Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

Air Quality Index Classification: Feature Selection for Improved Accuracy with Multinomial Logistic Regression

Rizky Caesar Irjayana*[1], Abdul Fadlil[2], Rusydi Umar[3]

[1]Master Program of Informatics, Ahmad Dahlan University, Indonesia
[2]Department of Electrical Engineering, Ahmad Dahlan University, Indonesia
[3]Department of Information System, Ahmad Dahlan University, Indonesia

Email: [1]caesar.sy23@gmail.com

Received : Jul 21, 2025; Revised : Sep 20, 2025; Accepted : Sep 23, 2025; Published : Oct 16, 2025

Abstract

Air pollution is a major public health concern, creating the need for accurate and interpretable Air Quality Index (AQI) classification models. This study aims to classify AQI into three categories—Good, Moderate, and Unhealthy—using Multinomial Logistic Regression (MLR) with feature selection. The dataset, obtained from public monitoring stations in Jakarta between 2021 and 2024, initially contained 4,620 daily records. After cleaning and outlier removal, 3,586 valid samples remained, from which 900 balanced records (300 per class) were selected for modeling. Key features included PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2, which were standardized using Max Normalization to ensure uniform feature scaling. The classification process applied k-fold cross-validation (k = 2–5), and performance was assessed using accuracy and Macro F1-score. Results show that including $PM_{2.5}$ improves performance by about 10%, with the best outcome at k = 5 (accuracy = 91.67%, Macro F1 = 91.45%). These findings confirm $PM_{2.5}$ as a decisive feature for AQI prediction and demonstrate that MLR provides a lightweight, transparent, and computationally efficient solution. Beyond environmental health, the contribution of this work lies in advancing data-driven decision support systems in Informatics, particularly for real-time monitoring and policy applications.

Keywords : *Air quality index, Classification, Data mining, K-fold cross validation, Multinomial logistic regression.*

1. INTRODUCTION

Air pollution in a given location is closely associated with various adverse health effects. The standard reference used to measure pollutant concentrations is known as the Air Quality Index (AQI), or ISPU in Indonesian [1]. Air pollution is a complex mixture of gases and particles originating from anthropogenic activities (such as transportation and industrial processes) as well as natural sources (such as windblown dust, sea salt spray, and emissions) and the chemical composition of pollutants varies depending on geographical location and time [2]. In line with global economic development, rapid growth has frequently been accompanied by worsening air pollution, which in turn generates substantial health burdens. Empirical evidence shows that economic growth is closely linked to environmental degradation, and the resulting pollution significantly increases healthcare expenditures, highlighting a persistent trade-off between development gains and public health costs [3]. In the early stages of economic expansion, this pattern is further reinforced by a surge in energy demand, largely driven by fossil fuels such as coal and oil, which intensify pollutant emissions and place additional strain on environmental and health systems [4]. Air pollution exposure, both short- and long-term, has direct adverse impacts on public health and is causally linked to increased mortality risks, even at pollutant levels below national standards [5]. Ambient air pollution is now recognized as one of the leading global risk factors for premature death and years of life lost, surpassing the impact of HIV/AIDS, parasitic

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

diseases, infectious illnesses, violence, and even cigarette smoking [6]. According to the World Health Organization (WHO), air pollution poses a serious environmental threat to human health and contributes to climate change. It is estimated that around 7 million deaths annually are attributable to air pollution, primarily due to respiratory and cardiovascular diseases [7]. Recognizing the critical role of air quality and its impact on public health, the Ministry of Home Affairs of Indonesia issued Instruction No. 2 of 2023 on Air Pollution Control in the JABODETABEK Area as an official response to address the escalating air pollution problem in the region [8]. It has been reported that air pollution not only harms public health but also poses strategic risks to national defense readiness. This is primarily due to the deterioration of soldiers' physical fitness [9].

Given the significant impact of air quality on public health and the environment, it is essential to monitor and determine AQI categories rapidly and accurately. With the continuous increase in AQI data across regions, there is a pressing need for decision-making tools capable of handling large and expanding volumes of data. Classification methods provide an effective solution in addressing this challenge. In the context of machine learning, classification generally involves two main phases: training and validation. The training phase aims to construct a predictive model using labeled input data, while the validation phase evaluates its performance in producing accurate predictions [10]. The application of classification techniques for AQI analysis not only accelerates data interpretation, but also supports the development of early warning systems and predictive models for identifying potential air pollution events in specific locations.

Several studies across different domains have demonstrated the robustness and flexibility of Multinomial Logistic Regression (MLR) in handling multiclass classification tasks, particularly when dealing with structured, high-dimensional, or domain-specific data. The classification process using the Multinomial Logistic Regression (MLR) algorithm has been previously examined in a study by [11], applied Multinomial Logistic Regression (MLR) to classify Air Quality Index (AQI) levels into eight categories using data from 145 observations containing $PM_{2.5}$ and PM_{10} values. The AQI classes were derived based on Sturges' rule, and the model achieved a classification accuracy of 83.75%. Similarly, [12] applied MLR to identify the critical parameters of the ISPU using meteorological factors such as wind direction, air temperature, humidity, and rainfall, showing that several meteorological variables significantly influenced the dominant pollutant with an overall classification accuracy of 53%. These studies highlights the effectiveness of MLR in handling multiclass classification problems for environmental monitoring, particularly air quality prediction. Building on this, a subsequent study by [13], further demonstrated the effectiveness of MLR in a different domain, clinical outcome prediction for colorectal cancer. Using high-dimensional omics and clinical data from 589 patients, MLR was used to classify patients into four distinct outcome categories. Notably, the model achieved an accuracy of 85.5%, outperforming Random Forest in terms of precision**,** recall, and F1 score. Further supporting the robustness of MLR, [14] proposed a multinomial logistic regression algorithm to classify patients with parkinsonian disorders in three classes based on FDG-PET brain imaging data. The model achieved ROC AUC values of up to 0.95, confirming its effectiveness in multiclass classification without relying on healthy controls.

