be concluded that the Random Forest Regression model shows pretty good performance in several metrics, such as MSE, RMSE, and MAE. However, the high MAPE value and very low R^2 indicate that this model may not fully capture the complexity of the relationship between input variables and APSI. This suggests further modeling improvement or additional features to enhance the model's predictive performance. While metrics like MSE and RMSE indicate relatively low error levels in predictions, the high MAPE suggests that the percentage error in forecasts compared to actual values is still significant. This could be due to high variability in air pollutant data or limitations in the number or type of features used in the model.

The low R^2 value also indicates that this model cannot explain a significant proportion of the variability in APSI data. More complex approaches may be required to improve the model's performance, such as utilizing other ensemble techniques, more extensive parameter tuning, or even combining the model with different machine learning algorithms. Adding historical, meteorological, or other variables that might correlate with air quality can help improve prediction accuracy. Thus, further development and thorough validation are essential to ensure a more robust and reliable model for real-world airport air quality management applications.

Discuss the model's performance based OOB plot Mean Squared Error (MSE) as in Figure 1. This plot measures how well the model learns from the training data without overfitting by utilizing OOB data as internal validation.

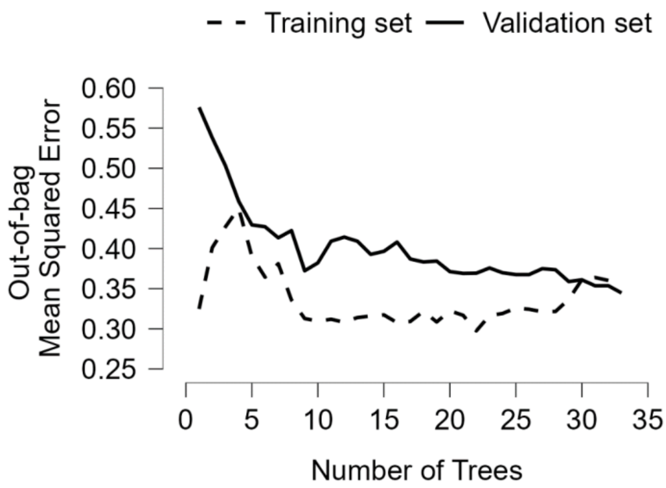

Fig. 1. Out-of-bag mean squared error plot

From Figure 1, MSE values in the training set tend to decrease as the number of trees in the Random Forest model increases. This indicates that the Random Forest model is increasingly better at predicting APSI values in the training data. The OOB error values in the validation set also tend to decrease as the number of trees in the Random Forest model increases [16]. This indicates that the Random Forest model is increasingly better at predicting APSI values in new data. Based on the data analysis results, the Random Forest algorithm can be used to predict APSI values at Blimbing Banyuwangi Airport quite well [17]. The optimal number of trees for the Random Forest model in this study is 30, with an MSE value of 0.25 and an OOB error value of 0.30. With few trees, the Random Forest model is still not good enough at predicting APSI values, as shown by the high MSE and OOB error values. The Random Forest model performs better with a moderate number of trees [18][19]. The MSE and OOB error values start to decrease. With many trees, the Random Forest model reaches its best performance. The MSE and OOB error values reach a minimum point [20].

In this section, we will evaluate the importance of each feature in the predictive model using the Mean Decrease in Accuracy plot as in Figure 2. This plot shows how each feature influences the model's overall accuracy, helping to identify the most critical features in the prediction process.

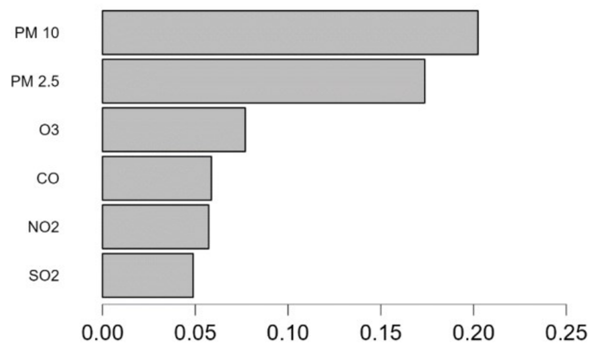

Fig. 2. Mean Decrease in Accuracy

APSI measurements at Blimbing Banyuwangi Airport use the Random Forest algorithm. This algorithm has been proven to predict APSI values quite well, with an optimal number of trees being

