Comparison of Ordinal Logistic Regression, XGBoost, and TabNet Models for Air Quality Classification in Jakarta City

Theophilus Nathan Tanudjaja[1*], *Syarifah* Diana Permai[1*], and *Diana*[2]

[1]Statistics Department, School of Computer Science, Bina Nusantara University Jakarta, Indonesia 11480
[2]Computer Science Department, School of Computer Science, Bina Nusantara University Jakarta, Indonesia 11480

Abstract. This study aims to compare the performance of statistical, machine learning, and deep learning models in solving the air quality classification problem in Jakarta province over the period from 2018 to 2024. The dataset used in this research was obtained from the 'Satu Data Indonesia' website, specifically the daily air quality data ISPU (Indeks Standar Pencemar Udara). Specifically, the models evaluated in this study include Ordinal Logistic Regression, XGBoost, and TabNet, and their performance is assessed using several evaluation metrics, such as precision, recall, F1-score, and accuracy. The research process is divided into several stages, including hyperparameter tuning, to enhance model performance. The dataset consists of daily air quality data, with variables such as PM_{10}, SO_2, CO, O_3, and NO_2. The results indicate that the XGBoost model outperforms the other models.

1 Introduction

Air pollution has become one of the most concerning global environmental issues. Over the past two decades, Indonesia has experienced significant changes in air quality, where the country has shifted from one of the countries with the cleanest air quality to one of the twenty countries with the highest pollution levels in the world[1]. In Jakarta, air pollution threatens the health of more than 10.5 million residents, with the highest recorded annual PM2.5 concentration among all cities in Indonesia[2]. Several human activities, such as cigarette smoke, industrial activities, transportation, land or forest burning, contribute significantly to the high levels of air pollution[3]. The impact of air pollution on health is substantial, with approximately 6.7 million deaths in 2019, nearly 85% of which were caused by non-communicable diseases[4]. Reducing particulate matter (PM) 2.5 levels could prevent 124,729 deaths annually in Europe [5]. Therefore, accurate measurement and monitoring of air quality are crucial to identifying pollution sources, understanding the spread patterns, and formulating effective policies to reduce its negative impact.

* Corresponding author: theophilus.tanudjaja@binus.ac.id, syarifah.permai@binus.ac.id

This study aims to contribute to the prevention of deaths due to air pollution, focusing on the analysis of the factors affecting air quality in the DKI Jakarta province and identifying the best model based on evaluation metrics. Based on previous research, several variables relevant to air quality, such as PM_{10}, $PM_{2.5}$, Sulfur Dioxide (SO_2), Carbon Monoxide (CO), Ozone (O_3), and Nitrogen Dioxide (NO_2), have been used [6-7]. These variables are also established by the World Health Organization as pollutant standards for air quality indicators[8].

2 Ordinal Logistic Regression

Ordinal Logistic Regression is a statistical model used to analyze the relationship between a dependent variable and independent variables, where the dependent variable has more than two categories and is ordered on a scale. In ordinal regres-sion, a cumulative logit model is used to generate predictions. This model com-pares cumulative odds, which is the probability that the response category Y is less than or equal to the jth response category given p predictor variables ex-pressed in vector X, i.e., $P(Y \leq j \mid X)$, with the probability that Y exceeds the jth response category, i.e., $P(Y > j \mid X)$. The cumulative odds, $P(Y \leq j \mid X)$, are defined as follows[9]:

$$P(Y \leq j|X) = \frac{exp\ (\theta_j + \sum_{k=1}^{p} \beta_k x_k)}{1 + exp(\theta_j + \sum_{k=1}^{p} \beta_k x_k)} \quad (1)$$

The formulation of the cumulative logit model:

$$logit\ P(Y \leq j|X) = log[\frac{P(Y \leq j\ |X)}{P(Y > j|X)}] \quad (2)$$

The general cumulative logit model when equation (1) is substituted into equation (2):

$$logit\ P(Y \leq j \,|\, X) = \alpha_m + \sum_{k=1}^{p} \beta_k X_k \quad (3)$$

Where:
$P(Y \leq j|X)$ = Cumulative probability
α_m = The intercept parameter for category m
β_k = The regression coefficient for predictor k
X_k = The k^{th} independent variable