Given the demonstrated effectiveness of MLR across diverse domains, this work implements the approach to assess and categorize AQI into three simplified categories: Good, Moderate, and Unhealthy. Unlike previous studies that often relied on fine-grained AQI scales or complex machine learning models, this research contributes by explicitly incorporating $PM_{2.5}$ as a critical feature and simplifying the classification scheme to enhance interpretability and enable real-time deployment in Jakarta's urban monitoring context. The classification model is developed using a set of key air pollutant features that are commonly monitored in urban environments, namely particulate matter (PM_{10} and $PM_{2.5}$), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and nitrogen dioxide (NO_2). To ensure reliable model

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

evaluation, we implement K-fold Cross-validation (K-fold CV), enabling robust assessment across multiple partitions. Additionally, the study explores feature selection by comparing two experimental scenarios: one using $PM_{2.5}$, and the other excluding it. This allows us to evaluate the contribution of $PM_{2.5}$ to overall classification performance. The final evaluation considers two key performances: accuracy and the F1 score, providing a more comprehensive measure of both overall correctness and the model's ability to balance precision and recall across the classes. The potential contribution of this research lies in demonstrating that MLR can deliver accurate, interpretable, and computationally efficient performance for AQI classification, even when constrained to fewer input features and simplified class definitions. The findings are expected to support the development of feature-aware early warning systems and real-time monitoring tools, particularly useful in urban environments and data-limited regions.

2. METHOD

This research was designed and carried out through a series of well-structured and sequential stages, with the aim of achieving the expected outcomes as formulated by the authors. Each phase of the study was developed in accordance with scientific principles, ensuring that the research process remains systematic, logical, and academically sound. To provide a clearer overview of how the study was conducted from start to finish, a visual representation of the research workflow has been prepared in Figure 1.

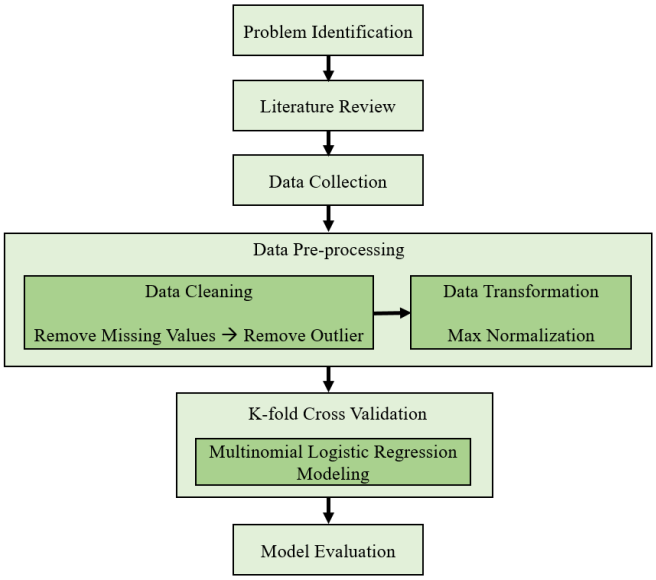

Figure 1. Research Workflow

Figure 1 illustrates the main steps involved in the research process, starting from data collection and pre-processing, through model development and evaluation. Detailed workflow of data acquisition, pre-processing, modeling, and evaluation stages. The core of the methodology lies in data pre-processing, which consists of two main procedures: data cleaning (removing missing values and outliers) and data transformation (applying max normalization to standardize feature scales). Following pre-processing, the dataset is subjected to K-fold cv combined with multinomial logistic regression modeling, the chosen classification algorithm for predicting AQI categories. Finally, model evaluation is conducted to assess the performance of the proposed approach using accuracy and Macro F1-score as the primary evaluation metrics.

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

2.1. Data Collection

The study's dataset was acquired through a public web-based source https://satudata.jakarta.go.id/open-data where the data can be accessed using the keyword tag "ISPU" in the search feature, which provides publicly accessible datasets on the AQI in the DKI Jakarta region, and was issued by the Environmental Agency of the DKI Jakarta Provincial Government. The dataset covers the period from 2021 to March 2024 and consists of a total of 4,620 raw daily records collected from monitoring stations. The regulation governing the AQI in Indonesia [15] is outlined in Regulation No. P.14/MENLHK/SETJEN/KUM.1/7/2020, issued by the Ministry of Environment and Forestry. This regulation establishes the updated AQI scale and classification thresholds, as summarized in Table 1.

Table 1. AQI Scale and Categories

AQI Range	Color Code	Category
1 - 50	Green	Good
51 – 100	Blue	Moderate
101 – 200	Yellow	Unhealthy
201 – 300	Red	Very Unhealthy
≥301	Black	Hazardous

Table 1 presents the current AQI ranges defined in the latest regulation, showing the threshold limits for each category. It replaces the previous regulation, No. 45 of 1997, by introducing two additional pollutant parameters: particulate matter ($PM_{2.5}$) and hydrocarbons (HC). These additions were made in recognition of the significant health impacts associated with both $PM_{2.5}$ and HC exposure. In line with this regulatory update, the present study aims to evaluate and compare the classification performance measured by accuracy and F1 score between two training scenarios: one that includes the $PM_{2.5}$ feature in the dataset and one that excludes it. This comparison is intended to explore whether the inclusion of $PM_{2.5}$ aligns with its recognized public health importance. Due to limitations in the available dataset, which only includes $PM_{2.5}$ as the newly added parameter and lacks hydrocarbon (HC) data, the experimental focus is restricted solely to $PM_{2.5}$. While this limitation does not invalidate the findings, it may introduce a potential bias since HC is a recognized pollutant contributing to air quality degradation and is included in several recent regulatory frameworks. The absence of HC could slightly reduce the comprehensiveness of the classification results, particularly in capturing the full spectrum of air pollution sources. Nevertheless, the selected features (PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2) remain the primary determinants of air quality and are sufficient to ensure the robustness of the classification outcomes. Moreover, the classification output in this study is limited to three AQI categories Good, Moderate, and Unhealthy as the dataset does not contain sufficient samples or thresholds to support more detailed class distinctions defined in the official AQI regulation.