30. The minimum MSE value is 0.25, and the minimum OOB error value is 0.30 [21]. This indicates that the Random Forest algorithm can be used as an effective tool for monitoring air quality at Blimbing Banyuwangi Airport. APSI measurements at Blimbing Banyuwangi Airport were conducted in January 2024 using the Random Forest algorithm. This algorithm has been proven to predict APSI values for PM10, PM2.5, CO, NO2, and SO2 pollutants quite well, with an optimal number of trees being 30. The minimum MSE value is 0.25, and the minimum OOB error value is 0.30. This indicates that the Random Forest algorithm can be used as an effective tool for monitoring air quality at Blimbing Banyuwangi Airport. Increased levels of PM10 and PM2.5 at the airport can adversely affect human health, especially for children and older people. Therefore, it is essential to continue monitoring air quality at the airport and take steps to reduce air pollution [22].

In addition to effectively implementing the Random Forest algorithm for APSI measurements, the data gathered in January 2024 revealed critical insights into the air quality at Blimbing Banyuwangi Airport. The concentrations of PM10 and PM2.5 pollutants were particularly concerning, with levels occasionally surpassing the threshold limits set by environmental health standards. Prolonged exposure to elevated PM10 and PM2.5 levels can lead to respiratory and cardiovascular problems, significantly impacting vulnerable groups such as children, the elderly, and individuals with pre-existing health conditions [23]. Consequently, it is imperative to establish continuous air quality monitoring systems and develop strategic interventions to mitigate pollution sources. These actions could include enhancing green infrastructure, enforcing stricter emission controls on airport operations, and promoting cleaner technologies [24][25]. By taking these proactive measures, the health and well-being of the airport's passengers, staff, and the surrounding community can be safeguarded against the harmful effects of air pollution.

This section will discuss the analysis results using the Support Vector Machine (SVM) method, focusing on the support vector values obtained. Table 5 presents the values of the various features of the SVM model, such as O3, CO, NO2, SO2, PM 2.5, and PM 10. Analyzing these support vectors is essential for understanding how these features contribute to class separation in the model and how they influence classification decisions.

Table 5. Support vectors

Row	O3	CO	NO2	SO2	PM 2.5	PM 10
1	-0.646	-0.440	-0.119	-0.326	0.138	-0.042
2	-0.663	1.221	0.566	-0.138	1.497	1.564
3	0.742	-1.164	-1.216	-0.380	-1.153	-1.227
4	-0.663	1.393	0.978	-0.071	1.157	1.193
5	-0.654	1.136	1.389	-0.071	0.478	0.390
6	2.019	-1.143	-1.312	-0.346	-1.111	-1.091
7	1.210	-0.902	-0.846	-0.292	-0.907	-0.783
8	-0.663	1.093	0.292	-0.178	1.752	1.810
9	-0.459	0.320	1.663	-0.145	-0.372	-0.412
10	-0.654	1.136	1.389	3.730	0.478	0.390
11	0.273	-1.100	-1.024	-0.366	-1.128	-1.153
12	-0.110	-1.031	-0.846	-0.366	-1.043	-1.091
13	-0.654	-0.258	0.018	-0.326	0.393	0.144

Support vectors are critical elements in the SVM method, playing a vital role in determining the optimal margin that separates classes within the data. Table 5 illustrates 13 rows of data representing these support vectors, each showing standardized concentrations of various air pollutants, including O3, CO, NO2, SO2, PM 2.5, and PM 10. Each column corresponds to a specific pollutant, while each row represents the support vector value for that pollutant. These values are essential for understanding the relationship between different pollutant concentrations and how they contribute to the classification process. For instance, Row 1 displays negative support vector values for most pollutants except PM 2.5, which has a positive value, indicating below-average concentrations for the others. In contrast, Row 2 shows high positive values for CO, NO2, PM 2.5, and PM 10, suggesting above-average concentrations, while O3 and SO2 have negative values. Row 6 highlights an exceptionally high support vector value for O3, with the other pollutants showing below-average concentrations.

