Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

Performance Analysis of Air Pollution Classification Prediction Map with Decision Tree and ANN

Rizky Fauzi Ramadhani, Sri Suryani Prasetiyowati*, Yuliant Sibaroni

Informatics, School of Computing, Telkom University, Bandung, Indonesia
Email: [1]rizkyfauzira@student.telkomuniversity.ac.id, [2,*]srisuryani@telkomuniversity.ac.id, [3]yuliant@telkomuniversity.ac.id
Submitted: **15/08/2022**; Accepted: **29/08/2022**; Published: **30/08/2022**

Abstrak− DKI Jakarta merupakan sebuah kota di Indonesia yang memiliki padat penduduk yang tinggi yang harus diperhatikan kondisi kesehatannya. Kualitas udara yang baik memberikan manfaat positif untuk menunjang kesehatan masyarakat agar dapat lebih produktif dalam beraktifitas dan menciptakan udara yang segar dan sehat. Studi ini menggunakan *Machine Learning* untuk meklasifikasikan udara berdasarkan atribut tertentu. Kemudian, pengembangan model prediksi klasifikasi berdasarkan data waktu dirancang untuk menghasilkan peta prediksi penyebaran polusi udara di area DKI Jakarta 3 tahun kedepan. Metode yang diterapkan adalah *Decision Tree* dan *Artificial Neural Network*. Hasilnya, model *Decision Tree* dan *Aritficial Neural Network* menunjukkan akurasi yang sangat baik untuk prediksi tahun 2024 hingga 2026. Di tahun 2024 model *Decision Tree* dan *Artificial Neural Network* mendapatkan akurasi 98% dan 94%. Di tahun 2025 model *Decision Tree* dan *Artificial Neural Network* mendapatkan akurasi 99% dan 93%. Di tahun 2026 model *Decision Tree* dan *Artificial Neural Network* mendapatkan akurasi 94% dan 93% yang dimana dapat dilihat model *Decision Tree* lebih unggul dibandingkan *Artificial Neural Network* dengan perbedaan 1 - 6 %.

Keywords: Polusi Udara; Peta Prediksi; *Decision Tree*; *Artificial Neural Network*; Jakarta

Abstract−Jakarta is a city in Indonesia that has a high population density that must pay attention to its health condition. Good air quality provides positive benefits to support public health so that they can be more productive at work and create fresh and healthy air. This study uses Machine Learning to classify air based on certain attributes. Then, the development of a prediction model based on time data is designed to produce a predictive map of air pollution in Jakarta area for the next 3 years. The methods applied are Decision Tree and Artificial Neural Networks. As a result, the Decision Tree and Artificial Neural Network models show very good accuracy for predictions from 2024 to 2026. The Decision Tree and Artificial Neural Network models get an accuracy of 98% and 94%. In 2025 the Decision Tree and Artificial Neural Network models get 99% and 93% accuracy. In 2026 the Decision Tree and Artificial Neural Network models get an accuracy of 94% and 93% which can be seen from the Decision Tree model which is superior to the Artificial Neural Network with a difference of 1 - 6%.

Keywords: Air Pollution; Prediction Map; Decision Tree; Artificial Neural Network; Jakarta

1. INTRODUCTION

DKI Jakarta has air pollution problems which is quite serious. Air pollution is damage to air quality. Poor air quality can endanger the health of humans, animals and plants, interfere with aesthetics and comfort, or properties [1]. Until now, air pollution is a problem that is quite difficult to overcome. This is because the main source of air pollutants in Jakarta comes from transportation, household activities, and transportation industries with relatively low chimneys. Based on the IQAir website which looks at the quality index. The air quality in various big cities represents that the DKI Jakarta air quality index can be categorized become a city with healthy – unhealthy air. The consequences of air pollution are very detrimental to humans because it can cause various problems with lung, heart, respiratory and even lung diseases cause death. Based on air pollution problems that occur and based on information on the *IQAir*.com website which represents the air quality index of DKI Jakarta is categorized into cities with good air healthy – unhealthy it is necessary to classify the air pollution map in the city of Jakarta so that it can be easy to know areas - areas that are prone to air pollution and harmful to humans. Currently, there are many machine learning methods used to classify pollution air based on the identification of the air quality index, for example by study [2] using the Genetic Algorithm (GA) for selecting inputs and designing high-level model architecture multi-layer perceptron to forecast hourly nitrogen dioxide concentrations in urban traffic busy one. Decision Tree and Naïve Bayes algorithms are also used to identify Air Quality Index (AQI), in study [3] which showed a high accuracy result in the use of the algorithm Decision Tree J48 is 91.9978% which is higher than the accuracy of using the Naïve Bayes method with an accuracy of 86.663%.