2.2. Pre-processing

Pre-processing is a crucial initial step in preparing raw data for modeling or further analysis. Raw data is often unsuitable for direct use due to the presence of noise, missing values, inconsistent formats, or other irregularities that can negatively impact model performance. Effective and efficient classification outcomes largely rely on the preprocessing techniques used during data preparation [16]. To address these issues and maximize the utility of the data, several pre-processing techniques are typically applied. These include data selection, data cleaning, and data transformation, which may vary depending on the specific objectives and requirements of the study.

Data selection involves identifying and excluding irrelevant or non-informative features that do not contribute meaningfully to the model, and in some cases, may even degrade its performance.

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

Data cleaning focuses on reducing errors, inconsistencies, and irrelevant entries. This includes removing missing values such as nulls, blanks, or improperly formatted entries that may hinder the learning process. Another important aspect of cleaning is outlier detection, which helps address imbalances in data distribution. Outliers can distort model learning by concentrating the data within narrow value ranges and limiting generalization. One commonly used method for detecting outliers is the Mean-standard Deviation technique, which treats values falling outside a specified range as anomalies [17]. This method allows for the exclusion of potentially harmful data points, improving classification model robustness. The formulation of the mean-standard deviation method is presented in (1).

$$Range = mean \pm 3 \times std.dev \quad (1)$$

The transformation of data ensures that raw inputs become standardized and coherent, meaningful, and machine-readable format, allowing it to be effectively processed by data mining or machine learning algorithms. One commonly used approach within this process is normalization, which aims to reduce scale differences among features and improve computational efficiency. Significant differences in feature scales may cause features with larger numeric values to overshadow those with smaller ones. Data normalization is applied to minimize this bias and ensure that all features contribute equally to the classification process [18]. Among the various normalization techniques, Max Normalization is a widely applied method. It works by dividing each value in a feature by the maximum value of that same feature, effectively mapping all values to a standardized range between 0 and 1. This transformation ensures that all features contribute proportionally to the learning process. The mathematical formulation of Max Normalization is presented in (2).

$$x_i^* = \frac{x_i}{x_{max}} \quad (2)$$

Where x_i is the original value, x_{max} is the maximum value of the feature, and x_i^* is the result of the normalization.

This method is considered simple and computationally efficient, making it suitable for data pre-processing tasks. Its effectiveness is further enhanced when applied after prior outlier detection and removal, as the performance of Max Normalization tends to improve significantly when extreme values have been eliminated. By reducing the influence of outliers, the remaining data can be more evenly distributed within the normalized range, allowing for a more balanced representation across features.

2.3. Data Mining

Data mining constitutes a methodical inquiry into large bodies of data, whereby investigators interrogate the information from multiple vantage points in order to detect anomalies, recurring structures, and interdependencies among the data points. The outcome of this analytic labor is the formulation of substantive insights that not only inform the construction of predictive models but also provide a sound empirical foundation for strategic planning and decision-making processes. This highlights the essential role of data mining in transforming data into actionable insights through pattern recognition and predictive modelling [19]. In recent years, data mining techniques have gained widespread application across numerous fields including the tourism sector owing to their capacity to reveal intricate and hidden patterns within high-volume data. Unlike traditional statistical methods, data mining is particularly effective at detecting non-linear and multi-dimensional relationships that might otherwise go unnoticed [20]. In this study, data mining techniques were employed to uncover hidden and underlying patterns within air quality datasets, enabling a more systematic and evidence-based classification of air quality levels. Following the process, this study proceeds to the classification stage

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

using the MLR algorithm, with performance evaluation conducted through K-fold CV to ensure generalizability and robustness of the model.

The MLR model is a statistical model used to predict the outcome of a categorical dependent variable with more than two levels. It estimates the probability of each possible outcome of the dependent variable as a function of the independent variables [21]. In this context, MLR is appropriate for modeling classification tasks such as air quality levels, where the output class falls into discrete categories like Good, Moderate, and Unhealthy. Mathematically, the MLR model estimates the probability of a given class i for an observation k based on a set of predictor variables X_j using the following equation (3).

$$P_k(Y = i|X) = \frac{e^{a_i + \sum_{j=1}^{J} \beta_{ij} X_j}}{\sum_{c=1}^{C} e^{a_c + \sum_{j=1}^{J} \beta_{cj} X_j}} \quad (3)$$

Where $P_k(Y=i|X)$ is the probability that observation k belongs to class i, α_i is the intercept term for class i, β_{ij} represents the coefficient for feature j in clas i, X_j denotes the value of the *j-th* predictor, and C is the total number of classes.

During the model evaluation stage, K-fold Cross Validation offers an effective strategy for assessing predictive performance across different data partitioning scenarios. In this method, the dataset is systematically divided into training and testing subsets, where each partition or *fold* is used once as the validation set while the remaining folds serve as the training set. This approach helps address variations in accuracy that often arise due to differing data splits or sequence order during training. By applying K-fold CV, the dataset is evenly distributed across k iterations, ensuring that each subset is utilized as the testing set exactly once [22]. Example of K-fold CV splitting process illustrating training and testing sets is shown in Figure 2.