3 XGBoost

XGBoost starts with a simple model, usually the mean of the target values in the dataset as the initial prediction. Next, the residuals are calculated, which repre-sent the difference between the actual values and the predictions for each data point in the training set, indicating the error of the initial model. The algorithm then builds a small decision tree (weak learner) to predict these residuals with the goal of minimizing the loss function, such as mean squared error in regression problems. The prediction from this tree is added to the previous prediction to update the model's approximation. New residuals are calculated based on the updated prediction, and a new tree is built to predict these latest residuals. This process is repeated for a specified number of iterations, known as boosting rounds. Finally, the results from all the trees, which are scores or logits, are com-bined and use the softmax function to generate predictions for each class.[10]

4 TabNet

TabNet provides a high-performance and interpretable deep learning architecture for tabular data. TabNet uses a sequential attention mechanism to determine which features to select at each step, allowing for high interpretability and effi-cient training. This attention mechanism enables the model to focus on the most relevant features for prediction, while minimizing resource usage, thereby im-proving training efficiency and accuracy[11]. TabNet is built by several main components:

4.1 Feature Transformer

This component is responsible for transforming input features into more informa-tive representations. The Feature Transformer consists of two parts: one that is shared across decision steps and used in all steps, and one that is decision step dependent and specific to each step. Both parts consist of a series of fully con-nected layers, batch normalization, and non-linear activation functions (RELU/GLU). Each block in the Feature Transformer also includes residual con-nections, which help maintain information from the previous layers and prevent the loss of information during the transformation process.

4.2 Attentive Transformer

This component uses sequential attention to select the most important features at each decision-making step. The Attentive Transformer is implemented using sparse matrix multiplication operations, making it efficient for tabular data. This mechanism operates through fully connected layers, batch normalization, and an attention mask formed using prior scales (accumulated information from previ-ous steps). The Sparsemax activation function is used to generate a sparse atten-tion distribution, which emphasizes only a small set of the most relevant features. This allows the model to make decisions more efficiently and interpretably.

5 Data and Methodology

5.1 Data

This study uses secondary data obtained from the Satu Data Indonesia website [12], covering the Jakarta region from 2018 to 2024, with a total of 7,076 data points. However, after preprocessing and removal of null data, 6,373 data points remain. The variables used are taken from the Air Pollution Standard Index (ISPU) Table 1.

Table 1. Research Variables

Heading level	Variable Type	Scale	Description
Category	Response	Ordinal	Air quality levels consisting of Good, Moderate, Unhealthy, and Very Unhealthy

PM_{10}	Predictor	Ratio	Amount of Particulate Matter with particle diameter less than 10 micrometers in micrograms per cubic meter
SO_2	Predictor	Ratio	Amount of Sulfur Dioxide in micrograms per cubic meter
CO	Predictor	Ratio	Amount of Carbon Monoxide in milligrams per cubic meter
NO_2	Predictor	Ratio	Amount of Nitrogen Dioxide in micrograms per cubic meter
O_3	Predictor	Ratio	Amount of Ozone in micrograms per cubic meter

5.2 Research Methods

Several research steps carried out:
- Data Preprocessing: In this stage, missing values are identified and the corresponding data is removed from the dataset.
- Exploratory Data Analysis (EDA): This is performed on the dataset to understand the data distribution and the correlation between variables.
- Data Splitting: The data is divided into two parts: 70% for training and 30% for testing.
- Building the Ordinal Logistic Regression Model
- Building and Hyperparameter Tuning for the XGBoost Model
- Building and Hyperparameter Tuning for the Tabular Network (TabNet) Model
- Comparing the Evaluation Results of the Ordinal Logistic Regression, XGBoost, and TabNet Models to Determine the Best Model

A Confusion Matrix is a method used to evaluate the performance of a classifica-tion model by comparing its predictions with the actual classes to determine whether the objects are classified correctly or incorrectly. This matrix provides information about the actual and predicted classification values. There are four terms that describe the classification outcomes using a Confusion Matrix: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). True Positive (TP) refers to cases where the model correctly predicts the positive class; True Negative (TN) refers to cases where the model correctly predicts the negative class; False Positive (FP) occurs when the model incorrectly predicts the positive class for an actually negative case; and False Negative (FN) occurs when the model incorrectly predicts the negative class for an actually positive case [13].