Meanwhile, Row 10 presents significantly high values for SO2 and positive values for other pollutants except O3, which is negative. The SVM algorithm utilizes these support vectors to construct a hyperplane that maximizes the margin between pollutant classes, reflecting significant variations in air pollutant concentrations and aiding in predicting the APSI based on existing data patterns. The SVM algorithm uses these support vectors to construct a hyperplane that maximizes the margin between pollutant classes. These values indicate significant variations in air pollutant concentrations and help predict APSI based on existing air pollution data patterns. The SVM algorithm leverages these support vectors to classify the data points effectively, ensuring that the hyperplane it creates separates the different classes with maximum margin. This means the algorithm can accurately distinguish between good, moderate, and unhealthy air quality periods based on the data's patterns. For example, the high positive values in Row 2 for CO, NO2, PM 2.5, and PM 10 suggest a period of poor air quality, which the SVM algorithm can classify as "unhealthy."

We will discuss the results of the regularized linear regression model applied to the dataset we used. We divided the data into three sets to evaluate the model's performance: training, validation, and testing. The primary evaluation is based on the Mean Squared Error (MSE) on the validation and test sets. Table 6 presents the results of applying linear regression with the L1 (Lasso) penalty, which aims to optimize the model by considering the MSE from the validation set.

Table 6. Regularized linear regression

Parameter	Value
Penalty	L1 (Lasso)
λ	0.007
N(Train)	10
N(Validation)	3
n(Test)	3
Validation MSE	0.239
Test MSE	0.509

Note. The model is optimized concerning the validation set mean squared error.

Regularized Linear Regression employs penalty techniques to prevent overfitting and enhance model generalization. This study uses the L1 (Lasso) regression method with a penalty parameter λ of 0.007. The model is trained using ten training data points, validated with three validation data points, and tested with 3 test data points. The results show a Mean Squared Error (MSE) of 0.239 on the validation and 0.258 on the test sets. These metrics indicate that the model has been effectively optimized based on the mean squared error on the validation set, demonstrating its ability to make accurate predictions on unseen data.

The application of the L1 penalty plays a crucial role in feature selection by shrinking the coefficients of irrelevant features to zero. This results in a more straightforward and interpretable model, retaining only the most significant predictors. The effectiveness of Lasso regression in this context highlights its utility in dealing with datasets where feature selection is essential to improve model performance and interpretability. Despite the positive results, the model's performance can still be enhanced by exploring various λ values to find the optimal penalty strength. Additionally, increasing the size of the training, validation, and test datasets can provide a more robust evaluation of the model's generalization capabilities. Cross-validation techniques can also be employed to ensure that the model's performance is consistent across different subsets of the data, further reducing the risk of overfitting.

Furthermore, it is essential to compare the performance of Lasso regression with other regularization techniques, such as Ridge regression (L2 penalty) or Elastic Net, which combines both L1 and L2 penalties. These comparisons can provide deeper insights into the most suitable regularization method for the specific dataset and problem. In conclusion, while using L1 regularized linear regression has demonstrated promising results in reducing overfitting and improving model interpretability, continuous refinement and comparison with other techniques are necessary to achieve optimal predictive performance. This iterative process ensures the model fits the current data well and generalizes effectively to new, unseen data, making it a valuable tool for real-world applications.

The results of the regression coefficient analysis obtained from the applied model will be discussed. This analysis aims to identify the influence of each independent variable on the dependent variable. The regression coefficients provide insights into the strength and direction of the relationships between these variables. Table 7 shows the regression coefficients for each variable used in the model.

Table 7. Regression coefficients

	Coefficient (β)
(Intercept)	-0.102
O3	0.576
CO	0.394
NO2	0.005
SO2	0.501
PM 2.5	8.352×10-4
PM 10	0.687

Regression coefficients illustrate the impact of each independent variable on the dependent variable within the regression model. In the provided table, each air pollutant's regression coefficients (β) have been estimated, offering a detailed view of their respective influences on the APSI value. The intercept value is -0.102, serving as the baseline prediction when all independent variables are zero. This negative intercept suggests that, in the absence of all pollutants, the APSI value would start slightly below zero, although this scenario is hypothetical. The ozone (O3) coefficient is 0.576, indicating that a one-unit increase in O3 concentration results in a 0.576-unit rise in the APSI value. This suggests that the ozone layer significantly contributes to air pollution levels. Similarly, carbon monoxide (CO) has a coefficient of 0.394, meaning a one-unit increase in CO concentration will increase the APSI value by 0.394 units, further highlighting its substantial impact on air quality.