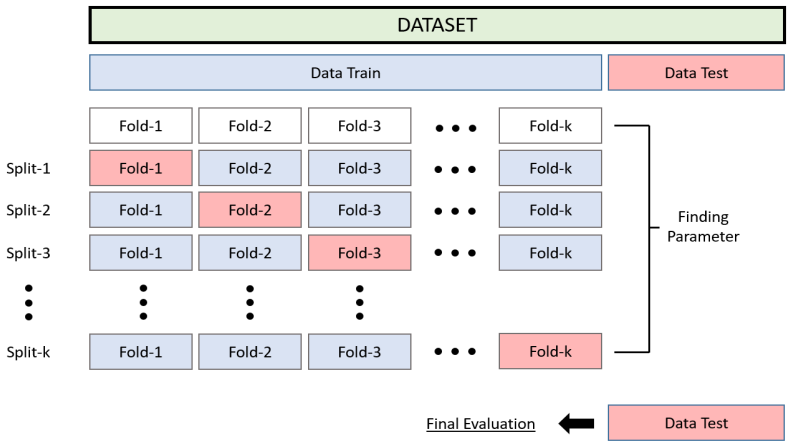

Figure 2. Illustration of K-fold CV

Figure 2 illustrates the general process of K-fold cv iteration, the number of iterations corresponds to the number of folds specified at the beginning of the process. For example, in the first iteration, the first fold acts as the test set while the remaining folds constitute the training set. In the second iteration, the second fold is used for testing, and so on this cycle continues until all folds have been used for validation. In many studies, common choices for k are 5 or 10, as these values offer a reasonable trade-off between bias, variance, and computational cost. Using a smaller number of folds (e.g., $k = 2$ or $k = 3$) can result in unstable models due to the limited size of the training data in each iteration, leading to higher variance and less generalizable outcomes. On the other hand, larger values (e.g., $k = 10$) tend to provide more stable results because each training set is larger, but this comes at the cost of increased

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

computation time, as more iterations are required. In this study, 5-fold Cross Validation is adopted as a balanced choice. It offers sufficient training data per fold while avoiding excessive computational burden, and is generally capable of producing evaluation results that are comparable in quality to those from higher fold values, making it both efficient and reliable.

2.4. Model Evaluation

Class balancing is a technique that aims to adjust the distribution of imbalanced data samples across different class labels before they are used to train a predictive model. Class imbalance frequently occurs in real-world datasets, where the majority class significantly outnumbers the minority class [23]. To address this issue, stratified selection and other balancing strategies are applied to improve classification accuracy across all classes and to reduce bias caused by imbalance. Since training on skewed data can result in unfair or distorted predictions, adopting class balancing ensures that each class is more equally represented, allowing all instances to contribute fairly to the learning process [24], [25]. To overcome the problem of class imbalance, several methods can be applied, one of the most common being undersampling. This technique reduces the number of majority class samples to achieve a more balanced class distribution. Essentially, undersampling is the process of reducing the number of instances in the majority class [26]. Compared to other balancing methods, undersampling provides notable advantages: it avoids the increase in dataset size that often occurs in oversampling, thereby reducing the risk of overfitting, and it significantly improves computational efficiency by shortening training time [27]. The approach is utilized to redistribute the dataset proportionally across the three target classes Good, Moderate, and Unhealthy to ensure that the classification model can learn more effectively from each category.

In this study, the performance of the model is evaluated using two key metrics: accuracy, as defined in Equation (4) [28], and the Macro F1-score, as defined in Equation (5). These indicators provide a comprehensive assessment of classification quality. Since the dataset employed in this research is balanced, the Macro F1-score is particularly well-suited, as it captures performance equally across all classes without being biased toward any specific class frequency [29].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Macro\ F1 = \frac{1}{N}\sum_{i=1}^{N} F1_i \quad (5)$$

Accuracy is a metric used to measure the percentage of correct predictions out of the total number of instances. In contrast, the F1-score, as presented in Equation (8), represents the harmonic mean of precision, as defined in Equation (6), and recall, as defined in Equation (7), thereby providing a single comprehensive measure that balances both metrics in evaluating classification performance.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = \frac{2.Precision.Recall}{Precision+Recall} \quad (8)$$

While precision reflects the proportion of correctly predicted positive cases among all predicted positives [30], and recall indicates the proportion of correctly identified positives among all actual positives [31], the F1-score captures both aspects in a single value. To evaluate multiclass classification tasks more comprehensively especially when class distributions are balanced the Macro F1-score is

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

commonly used. It calculates the F1-score for each class independently and then takes their unweighted average, providing an equal emphasis on all classes regardless of their sample size.

3. RESULT

3.1. Data

The initial raw dataset contained 4,620 records, but a substantial number of them included missing values, inconsistent formats, and irrelevant features that were not suitable for this research. These issues could significantly affect the model's evaluation performance; therefore, appropriate preprocessing measures were necessary to ensure data quality and integrity. The first step in addressing these issues was feature selection, aimed at identifying the most relevant attributes to be used in the modeling process. Description of selected features used for AQI classification is presented in Table 2.

Table 2. Feature Selection and Description

Feature	Data Type	Used
periode_data	int	No
tanggal	str	No
pm_10	int	Yes
pm_2,5	int	Yes
co	Int	Yes
o3	int	Yes
max	int	No
critical	str	No
categori	str	Yes
lokasi_spku	str	No

Table 2 presents the description of all features contained in the raw dataset to be processed in this study. The original data were stored in multiple .csv files, separated by month. These files were consolidated into a single .xlsx file using Microsoft Excel, based on the selected features listed in Table 2: PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2. These features were chosen not only for their direct relevance to air quality assessment but also because they represent the primary factors influencing air quality. Other features beyond these were considered irrelevant or redundant, and their inclusion could potentially degrade the performance of the classification model.

The second step was data cleaning, which included the processes of removing missing values and eliminating outliers. Records with missing values defined as those lacking entries in one or more selected features were excluded from the dataset. As a result, 879 records were removed during this stage. Examples of data categorized as missing values are shown in Table 3.

Table 3. Example of Missing Value Data

pm_10	pm_2,5	so2	co	o3	no2
57		21	23	36	16
15		16	5	20	---
28		19	11	28	---
67	85	26	N/A	7	5
57	74	25	N/A	8	5

Table 3 above shows several samples categorized as missing values, where a data row contains an empty entry or even an incorrect format that should not be present in the corresponding feature, and thus is intended to be removed. Outliers were treated as anomalous entries in the dataset. To identify

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

them, the mean-standard deviation method was applied, which calculates the upper and lower thresholds for each feature to define the acceptable data range. These calculated ranges are shown in Table 4.