Model evaluation can be analyzed using various metrics such as precision, recall, F1-score and accuracy. Each of these metrics can be calculated using the follow-ing formulas:

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1\ Score = 2\ \frac{Precision*Recall}{Precision+Recall} \tag{6}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

6 Result and Discussion

Before modeling the dataset, data exploration needs to be performed first. The exploration is carried out for both numerical and categorical data.

 https://doi.org/10.1051/e3sconf/202568000119

6.1 Exploratory Data Analysis

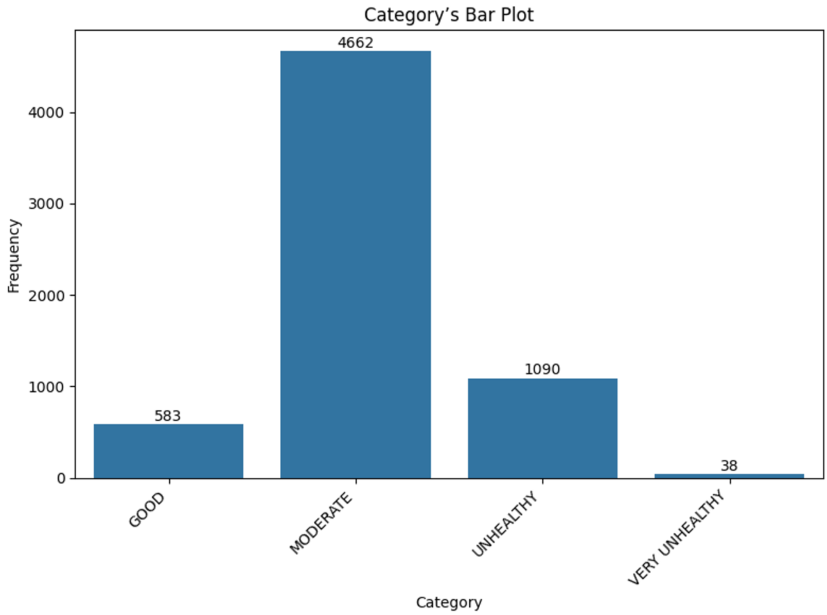

Fig. 1. Category’s Bar Plot

The data distribution shows a significant imbalance, with the 'Moderate' category dominating. This imbalance needs to be considered as it may affect the conclusions, but it can be addressed using models or analysis techniques that handle imbalances, such as weighting, resampling, or algorithms that are robust to dataset imbalance.

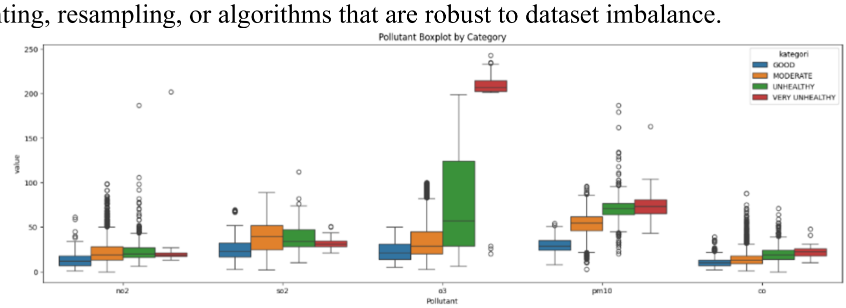

Fig. 2. Independent Variables Box Plot

The boxplot highlights the variation in pollutant levels among these categories, with O3 showing the most significant differences in distribution between categories, especially for the 'VERY UNHEALTHY' group, which shows much higher values and a larger interquartile range. Other pollutants, such as NO_2, SO_2, and PM_{10}, have relatively consistent distributions across categories, although there are still significant outliers in some of them.