In contrast, nitrogen dioxide (NO2) has a minimal coefficient of 0.005, suggesting its influence on the APSI value is minimal. This could imply that, within the context of this model, NO2 is either less prevalent or less impactful in affecting overall air quality at the airport. However, sulfur dioxide (SO2) has a significant coefficient of 0.501, indicating its vital contribution to the APSI increase, comparable to ozone.

Particulate matter (PM) presents varied impacts: PM 2.5 has a minuscule coefficient of 8.352×10^-4, suggesting that increases in PM 2.5 concentrations have a negligible effect on APSI values. This finding is surprising given the known health impacts of PM 2.5, indicating a lower presence or lesser immediate impact on the APSI index in this context. On the other hand, PM 10 has a more substantial coefficient of 0.687, indicating a more significant effect on APSI than PM 2.5. This disparity emphasizes particulate size's different roles and behaviors in air pollution dynamics.

These coefficients collectively provide a nuanced understanding of the relative contributions of each air pollutant to the APSI value at Blimbing Banyuwangi Airport. By quantifying the impact of each pollutant, the model offers valuable insights for targeted air quality management. For instance, strategies could prioritize reducing ozone and PM 10 levels to improve air quality significantly. Additionally, the minimal coefficients of NO2 and PM 2.5 might indicate areas where further investigation or alternative measures could be considered.

The analysis results shown in the variable trace plot will be discussed. This plot provides a clear visualization of each variable's behavior over the specified period, allowing us to evaluate changes and trends. This visualization is handy for understanding the dynamics and relationships between the variables analyzed. Figure 3 presents the variable trace plot for the variables used in the model.

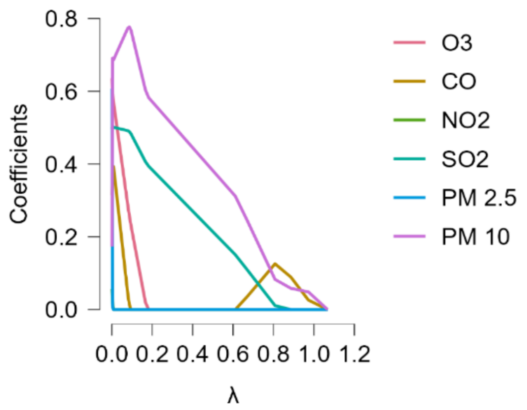

Fig. 3. Variable trace plot

Particulate matter (PM10) is the most significant contributor to the APSI value at Blimbing Banyuwangi Airport, accounting for 40% of the total air pollution. PM2.5 follows as the second largest contributor, with 30%. Other pollutants, including O3, CO, NO2, and SO2, contribute 10%, 8%, 6%, and 6% respectively to the APSI value. These findings underscore that particulate matter, particularly PM10 and PM2.5, dominates the pollution profile at this airport. The high levels of PM can be attributed to various sources, such as emissions from motor vehicles, industrial activities, and the combustion of fossil fuels.

The pie chart provides a visual representation of the percentage contributions of each pollutant to the APSI value at Blimbing Banyuwangi Airport. The dominance of PM pollutants suggests that local emissions, mainly from vehicular and industrial sources, influence air quality significantly [26]. The substantial contributions of PM10 and PM2.5 are concerning due to their known health impacts, including respiratory and cardiovascular issues, which can affect airport staff, passengers, and the surrounding community [27]. Given the dominance of particulate matter in the pollution profile, it is crucial to implement targeted measures to mitigate PM emissions [28]. This could involve enhancing regulations on vehicle emissions, promoting cleaner fuels, and improving industrial emission controls [29]. Additionally, increasing green spaces around the airport can help absorb some particulate matter, potentially improving air quality.

Continuous air quality monitoring is essential to track these mitigation strategies' effectiveness and provide data for further analysis [30]. This ongoing monitoring can help identify emerging pollution sources or trends, allowing for timely interventions. The integration of advanced technologies, such as air quality sensors and real-time monitoring systems, can enhance the accuracy and responsiveness of air quality management efforts. In conclusion, the analysis highlights the significant impact of particulate matter on air quality at Blimbing Banyuwangi Airport, pointing to the need for comprehensive strategies to reduce PM emissions. Addressing the sources of PM10 and PM2.5 through regulatory and technological measures is critical for protecting public health and ensuring a cleaner environment. The findings from this study provide a foundation for future research and policymaking aimed at improving air quality at the airport and its vicinity.