Table 4. Upper and Lower Outlier Boundaries

The Upper Bound	Feature	The Lower Bound
98,806	pm_10	7,280
146,71	pm_2,5	5,898
78,819	so2	-0,828
31,800	co	-6,343
84,368	o3	-21,839
57,292	no2	-14,753

Table 4 shows the upper and lower thresholds for each feature. Accordingly, any feature row with a value outside the defined upper and lower limits is directly considered an outlier. Based on this method, 155 records were identified as outliers and subsequently eliminated. After completing the data cleaning process, a total of 3,586 valid records remained. The distribution of instances across the target classes was as follows: 304 records for the *Good* class, 2,857 records for *Moderate*, and 425 records for *Unhealthy*.

The final step in the data pre-processing stage involved data transformation, specifically through the application of Max Normalization. This method scales all feature values to a standardized range between 0 and 1 by dividing each value by the maximum value of its respective feature. The results of the normalization process are presented in Table 5.

Table 5. Example of Max Normalization

Feature	Before Max Normalization	After Max Normalization
pm_10	21	0,2234
pm_2,5	33	0,2260
so2	50	0,6494
co	4	0,1290
o3	16	0,1905
no2	3	0,0526

Table 5 above refers to an example of applying max normalization to the values of a single data row based on the highest value of each respective feature. From the total of 3,586 cleaned records, a subset of 900 records was selected to ensure balanced class representation, consisting of 300 instances each for the *Good*, *Moderate*, and *Unhealthy* categories. This selection was made by applying undersampling to prevent issues caused by class imbalance, which can lead to biased model training, reduced predictive performance for minority classes, and an overall decrease in classification accuracy. The final dataset of 900 records was then converted into a .csv file format to facilitate ease of implementation in subsequent modeling procedures.

3.2. Classification and Model Evaluation

To facilitate a clearer understanding of this study, the testing scenarios are organized into several experiments, as presented in Table 6.

Table 6 summarizes the entire sequence of experimental scenarios conducted in this study, including the number of k-folds used and the features incorporated.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

Table 6. Research Scenario

Experiment	K-fold CV	Feature Selection
1	2	PM_{10}, $PM_{2.5}$, SO_2, CO, O_3 and NO_2
2	3	PM_{10}, $PM_{2.5}$, SO_2, CO, O_3 and NO_2
3	4	PM_{10}, $PM_{2.5}$, SO_2, CO, O_3 and NO_2
4	5	PM_{10}, $PM_{2.5}$, SO_2, CO, O_3 and NO_2
5	2	PM_{10}, SO_2, CO, O_3 and NO_2
6	3	PM_{10}, SO_2, CO, O_3 and NO_2
7	4	PM_{10}, SO_2, CO, O_3 and NO_2
8	5	PM_{10}, SO_2, CO, O_3 and NO_2

The model evaluation process was carried out using Python-based libraries from Scikit-learn to test all experimental scenarios. The main source code used to execute all scenarios presented in Table 6 can be seen in Figure 3.

```
18  # 2. Model
19  model = LogisticRegression(multi_class='multinomial', solver='lbfgs', max_iter=1000)
20
21  # 3. Evaluation k-fold cv k = 2, 3, 4, 5
22  for k in [2, 3, 4, 5]:
23      kf = KFold(n_splits=k, shuffle=True, random_state=42)
24
25      y_true_all, y_pred_all = [], []
26      acc_scores = []
27      f1_macro_scores = []
28
29      for train_idx, test_idx in kf.split(X):
30          X_train, X_test = X.iloc[train_idx], X.iloc[test_idx]
31          y_train, y_test = y[train_idx], y[test_idx]
32
33          model.fit(X_train, y_train)
34          y_pred = model.predict(X_test)
35
36          y_true_all.extend(y_test)
37          y_pred_all.extend(y_pred)
38
39          acc_scores.append(accuracy_score(y_test, y_pred))
40          f1_macro_scores.append(f1_score(y_test, y_pred, average='macro'))
```

Figure 3. Source Code (python) for Model Evaluation

Figure 3 above illustrates the entire process, where the classification using MLR is implemented on line 19 of the source code, followed by k-fold cross-validation modeling on line 22 using folds of 2, 3, 4, and 5. The evaluation results are then obtained through the calculation of accuracy and macro F1-score, presented on lines 39 and 40, respectively. The complete results from all experimental iterations are summarized in Table 7.

Table 7. Experimental Result

Experiment	Accuracy %	Precision %	Recall %	Macro F1-score %
1	89,22	89,47	89,57	88,86
2	90,78	91,18	90,91	90,54
3	91,44	91,67	91,63	91,19
4	91,67	91,97	91,89	91,45
5	79,78	80,15	80,60	78,98
6	80,78	80,63	81,18	80,12
7	81,11	80,84	81,54	80,41
8	81,67	81,24	81,91	81,04

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

Table 7 presents the overall results of the experiments conducted. Experiments 1 through 4 were performed using the $PM_{2.5}$ feature, while experiments 5 through 8 were conducted without it. These two major scenarios yield their best representations in experiments 4 and 8, which achieved the highest accuracy and macro F1-score in their respective groups. To summarize and facilitate the interpretation of the experimental results, a graphical visualization of the evaluation metrics comparison is presented in Figure 4.