6.2 Ordinal Logistic Regression

6.2.1 Model 1

The first model in this analysis will include all available independent variables as input variables to build the ordinal logistic regression model. The initial step taken is to perform a simultaneous significance test using the likelihood ratio test. This test compares the full model, which includes all independent variables, with the null model that has no predictor variables [9]

Table 2. Results of the Likelihood Ratio Test for Model 1

#Df	LogLik	Df	Chisq	Pr(>Chisq)
8	-3481.6			
3	-1906.0	5	3151.2	$< 2.2 \times 10^{-16}$

H0: $\beta_{pm10} = \beta_{so2} = \beta_{co} = \beta_{o3} = \beta_{no2} = 0$ (no independent variable affects the de-pendent variable)
H1: At least one $\beta_i \neq 0$ for i = 1, 2, ..., k (at least one independent variable affects the dependent variable)

Since the obtained p-value is < 2.2*10-16, which is smaller than the significance level α = 0.05, it can be concluded that the null hypothesis is rejected. This indi-cates that at least one independent variable has an effect on the dependent varia-ble. Next, a partial significance test will be conducted using the Wald test. This test aims to determine the significance of each regression coefficient in the mod-el, to see whether each independent variable significantly affects the dependent variable.

Table 3. Result of Wald test for Model 1

Variable	Wald_Statistic	P_value
PM_{10}	855.4789795	0
SO_2	71.4044827	0
CO	0.7753585	0.3785644
O_3	504.5348213	0
NO_2	101.9479885	0
0|1	690.7793544	0
1|2	1427.4469104	0
2|3	1385.5662377	0

H0: $\beta_i = 0$ (The variable does not significantly affect the dependent variable)
H1: $\beta_i \neq 0$ (The variable significantly affects the dependent variable)

Since the p-value for the variables PM_{10}, SO_2, O_3, and NO_2 obtained is 0, which is smaller than the significance level $\alpha = 0.05$, it can be concluded that the null hy-pothesis is rejected. This indicates that the variables PM_{10}, SO_2, O_3, and NO_2 sig-nificantly affect the dependent variable. On the other hand, the CO variable does not significantly affect the dependent variable.

6.2.2 Model 2

The second model in this analysis will use the variables as PM_{10}, SO_2, O_3, and NO_2, as the CO variable did not pass the Wald test (t-test). Therefore, in the sec-ond model, the CO variable is removed to build the ordinal logistic regression model. The initial step taken is to perform a simultaneous significance test using the likelihood ratio test [9].

Table 4. Results of the Likelihood Ratio Test for Model 2

#Df	LogLik	Df	Chisq	Pr(>Chisq)
7	-3481.6			
3	-1906.4	4	3150.4	$< 2.2 \times 10^{-16}$

H0: $\beta_{pm10} = \beta_{so2} = \beta_{o3} = \beta_{no2} = 0$ (no independent variable affects the dependent variable)
H1: At least one $\beta_i \neq 0$ for i = 1, 2, ..., k (at least one independent variable affects the dependent variable).

Since the obtained p-value is < 2.2*10-16, which is smaller than the significance level α = 0.05, it can be concluded that the null hypothesis is rejected. This indi-cates that at least one independent variable has an effect on the dependent varia-ble. Next, a partial significance test will be conducted using the Wald test. This test aims to determine the significance of each regression coefficient in the mod-el, to assess whether each independent variable significantly affects the depend-ent variable.

Table 5. Results of Wald test for Model 2

Variable	Wald_Statistic	P_value
PM_{10}	884.71544	0
SO_2	72.95453	0
O_3	563.86551	0.3785644
NO_2	109.37657	0
0|1	721.48559	0
1|2	1449.85431	0

2|3	1380.73769	0

H_0: $\beta_i = 0$ (The variable does not significantly affect the dependent variable)
H_1: $\beta_i \neq 0$ (The variable significantly

Since the p-value for all independent variables obtained is 0, which is smaller than the significance level $\alpha = 0.05$, it can be concluded that the null hypothesis is rejected. This indicates that all independent variables significantly affect the dependent variable.
The Variance Inflation Factor (VIF) test is used to detect multicollinearity among independent variables in the regression model, which could cause instability in coefficient estimates and reduce the reliability of the analysis results [15].
Based on the VIF test results, all independent variables are below the threshold of 10. This indicates that there is no significant indication of multicollinearity among the independent variables in the model. Specifically, the VIF for PM_{10} is 1.0047, for SO_2 is 1.1221, for O_3 is 1.1928, and for NO_2 is 1.1270. Next, the model must satisfy the proportional odds assumption, which can be tested using the Brant test, to ensure that the effect of the independent variables on the odds ratio is constant across all ordinal response categories [16].