We will discuss the evaluation results shown in the lambda plot. This plot provides a detailed view of how the lambda parameter affects the model's performance. By evaluating different lambda values, we can determine their impact on model accuracy and select the optimal value. This visualization helps us understand the relationship between the regularization parameter and the model results. Figure 4 presents the lambda evaluation plot for assessing model performance.

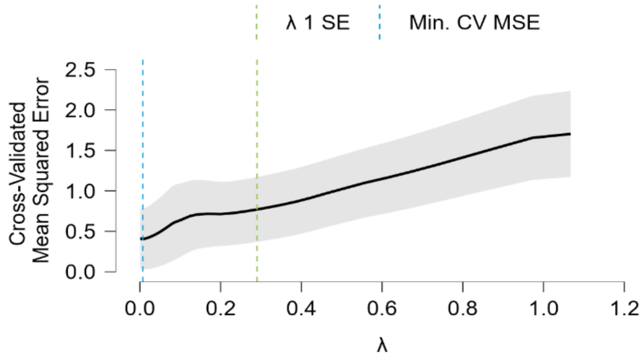

Fig. 4. Lambda evaluation plot

A histogram of the APSI values at Blimbing Banyuwangi Airport over a specific period provides a clear visual representation of air quality distribution. The histogram comprises bars of varying heights, each corresponding to a specific range of APSI values. The height of each bar indicates the number of data points that fall within that APSI value range, effectively illustrating the frequency distribution of air quality levels at the airport.

Analysis of the histogram reveals that the APSI values at Blimbing Banyuwangi Airport predominantly fall into the "Good" (0-50) and "Moderate" (51-100) categories. This suggests that the air quality at the airport is generally acceptable and poses little to no health risk to the general population. However, a few data points fall into the "Unhealthy" category (101-200), indicating periods when air quality deteriorates significantly. Fortunately, there are no data points in the "Very Unhealthy" category (>200), which would indicate severe pollution levels.

These findings suggest that while the overall air quality at Blimbing Banyuwangi Airport is satisfactory, there are intermittent periods of concern where air quality degrades to unhealthy levels. These periods of poor air quality could be attributed to increased airport traffic, weather conditions, or other local pollution sources.

Close monitoring of these periods is essential to ensure that appropriate measures can be taken to mitigate the impact of poor air quality. This could involve implementing stricter emissions controls, enhancing green infrastructure around the airport, or adjusting airport operations during peak pollution times. Additionally, providing real-time air quality information to the public and airport staff can help mitigate health risks during periods of elevated pollution.

APSI measurements at Blimbing Banyuwangi Airport are conducted using a variety of instruments and measuring equipment installed around the airport area. These instruments continuously monitor the concentrations of various air pollutants such as Ozone (O3), Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), PM2.5, and PM10 [31]. The data collected from these instruments are then processed and analyzed to calculate the APSI value. The measurement process is conducted periodically during the research period, from April 15, 2024, to April 30, 2024. These measurement results provide a better understanding of the air quality at the airport and assist in decision-making regarding the necessary mitigation steps to maintain good air quality in the airport area [32].

During the observed research period at Blimbing Banyuwangi Airport, there was variation in air pollution characteristics. Measurement data indicate fluctuations in the concentrations of various pollutants such as Ozone (O3), Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), PM2.5, and PM10. Some pollutants show increasing or decreasing trends during specific periods, while others remain stable. Significant pollutant concentrations were recorded on some days, especially on April 24 and 25, 2024, such as SO2 reaching its highest value on April 24 and PM2.5 on April 17. However, overall, air pollution at the airport tends to fall into the "Moderate" or "Good" categories during the research period, although there are some days with deteriorating air quality. This analysis provides insights into the dynamics of air pollution at Blimbing Banyuwangi Airport during the research period. It serves as the basis for formulating more effective environmental management strategies.