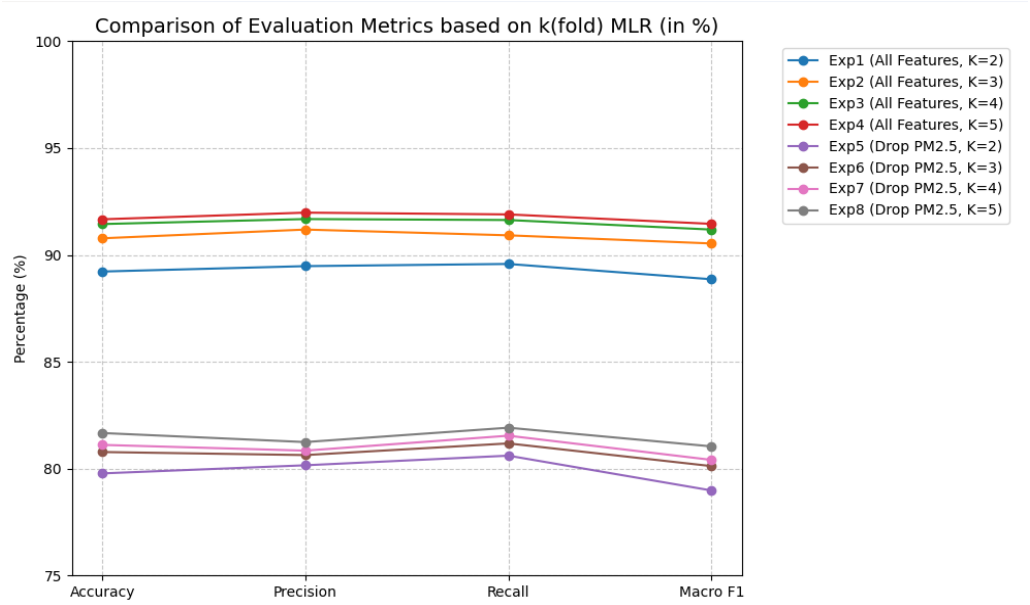

Figure 4. Comparison of Evaluation Matrics

Figure 4 illustrates a contrasting comparison between the two main scenarios in this study, namely with and without the inclusion of the $PM_{2.5}$ feature. The upper part represents all experiments involving $PM_{2.5}$, while the lower part shows those conducted without it. The results clearly demonstrate that the inclusion of $PM_{2.5}$ significantly influences the final outcomes, improving accuracy by approximately 10 percent. This comparison is evident from the representative best-performing experiments under each scenario, namely Experiment 4 and Experiment 8. The confusion matrix comparison between these two experiments is illustrated in Figure 5.

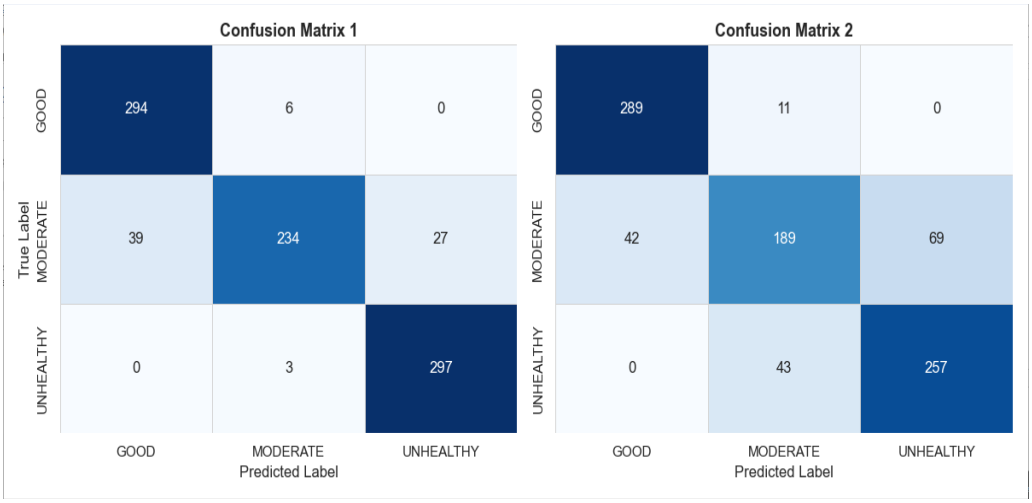

Figure 5. Comparison Between Confusion Matrix Experiment 4 (left) and 8 (right)

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

Figure 5 presents a comparison between the confusion matrices of Experiment 4 (left) and Experiment 8 (right). The results indicate that Experiment 4, which includes the $PM_{2.5}$ feature, demonstrates a more balanced classification across categories, achieving higher accuracy overall. In contrast, Experiment 8, which excludes $PM_{2.5}$, shows a noticeable decline in classification performance, particularly in distinguishing the Moderate and Unhealthy classes.

4. DISCUSSION

This study highlights the classification performance by comparing accuracy and macro F1-scores across experimental scenarios of $PM_{2.5}$ feature selection using Multinomial Logistic Regression (MLR). The analysis demonstrates that including the $PM_{2.5}$ feature significantly improves classification results, confirming its importance in predicting AQI categories. Furthermore, this research was motivated by the need to examine the relationship between the latest AQI regulations in Indonesia and whether the classification outcomes are consistent with current policy, particularly regarding the role of $PM_{2.5}$ as a critical parameter in air quality assessment.

The performance results obtained using the K-fold cross-validation technique with varying values of k (folds) are presented in Table 6 in the previous section. The research scenarios were divided into two major scenarios: Experiments 1 to 4, which included the PM2.5 feature, and Experiments 5 to 8, which excluded the PM2.5 feature. The overall results of all experiments are summarized in Table 7. The best performance was achieved in Experiment 4 with k = 5, yielding an accuracy of 91.67% and a macro F1-score of 91.45%, representing the scenario that included the PM2.5 feature. In contrast, for the experiments without the PM2.5 feature, the best performance was obtained in Experiment 8, with an accuracy of 81.67% and a macro F1-score of 81.04%.

These findings demonstrate that the choice of k directly influences the final performance, with k = 5 producing the most optimal results across both major scenarios of this study. However, Peryanto et al. (2020) [20] reported that the best performance was achieved with k = 3 despite also testing k = 5. This suggests that the optimal number of folds may vary depending on the classification algorithm and methodology applied. Nevertheless, using k = 5 generally provides a reliable balance between bias and variance, and is therefore sufficient to determine the most appropriate fold configuration in this study.

In the context of algorithmic comparison, when compared to the previous study by Irjayana et al. (2025) [1], the Naïve Bayes algorithm achieved an accuracy of 93%, which is slightly higher than the result obtained in this study, with a difference of approximately 3%. Nevertheless, this indicates that MLR remains a reliable method for achieving good performance in AQI classification tasks.