Table 6. Results of Brant Test for Model 2

Test for	X2	df	Probability
Omnibus	118.72	8	0
X_train$pm10	0.29	2	0.87
X_train$so2	111.11	2	0
X_train$o3	5.57	2	0.06
X_train$no2	4.04	2	0.13

H0: The regression coefficients are proportional or parallel across all categories (the proportional odds assumption is satisfied).
H1: The regression coefficients are not proportional across all categories (the proportional odds assumption is violated).

Since the p-value from the omnibus test obtained is 0, which is smaller than the significance level $\alpha = 0.05$, it can be concluded that the null hypothesis is rejected. This indicates that the regression coefficients are not proportional across all categories. Model 2 generates cumulative probability estimates for each category, which can be used in the prediction process. The equation for the cumulative probabilities is given as follows:

$$P(Y \leq \text{“Good”}) = \frac{\exp\begin{bmatrix} 5.58696753 + (0.12258561 * PM10) + \\ (0.02613150 * SO2) + (0.03839146 * NO2) \end{bmatrix}}{1 + \exp\begin{bmatrix} 5.58696753 + (0.12258561 * PM10) + \\ (0.02613150 * SO2) + (0.03839146 * NO2) \end{bmatrix}} \quad (8)$$

$$P(Y \leq \text{“Moderate”}) = \frac{\exp\begin{bmatrix} 13.07422906 + (0.12258561 * PM10) + \\ (0.02613150 * SO2) + (0.03839146 * NO2) \end{bmatrix}}{1 + \exp\begin{bmatrix} 13.07422906 + (0.12258561 * \mathrm{PM10}) + \\ (0.02613150 * SO2) + (0.03839146 * NO2) \end{bmatrix}} \quad (9)$$

$$P(Y \leq \text{"Unhealty"}) = \frac{\exp\begin{bmatrix} 19.72409068 + (0.12258561*PM10) + \\ (0.02613150*SO2) + (0.03839146*NO2) \end{bmatrix}}{1+\exp\begin{bmatrix} 19.72409068 + (0.12258561*PM10) + \\ (0.02613150*SO2) + (0.03839146*NO2) \end{bmatrix}} \quad (10)$$

$$P(Y \leq \text{"Very Unhealty"}) = 1 \quad (11)$$

6.3 XGBoost

XGBoost fine-tuning is done by optimizing several key parameters, namely max_depth (MD), which controls the depth of the decision tree, affecting the model's complexity; learning_rate (LR), which determines the contribution of each new tree to the overall model, influencing convergence speed and learning stability; and n_estimators (NE), which controls the total number of trees built in the model, thus affecting the complexity and generalization ability of the model.

Table 7. Hyperparameter tunning for XGBoost

MD, LR, NE	Precision	Recall	F1- Score	Accuracy
25, 0.001, 50	87.61	80.16	83.5	88.55
25, 0.01, 100	87.06	79.54	82.91	88.18
10, 0.001, 100	87.06	79.24	82.77	88.02
25, 0.001, 100	87.29	81.35	83.99	88.55
10, 0.001, 100	87.06	79.24	82.77	88.02
25, 0.0001, 100	87.44	79.58	83.12	88.28
25, 0.001, 150	87.44	79.58	83.12	88.28

From Table 7, an experiment was conducted to test various XGBoost models with different parameter variations: max_depth (50, 25, 10), learning_rate (0.01, 0.001, 0.0001), and n_estimators (50, 100, 150). The goal of this experiment was to find the parameter combination that produces the best performance. The evaluation results showed that the combination of max_depth = 25, learning_rate = 0.001, and n_estimators = 100 provided the most optimal results. This choice is considered the best because the selected parameter values are in the middle of the tested value ranges. For example, max_depth = 25 is the midpoint between the options 10 and 50.

Table 8. Comparison of evaluation metrics for training and testing with XGBoost's best parameters

Dataset	Precision	Recall	F1- Score	Accuracy
Training	94.26	92.62	93.36	94.4
Testing	87.29	81.35	83.99	88.55

The results of precision, recall, F1-score, and accuracy from the training set also show that overfitting did not occur, as the difference between performance on the training and testing data is relatively small, with testing accuracy remaining high and not significantly decreasing compared to training.