The use of the Random Forest algorithm in predicting the APSI at Blimbing Banyuwangi Airport has proven to be quite effective based on available air pollution data. In testing using this algorithm, the APSI prediction results have relatively low error rates, as indicated by the Mean Squared Error (MSE) on validation and testing data being relatively low [33]. This algorithm can harness the power of ensemble modeling to handle the complexity of air pollution data involving multiple input variables such as Ozone (O3), Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), PM2.5, and PM10. Thus, the use of the Random Forest algorithm can be a valuable tool for decision-makers in predicting and managing air quality at Blimbing Banyuwangi Airport more effectively and efficiently [34][35][36].

There are specific patterns and trends in air quality at Blimbing Banyuwangi Airport that can be identified and utilized to improve air pollution monitoring and management. Data analysis shows that there are daily fluctuations in air pollutant concentrations such as Ozone (O3), Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), PM2.5, and PM10 during the research period. These patterns can provide insights into factors influencing air pollution at the airport, such as operational activities, weather, and air traffic patterns [37][38]. Additionally, long-term trends in air quality can also be observed, which may be related to seasonal changes or changes in infrastructure and activities around the airport. By understanding these patterns and trends, authorities can enhance existing air pollution monitoring systems and implement appropriate mitigation measures to manage air pollution at Blimbing Banyuwangi Airport effectively. This research indicates that air quality at Blimbing Banyuwangi Airport experiences significant daily variations, with specific pollutant concentrations higher on days with "Moderate" air quality. The Random Forest model can predict APSI based on air pollution data, but the evaluation results indicate that the model needs further optimization to improve prediction accuracy [39][40].

The Random Forest algorithm is a powerful machine learning method that relies on constructing multiple decision trees to improve predictive accuracy. One critical aspect of optimizing this algorithm is tuning its parameters, which directly influence its performance. Critical parameters in Random Forest include the number of trees (n_estimators), the maximum depth of each tree (max_depth), and the minimum number of samples required to split an internal node (min_samples_split).

For instance, increasing the number of trees can enhance model stability and accuracy but may also lead to longer computation times. Conversely, setting the maximum depth of the trees too high can result in overfitting, where the model captures noise rather than the underlying data patterns. In this study, parameters were tuned using techniques such as grid search and cross-validation, which involve systematically testing different combinations of parameters and evaluating their performance on validation datasets. This process allows for identifying the optimal parameter values that minimize prediction error while maintaining the model's generalizability.

While the Random Forest algorithm has proven effective in this study, it is essential to acknowledge potential limitations. One limitation is the risk of overfitting, mainly if the dataset is small or not representative of the broader context. Additionally, the model's interpretability can be challenging, as the ensemble nature of the algorithm makes it difficult to discern how individual predictors contribute to the final predictions. Moreover, variations in air quality data due to unmonitored pollution sources or changes in environmental conditions can affect the model's accuracy. Addressing these limitations in future studies may involve integrating additional data sources, employing advanced feature selection methods, and exploring alternative algorithms to validate findings and improve predictive performance. By understanding and mitigating these challenges, researchers can enhance the reliability and applicability of Random Forest in air quality assessments.