In terms of feature selection, this study provides a significant contribution to previous research employing the MLR algorithm. Pratiwi et al. (2024) [10] utilized meteorological variables (wind direction, temperature, humidity, rainfall, and wind speed) to identify critical ISPU parameters, but the classification accuracy achieved was only 53%. Meanwhile, Ali et al. (2022) [9] reported an accuracy of 83.75% by using limited pollutant features, namely PM_{10} and $PM_{2.5}$, to classify air quality index levels. These findings suggest that expanding the set of relevant features can enhance predictive performance. Accordingly, this study incorporates PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2, resulting in a classification accuracy of 91.67%. Thus, the contribution of this study lies in demonstrating the importance of representative feature selection for air quality assessment while reinforcing the effectiveness of MLR in environmental monitoring.

In addition to confirming the relevance of regulations related to $PM_{2.5}$ inclusion, the findings of this study also have a number of practical implications. The simplicity and interpretability of the MLR model make it suitable for application in real-time monitoring systems, particularly in the form of public dashboards or mobile applications that require quick but transparent decisions. MLR provides

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

coefficients that can be directly understood by policymakers and environmental agencies, thereby facilitating evidence-based decision-making.

Nevertheless, the findings of this study are still limited to data from DKI Jakarta, so generalizing the results to other regions with different pollution characteristics and seasonal variations needs to be further tested, for example through time series analysis and temporal validation. From a public health perspective, the approximately 10% increase in accuracy by including $PM_{2.5}$ is significant, because classification errors can lead to inaccurate estimates of exposure risk for the population. Therefore, future research directions should focus on expanding the scope of datasets, exploring broader feature sets, and refining computational strategies to ensure that the proposed approach remains robust and adaptable in diverse real-world contexts.

5. CONCLUSION

Classification results reveal that Multinomial Logistic Regression (MLR), when combined with feature selection and k-fold cross-validation, provides an effective and computationally efficient approach for Air Quality Index (AQI) classification. The best performance was achieved at k = 5, with an accuracy of 91.67% and a macro F1-score of 91.45% when $PM_{2.5}$ was included, compared to only 81.67% accuracy and 81.04% macro F1-score without it. This nearly 10% improvement highlights the decisive role of $PM_{2.5}$ both as a regulatory indicator and a predictive feature. Beyond environmental health, these findings also emphasize the value of MLR as a lightweight and interpretable model for Informatics, particularly in decision support systems that require real-time and transparent outputs. From a computer science perspective, the contribution of this work lies in advancing computational modeling strategies that balance predictive accuracy with interpretability, offering a scalable method applicable to data-driven policy and public applications. Future research should expand this framework by applying larger and more diverse datasets across multiple regions, integrating broader feature sets such as hydrocarbons and meteorological variables, and extending the classification scheme to include Very Unhealthy and Hazardous categories. Furthermore, it is essential to compare MLR with other algorithms, including traditional methods such as Support Vector Machine (SVM) and Random Forest, as well as modern approaches like ensemble methods and deep learning. Such comparisons would improve robustness and generalizability, while also strengthening the role of computational informatics in building adaptive and accurate decision support systems for real-time air quality monitoring.

CONFLICT OF INTEREST

The authors wish to affirm that no conflicts of interest are present among the authors or between the authors and the subject of the research reported in this paper.

REFERENCES

[1] R. C. Irjayana, A. Fadlil, and R. Umar, "Pengaruh Seleksi Fitur Terhadap Akurasi Klasifikasi Indeks Standar Pencemar Udara Menggunakan Naïve Bayes", *Insect (Informatics and Security)*, vol. 11, no. 1, pp. 67–78, 2025, doi: 10.33506/insect.v11i1.4303.

[2] I. S. Mudway, F. J. Kelly, and S. T. Holgate, "Oxidative stress in air pollution research," *Free Radical Biology and Medicine,* vol. 151, pp. 2-6, 2020, doi: 10.1016/j.freeradbiomed.2020.04.031.

[3] F. Chen and Z. Chen, "Cost of Economic Growth: Air Pollution and Health Expenditure," *Science of the Total Environment*, vol. 755, Part 1, p. 142543, 2021, doi: 10.1016/j.scitotenv.2020.142543.

[4] J. Y. Xie, D. H. Suh, and S. -K. Joo, "A Dynamic Analysis of Air Pollution: Implications of Economic Growth and Renewable Energy Consumption", *International Journal of Environmental Research and Public Health*, vol. 18, no. 18, p. 9906, 2021, doi: 10.3390/ijerph18189906.

[5] Y. Wei, Y. Wang, X. Wu, Q. Di, L. Shi, P. Koutrakis, A. Zanobetti, F. Dominici, and J. D.

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

Schwartz, "Causal Effects of Air Pollution on Mortality Rate in Massachusetts," *American Journal of Epidemiology,* vol. 189, no. 11, pp. 1316-1323, 2020, doi: 10.1093/aje/kwaa098.

[6] J. Lelieveld, A. Pozzer, U. Pöschl, M. Fnais, A. Haines, and T. Münzel, "Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective," *Cardiovascular Research*, vol. 116, no. 11, pp. 1910–1917, 2020, doi: 10.1093/cvr/cvaa025.

[7] WHO, "Nearly 50 million people sign up call for clean air action for better health," [Online]. Available: https://www.who.int/news/item/17-03-2025-nearly-50-million-people-sign-up-call-for-clean-air-action-for-better-health, Accessed: Jul. 10, 2025.

[8] Kementerian Dalam Negeri Republik Indonesia, "Inmendagri Tahun 2023," [Online]. Available: https://ditjenbinaadwil.kemendagri.go.id/halaman/detail/inmendagri-tahun-2023, Accessed: Jul. 10, 2025.

[9] A. A. Anandari, A. F. Wadjdi and G. Harsono, "Dampak Polusi Udara terhadap Kesehatan dan Kesiapan Pertahanan Negara di Provinsi DKI Jakarta," *Journal on Education*, vol. 6, no. 2, pp. 10868-10884, 2024, doi: 10.31004/joe.v6i2.4880.