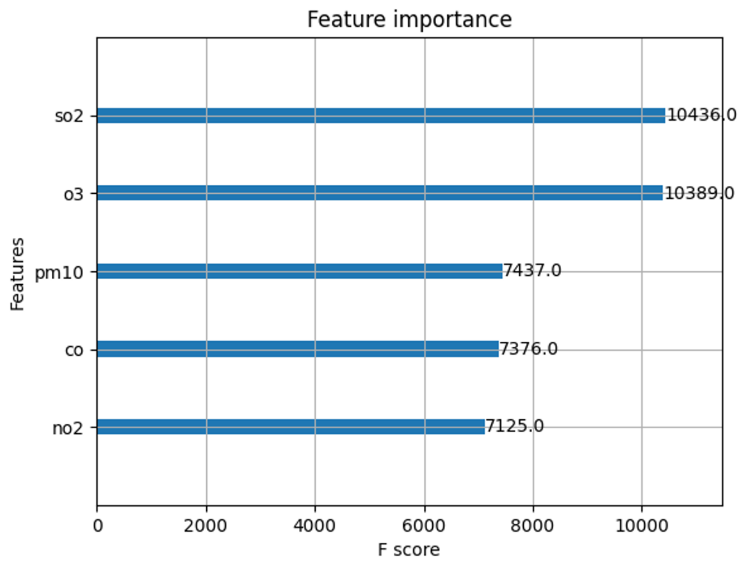

Fig. 3. Feature Importance of XGBoost

Feature importance shows that the XGBoost model heavily relies on the variables SO_2 and O_3 in making predictions, while PM_{10}, CO, and NO_2 also contribute significantly, but not as much as the first two features.

6.4 TabNet

Fine-tuning TabNet is done by optimizing several key parameters, n_d, which controls the feature dimension for decision prediction. Larger values provide more capacity to the model, with a risk of overfitting, n_a, which controls the embedding dimension for attention. Generally, setting n_a = n_b provides the best results, n_steps, which is the number of attention and feature selection steps, mask_type is the type of attention masking function, with options of 'sparsemax' or 'entmax'.

Table 9. Hyperparameter tunning for TabNet

n_d, n_a, n_steps, mask_type	Precision	Recall	F1- Score	Accuracy
6,6,5,sparsemax	71.99	80.34	75.18	86.56
8,8,3,entmax	76.39	76.56	75.69	88.02
6,6,5,entmax	79.81	76.18	77.68	88.08
8,8,5,entmax	87.58	74.57	79.88	88.39

10,10,5,entmax	84.02	75.62	79.24	88.28
10,10,5,sparsemax	85.94	77.89	81.35	87.66
8,8,5,sparsemax	81.62	72.73	76.56	87.19

From Table 9, an experiment was conducted to test various TabNet models with different parameter variations: n_d (6, 8, 10), n_a (6, 8, 10), n_steps (3, 5, 10), and mask_type (entmax, sparsemax). The goal of this experiment was to find the parameter combination that produces the best performance. The evaluation results showed that the combination of n_d and n_a = 8, n_steps = 5, and mask_type = 'entmax' provided the most optimal results. This choice is considered the best because the selected parameter values are in the middle of the tested value ranges. For example, n_steps = 5 is the midpoint between the options 3 and 10.

Table 10. Comparison of evaluation metrics for training and testing with Tabnet's best parameters

Dataset	Precision	Recall	F1- Score	Accuracy
Training	88.66	79.88	83.8	89.31
Testing	87.58	74.57	79.88	88.39

In addition, the results of precision, recall, F1-score, and accuracy from the training set also show that overfitting did not occur, as the difference between performance on the training and testing data is very small.