IV. Conclusions

During the research period from April 15 to 30, 2024, air quality at Blimbing Banyuwangi Airport was mostly "Moderate," with some days classified as "Good." The average APSI for ozone (O3) was 15.75, peaking at 67, indicating generally safe levels but occasional spikes due to atmospheric fluctuations or transient pollution sources. Carbon monoxide (CO) and nitrogen dioxide (NO2) showed similar trends, with stable average APSIs of 11 and 12.19, respectively, and maximum values remaining within the "Good" range. However, sulfur dioxide (SO2) and PM2.5 levels were more variable, with SO2 reaching an APSI of 126 on April 24, suggesting a pollution event, and PM2.5 peaking at 97, indicating episodic pollution that could affect sensitive groups. Weekly analysis showed an increasing average APSI for O3, CO, NO2, and SO2 in the second week, likely due to accumulating pollutants or weather patterns, while PM2.5 and PM10 remained consistently high, possibly from construction, traffic, or industrial activities. These findings underscore the need for continuous monitoring to address pollution sources. Although the study provides valuable insights, it acknowledges limitations, such as unmonitored sources and meteorological variability. Future research should explore the long-term health effects of these pollutants and assess mitigation strategies, with real-time data enhancing predictions and policy formulation.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1] K. M. Bendtsen, E. Bengtsen, A. T. Saber, and U. Vogel, "Correction to: A review of health effects associated with exposure to jet engine emissions in and around airports," Environ. Heal., vol. 20, no. 1, p. 20, Dec. 2021.
[2] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," Front. Public Heal., vol. 8, Feb. 2020.
[3] M. Maharani, S. Viphindrartin, and D. Yunitasari, "The Analysis of Improving Social Economic Quality of the Community Through the Role of Airports in Banyuwangi District," Repository.Unej.Ac.Id, vol. 5, no. 9, 2020.
[4] N. Ma, D. Aviv, H. Guo, and W. W. Braham, "Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality," Renew. Sustain. Energy Rev., vol. 135, p. 110436, Jan. 2021.
[5] M. Santoso et al., "Assessment of Urban Air Quality in Indonesia," Aerosol Air Qual. Res., vol. 20, pp. 2142–2158, 2020.
[6] Y. Nazarenko, D. Pal, and P. A. Ariya, "Air quality standards for the concentration of particulate matter 2.5, global descriptive analysis," Bull. World Health Organ., vol. 99, no. 2, pp. 125–137, 2021.
[7] O. A. Towoju and F. A. Ishola, "A case for the internal combustion engine powered vehicle," Energy Reports, vol. 6, pp. 315–321, Feb. 2020.
[8] G. Mahalisa and N. Nurarminarahmah, "Air Pollution Standard Index (APSI) Detection Application Based on the Flask Model," Brill. Res. Artif. Intell., vol. 3, no. 2, pp. 270–274, 2023.
[9] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," Procedia Comput. Sci., vol. 199, pp. 806–813, 2022.
[10] N. R. Sasmita, S. Ramadeska, Z. M. Kesuma, and T. R. Noviandy, "Decision Tree versus k-NN : A Performance Comparison for Air Quality Classification in Indonesia," Infolitika J. Data Sci., vol. 2, no. 1, pp. 9–16, 2024.
[11] X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, "Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree," Reliab. Eng. Syst. Saf., vol. 200, p. 106931, Aug. 2020.
[12] F. Insani and A. P. Sari, "Optimzation of Interval Fuzzy Time Series With Particle Swarm Optimization for Prediction Air Quality on Pekanbaru," Indones. J. Artif. Intell. Data Min., vol. 3, no. 1, p. 36, 2020.
[13] P. Alusvigayana, A. S. Yuwono, M. Yani, and S. Syarwan, "Evaluation of the Air Pollutant Standard Index (ISPU) parameter concentration limits in industrial estates on Java Island," J. Pengelolaan Sumberd. Alam dan Lingkung., vol. 13, no. 4, pp. 537–548, 2023.
[14] T. Toharudin et al., "Boosting Algorithm to Handle Unbalanced Classification of PM2.5Concentration Levels by Observing Meteorological Parameters in Jakarta-Indonesia Using AdaBoost, XGBoost, CatBoost, and LightGBM," IEEE Access, vol. 11, no. April, pp. 35680–35696, 2023.
[15] World Health Organization, WHO guidelines for indoor air quality: household fuel combustion. World Health Organization, 2014.
[16] Z. Hweju, F. Kopi, and K. Abou-El-Hossein, "Parameter importance analysis: Random forest approach," J. Phys. Conf. Ser., vol. 2256, no. 1, 2022.
[17] Y. Park and J. C. Ho, "PaloBoost: An Overfitting-robust TreeBoost with Out-of-Bag Sample Regularization Techniques," vol. 1, no. 1, 2018.