[10] Y. Shino, Y. Durachman, and N. Sutisna, "Implementation of Data Mining with Naive Bayes Algorithm for Eligibility Classification of Basic Food Aid Recipients," *International Journal of Cyber and IT Service Management (IJCITSM),* vol. 2, no. 2, pp. 154-162, 2022, doi: 10.34306/ijcitsm.v2i2.114.

[11] A. N. Ali, G. Nassreddine, and J. Younis, "Air Quality prediction using Multinomial Logistic Regression," *Journal of Computer Science and Technology Studies*, vol. 4, no. 2, pp. 71-78, 2022, doi: 10.32996/jcsts.2022.4.2.9.

[12] R. Pratiwi, R. Widyasari, and M. Fathonni, "Analisis Regresi Logistik Multinomial Dalam Estimasi Parameter Kritis Indeks Standar Pencemar Udara," *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 5, no. 1, pp. 499-513, 2024, doi: 10.46306/lb.v5i1.588.

[13] C. H. Feng, M. L. Disis, C. Cheng, and L. Zhang, "Multimetric feature selection for analyzing multicategory outcomes of colorectal cancer: random forest and multinomial logistic regression models," *Laboratory Investigation*, vol. 102, no. 3, pp. 236-244, 2022, doi: 10.1038/s41374-021-00662-x.

[14] E. Štokelj, T. Rus, J. Jamšek, M. Trošt, and U. Simončič, "Multinomial logistic regression algorithm for the classification of patients with parkinsonisms," *EJNMMI Research*, vol. 15, no. 24, 2025, doi: 10.1186/s13550-025-01210-0.

[15] Database Peraturan, "Indeks Standar Pencemar Udara," [Online]. Available: https://peraturan.bpk.go.id/Details/163466/permen-lhk-no-14-tahun-2020, Accessed: Jul. 10 2025.

[16] E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. M. F. S. El-Salhi, "The Effect of Preprocessing Techniques, Applied to Numeric Features, on Classification Algorithms' Performance," *Data*, vol. 6, no. 2, p. 11, 2021, doi: 10.3390/data6020011.

[17] H. A. Ahmed, P. J. M. Ali, A. K. Faeq and S. M. Abdullah, "An Investigation on Disparity Responds of Machine Learning Algorithms to Data Normalization Method," *ARO-The Scientific Journal Of Koya University*, vol. 10, no. 2, pp. 29-37, 2022, doi: 10.14500/aro.10970.

[18] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, Part B, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.

[19] R. R. Asaad and R. M. Abdulhakim, "The Concept of Data Mining and Knowledge Extraction Techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 17-20, 2021, doi: 10.48161/qaj.v1n2a43.

[20] Sunardi, A. Fadlil and N. M. P. Kusuma, "Implementasi Data Mining dengan Algoritma Naïve Bayes untuk Profiling Korban Penipuan Online di Indonesia," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, pp. 1562-1572, 2022, doi: 10.30865/mib.v6i3.3999.

[21] G. Shiran, R. Imaninasab, and R. Khayamim, "Crash Severity Analysis of Highways Based on Multinomial Logistic Regression Model, Decision Tree Techniques, and Artificial Neural Network: A Modeling Comparison," *Sustainability*, vol. 13, no. 10, 2021, doi: 10.3390/su13105670.

Vol. 6, No. 5, October 2025, Page. 3265-3279
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5155

[22] A. Peryanto, A. Yudhana, and R. Umar, "Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation," *Journal of Applied Informatics and Computing (JAIC)*, vol. 4, no. 1, pp. 45-51, 2020, doi: 10.30871/jaic.v4i1.2017.
[23] L. Sha, M. Raković, A. Das, D. Gašević, and G. Chen, "Leveraging Class Balancing Techniques to Alleviate Algorithmic Bias for Predictive Tasks in Education," *IEEE Transactions on Learning Technologies*, vol. 15, no. 4, pp. 481-492, 2022, doi: 10.1109/TLT.2022.3196278.
[24] M. Mahmood, F. M. Jasem, A. A. Mukhlif, and B. Al-Khateeb, "Classifying cuneiform symbols using machine learning algorithms with unigram features on a balanced dataset," *Journal of Intelligent Systems*, vol. 32, no. 1, p. 20230087, 2023, doi: 10.1515/jisys-2023-0087.
[25] S. George and V. Srividhya, "Performance Evaluation of Sentiment Analysis on Balanced and Imbalanced Dataset Using Ensemble Approach," *Indian Journal of Science and Technology*. vol.15, no. 17, pp. 790-797, 2022, doi: 10.17485/IJST/v15i17.2339.
[26] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," *2020 11th International Conference on Information and Communication Systems (ICICS)*, 2020, pp. 243-248, doi: 10.1109/ICICS49469.2020.239556.
[27] D. Devi, S. K. Biswas, and B. Purkayastha, "A review on solution to class imbalance problem: Undersampling approaches," *2020 International Conference on Computational Performance Evaluation (ComPE)*, 2020, pp. 626-631, doi: 10.1109/ComPE49325.2020.9200087.
[28] I. Riadi, A. Fadlil, and B. A. Prabowo, "MAC Address Classification in Privacy Issue Using Gaussian Naïve Bayes," *JUITA: Jurnal Informatika*, vol. 12, no. 2, pp. 235-242, 2024, doi: 10.30595/juita.v12i2.22571.
[29] J. Opitz and S. Burst, "Macro F1 and Macro F1," arXiv, 2021, doi: 10.48550/arXiv.1911.03347.
[30] B. Wang, "A Parallel Implementation of Computing Mean Average Precision," *arXiv preprint arXiv:2206.09504*, p. 2, 2022, doi: 10.48550/arXiv.2206.09504.
[31] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 6, pp. 1–13, 2020. doi: 10.1186/s12864-019-6413-7.