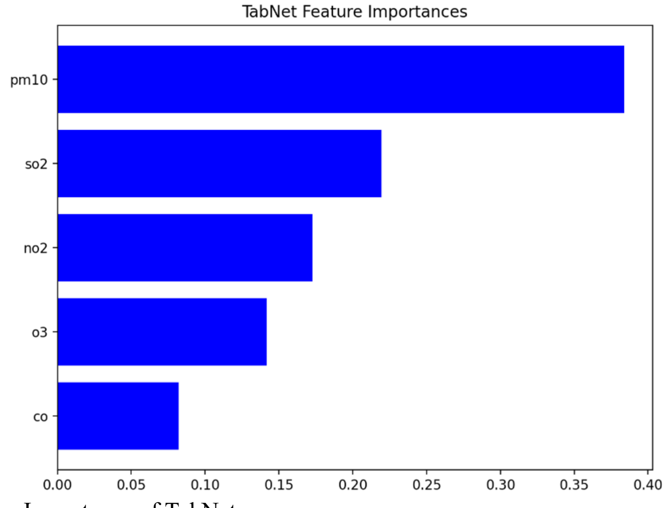

Fig. 4. Feature Importance of TabNet

The feature importance analysis results show that the TabNet model heavily relies on the PM_{10} variable in making predictions. Meanwhile, the SO_2, NO_2, and O_3 variables contribute significantly, although not as much as PM_{10}. On the other hand, the CO variable has a relatively small contribution, with a value below 10%.

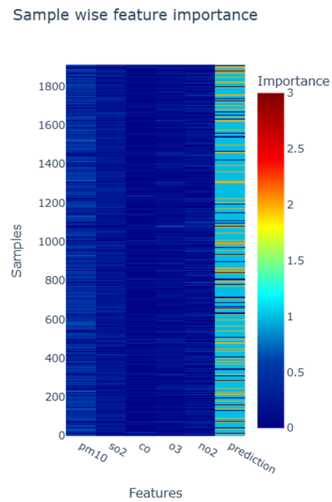

Fig. 5. Sample Wise Feature Importance

The TabNet model is capable of providing feature importance for each sample in the dataset through the attention mechanism. The colors shown in Figure 5 indicate the level of importance of each feature for the prediction result of each sample, where blue represents a low level of importance, while red or yellow indicates a high level of importance. The results in Figure 5 support the findings in Figure 4, where the CO variable does not make a significant contribution, as its importance value is close to zero in almost every sample.

6.5 Model Comparison

After conducting analysis on the Ordinal Logistic Regression, XGBoost, and TabNet models, the next step is to compare the results from all the models. Evaluation is carried out using various performance measurement metrics such as precision, recall, F1-score, and accuracy. This way, we can select the model that provides the best results.

Table 11. Comparison of evaluation metrics for All Models

Model	Precision	Recall	F1- Score	Accuracy
Ordinal Logistic Regression	64	53.66	57.68	82.22
XGBoost	87.29	81.35	83.99	88.55
TabNet	87.58	74.57	79.88	88.39

The results from Table 11 show the evaluation metrics for the four models tested. It can be concluded that XGBoost is the best model among the others, as XGBoost achieved the best results in three out of the four metrics used. The recall value is 81.35%, the F1-score reaches 83.99%, and the accuracy is 88.55%. In fact, the precision, recall, F1-score, and accuracy values for XGBoost are all above 80%, a performance that was not achieved by any of the other models.

7 Conclusion

In the Ordinal Logistic Regression model, the results of the partial significance analysis show that the variables PM_{10}, SO_2, O_3, and NO_2. significantly affect the dependent variable. In contrast, the CO variable does not significantly affect the dependent variable. Additionally, the results of Feature Importance in the XGBoost and TabNet models also show that the CO variable has a low contribution to air quality prediction, indicating that these models do not consider CO as an important feature in air quality classification. These findings have important implications for public health and environmental policy. They suggest that government agencies should reconsider current budget allocation shifting resources from CO reduction initiatives toward controlling pollutants that have a greater and statistically significant impact on air quality, such as PM_{10}, SO_2, O_3, and NO_2. Based on the comparison of the Ordinal Logistic Regression, XGBoost, and TabNet models, it can be concluded that two models perform quite well: XGBoost and TabNet. However, based on the evaluation metrics, XGBoost gives the best results. The XGBoost model achieved a precision of 87.29%, recall of 81.35%, F1-score of 83.99%, and accuracy of 88.55%.