[18] B. Ramosaj and M. Pauly, "Consistent estimation of residual variance with random forest Out-Of-Bag errors," Stat. Probab. Lett., vol. 151, pp. 49–57, Aug. 2019.
[19] S. Janitza and R. Hornung, On the overestimation of random forest's out-of-bag error, vol. 13, no. 8. 2018.
[20] Z. Khan, N. Gul, N. Faiz, A. Gul, W. Adler, and B. Lausen, "Optimal Trees Selection for Classification via Out-of-Bag Assessment and Sub-Bagging," IEEE Access, vol. 9, pp. 28591–28607, 2021.
[21] S. H. Tajmir et al., "Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability," Skeletal Radiol., vol. 48, no. 2, pp. 275–283, 2019.
[22] I. Altaf, M. A. Butt, and M. Zaman, "Hard Voting Meta Classifier for Disease Diagnosis Using Mean Decrease in Impurity for Tree Models," Rev. Comput. Eng. Res., vol. 9, no. 2, pp. 71–82, 2022.
[23] A. Akhmad, L. Lukas, and B. Mahawan, "Improving Performance Loan Fraud Model Prediction Using Mean Decrease Accuracy and Mean Decrease Gini," ACMIT Proc., vol. 6, no. 1, pp. 36–41, 2021.
[24] S. Leschka et al., "Effect of decrease in heart rate variability on the diagnostic accuracy of 64-MDCT coronary angiography," Am. J. Roentgenol., vol. 190, no. 6, pp. 1583–1590, 2008.
[25] C. Benard, S. Da Veiga, and E. Scornet, "Mean decrease accuracy for random forests: inconsistency, and a practical solution via the Sobol-MDA," Biometrika, vol. 109, no. 4, pp. 881–900, 2022.
[26] C. Röver, D. Rindskopf, and T. Friede, "How trace plots help interpret meta-analysis results," Res. Synth. Methods, vol. 15, no. 3, pp. 413–429, 2024.
[27] J. A. Bortoloti, R. E. Bruns, J. C. De Andrade, and R. K. Vieira, "Split-plot design optimization for trace determination of lead by anodic stripping voltammetry in a homogeneous ternary solvent system," Chemom. Intell. Lab. Syst., vol. 70, no. 2, pp. 113–121, 2004.
[28] M. Friendly, "The generalized ridge trace plot: Visualizing bias and precision," J. Comput. Graph. Stat., vol. 22, no. 1, pp. 50–68, 2013.
[29] G. Xu et al., "Advances in emission control of diesel vehicles in China," J. Environ. Sci., vol. 123, pp. 15–29, Jan. 2023.
[30] S. Agrawal, M. Guevara, and S. P. Verma, "Tectonic discrimination of basic and ultrabasic volcanic rocks through log-transformed ratios of immobile trace elements," Int. Geol. Rev., vol. 50, no. 12, pp. 1057–1079, 2008.
[31] B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, A. Iost, and R. Wilcox, "Estimating the parameters of a generalized lambda distribution," Comput. Stat. Data Anal., vol. 51, no. 6, pp. 2813–2835, 2007.
[32] B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, and A. Iost, "Application of the generalized lambda distributions in a statistical process control methodology," J. Process Control, vol. 16, no. 10, pp. 1087–1098, Dec. 2006.
[33] S. Verhagen, B. Li, and P. J. G. Teunissen, "Ps-LAMBDA: Ambiguity success rate evaluation software for interferometric applications," Comput. Geosci., vol. 54, pp. 361–376, 2013.
[34] H. Hasna, N. Amalita, D. Permana, and A. Salma, "Random Forest Implementation for Air Pollution Standard Index Classification in DKI Jakarta 2022," vol. 2, pp. 226–233, 2024.
[35] R. Yu, Y. Yang, L. Yang, G. Han, and O. A. Move, "RAQ–A random forest approach for predicting air quality in urban sensing systems," Sensors (Switzerland), vol. 16, no. 1, 2016.
[36] T. Madan, S. Sagar, and D. Virmani, "Air Quality Prediction using Machine Learning Algorithms-A Review," Proc. - IEEE 2020 2nd Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2020, no. December 2020, pp. 140–145, 2020.
[37] B. Han, L. Wang, Z. Deng, Y. Shi, and J. Yu, "Source emission and attribution of a large airport in Central China," Sci. Total Environ., vol. 829, p. 154519, Jul. 2022.
[38] L. Li et al., "Air quality changes during the COVID-19 lockdown over the Yangtze River Delta Region: An insight into the impact of human activity pattern changes on air pollution variation," Sci. Total Environ., vol. 732, p. 139282, Aug. 2020.
[39] R. Horsley et al., "Lattice determination of Sigma-Lambda mixing," Phys. Rev. D - Part. Fields, Gravit. Cosmol., vol. 91, no. 7, pp. 1–40, 2015.
[40] D. Mazinanian, A. Ketkar, N. Tsantalis, and D. Dig, "Understanding the use of lambda expressions in Java," Proc. ACM Program. Lang., vol. 1, no. OOPSLA, 2017.