However, this study has several limitations. First, the Ordinal Logistic Regression model was trained on ordinal data, while the other models (XGBoost and TabNet) were treated as if they were handling nominal data. This discrepancy in data processing may lead to an unfair comparison, as ordinal relationships in the dependent variable are treated differently from nominal data. Future research could explore models specifically designed for nominal data, such as Logistic Regression, for potentially better performance. Second, the study did not utilize K-fold cross-validation, which is a widely accepted technique for model validation. K-fold cross-validation limits the robustness of model evaluation, as the models were not tested for their generalization ability on multiple subsets of the dataset.

Data Availability Statement
The data used in this research were obtained from Satu Data Indonesia website: https://data.go.id/instantion/provinsi-dki-jakarta?q=ispu

Credit Authorship Contribution Statement
Theophilus Nathan Tanudjaja: Data Curation, Visualization, Formal analysis, Literature Review. **Syarifah Diana Permai**: Conceptualization, Supervision, Methodology, Writing - Review & Editing. **Diana**: Supervision, Writing - Review & Editing.

References

1. Michael Greenstone, Qing (Claire) Fan (2019) Kualitas Udara Indonesia yang Memburuk dan Dampaknya terhadap Harapan Hidup
2. Syuhada G, Akbar A, Hardiawan D, et al (2023) Impacts of Air Pollution on Health and Cost of Illness in Jakarta, Indonesia. Int J Environ Res Public Health 20:. https://doi.org/10.3390/ijerph20042916
3. Lauri Myllyvirta, Isabella Suarez, Erika Uusivuori, Hubert Thieriot (2020) Pencemaran Udara Lintas Batas di provinsi Jakarta, Banten, dan Jawa Barat
4. World Health Organization (2024) Health consequences of air pollution. https://www.who.int/news/item/25-06-2024-what-are-health-consequences-of-air-pollution-on-populations. Accessed 2 Jan 2025

5. Khomenko S, Cirach M, Pereira-Barboza E, et al (2021) Premature mortality due to air pollution in European cities: a health impact assessment. Lancet Planet Health 5:e121–e134. https://doi.org/10.1016/S2542-5196(20)30272-2
6. Damayanti TV, Handriyono RE (2022) Monitoring Kualitas Udara Ambien Melalui Stasiun Pemantau Kualitas Udara Wonorejo, Kebonsari Dan Tandes Kota Surabaya. Environmental Engineering Journal ITATS 2:11–18. https://doi.org/10.31284/j.envitats.2022.v2i1.2897
7. Apriani NF, Salampessy JEB, Kusumadewi S, et al (2024) Classification and Forecasting Air Pollution Using Naive Bayes and Prophet: A Use Case of Air Quality Index in Jakarta. In: 2024 International Conference on Computer, Control, Informatics and its Applications (IC3INA). IEEE, pp 279–284
8. Geneva: World Health Organization (2021) WHO global air quality guidelines. Particulate matter (PM25 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide
9. Pentury T, Aulele SN, Wattimena R (2016) ANALISIS REGRESI LOGISTIK ORDINAL. BAREKENG: Jurnal Ilmu Matematika dan Terapan 10:55–60. https://doi.org/10.30598/barekengvol10iss1pp55-60
10. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794. https://doi.org/10.1145/2939672.2939785
11. Arik SO, Pfister T (2019) TabNet: Attentive Interpretable Tabular Learning. https://doi.org/10.48550/arXiv.1908.07442
12. Provinsi DKI Jakarta. https://data.go.id/instantion/provinsi-dki-jakarta?q=ispu. Accessed 9 Oct 2025
13. Valero-Carreras D, Alcaraz J, Landete M (2023) Comparing two SVM models through different metrics based on the confusion matrix. Comput Oper Res 152:106131. https://doi.org/10.1016/j.cor.2022.106131
14. Farhadpour S, Warner TA, Maxwell AE (2024) Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. Remote Sens (Basel) 16:. https://doi.org/10.3390/rs16030533
15. Mayawi M, Nurhayati N, Talib T, et al (2023) Ordinal Logistic Regression Analysis of Factors that Affecting the Blood Sugar Levels Diabetes Mellitus Patients. Pattimura International Journal of Mathematics (PIJMath) 2:33–42. https://doi.org/10.30598/pijmathvol2iss1pp33-42
16. Brant R (1990) Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. Biometrics 46:1171. https://doi.org/10.2307/2532457