Prediction of air pollution in Brazil was also carried out in the study [4]. By implementing Artificial Neural Network (ANN) method is more effective to solve non-linear problems. The same case with study [5] which uses Artificial Neural Network and Gradient Boost methods to predict air pollution traffic is like a moving camera to collect air traffic in real-time. From these 4 studies, it has been proven that machine learning can be an option for obtain information in the form of predictions, classifications, and others. In this research will be carried out prediction using Decision Tree algorithm and Artificial Neural Network (ANN). Because algorithm Decision Trees can be used to examine the relationship between air pollution concentrations and variables affected [6]. As a comparison algorithm, the Artificial Neural Network (ANN) algorithm is efficient to solve non-linear and collection problems, combining multiple outputs models [4].

In research [6], which aims to predict environmental variations in ozone (O3) and PM 2.5 concentrations using the Decision Tree and LUR methods. The results obtained indicate that the Decision Tree is higher than

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

LUR. Where Decision Tree produces an accuracy of 0.85 and 0.76, while the LUR accuracy is 0.71 and 0.58 for ozone and PM 2.5. However, this study only predicts the concentration of ozone (O3) and PM2.5. The ANN algorithm is also implemented in research [7], which is proposed for air pollution prediction research by implementing multi-layer perceptron and ensemble type using PM 10, PM 2.5, nitrogen, and ozone attributes. The development of the ANN model is an important advance because it not only acts as an alternative to traditional modelling approaches, but also as a complementary model that can be used to improve predictive performance. The ANN and Random Forest in research [8] were also implemented to compare the two algorithms to produce more accurate predictions. The results showed that the ANN algorithm is more than the RF model with an accuracy of 0.9639 and 0.7028, respectively. In the study [9] compared three algorithms, namely Decision Tree, ANN, and Multiple Regression to predict the composition of the carcass network of young goats. The results of these studies can predict the composition of the carcass well. The highest accuracy lies in the ANN algorithm 0.711 – 0.893, while DT with an accuracy of 78.9 – 85.2% and Multiple Regression with an accuracy of 51%. ANN and SVM algorithms are also implemented in research [10] to predict high seas footprints for water resource management. The results of the study stated that the ANN algorithm can predict better than the SVM algorithm, where the ANN accuracy result is 0.9807 and the SVM accuracy result is 0.9588. While in research [11] compares the ANN and Decision Tree in implementing stock price predictions for digital game content. With the highest average ANN accuracy of 15.31% and Decision Tree with 14.06% accuracy, the study states that ANN is a more stable method in predicting stock prices in the volatile post-crisis stock market. A study [12] proposes that the use of ANN to predict is still unsatisfactory and hopes that the use of ANN can be further developed. However, ANN can be used to predict soil temperature well with an accuracy of 0.9994.

This research predicts the classification of the city of Jakarta based on 5 air monitoring stations consisting of Central Jakarta, South Jakarta, West Jakarta, East Jakarta, and North Jakarta using the Decision Tree and Artificial Neural Network (ANN). The data to be processed in the study contains a combination of air condition data from 2016 to 2021. Feature selection in this study is also carried out by comparing the performance of the model with attributes 3, 4, and 5 years previously. The method used also uses the ANOVA method. In addition, the visualization of predictions in the form of maps will be made using the classification results that are inputted into the ArcMap application. The purpose of this research is to compare the performance between the two implemented algorithms, namely Decision Tree and Artificial Neural Network (ANN) in obtaining a prediction model for the years 2024 to 2026 in DKI Jakarta. It is hoped that the output of this prediction map can also help parties within the scope of DKI Jakarta to prevent air pollution in polluted areas.

2. RESEARCH METHODOLOGY

2.1 Research Stages

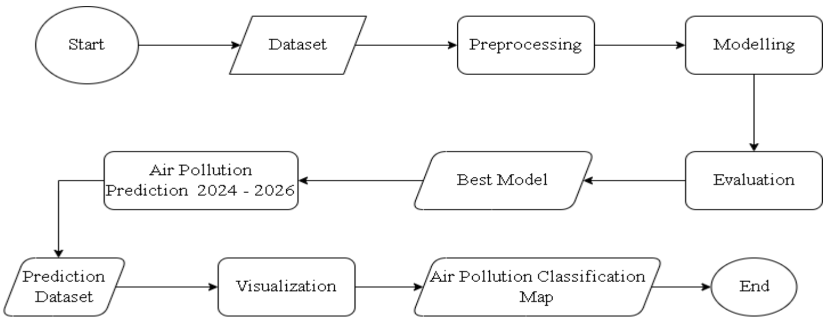

Figure 1. System Design Flowchart

Figure 1 shows how the flow of the system built in this study. The first step is to prepare the dataset. The second stage preprocessing the dataset, which consists of encoding, scaling, and sampling to change the raw data to make it easier to understand and ready to be processed by the system at the next stage. The third stage is modelling the dataset to show which model with the highest accuracy. The fourth stage is evaluation to evaluate the model. The fifth stage is selecting the best model to be processed to the next stage. The sixth stage is making air pollution predictions from 2024 to 2026. The seventh stage is prediction that will be used to the next stage. The eight stage is visualization in this stage the prediction will be visualize. The last stage is a classification map the value of the classification process and the processes carried out.

2.2. Dataset

Table 1. Dataset

Notation	Features Name
X1	PM 10
X2	SO2

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

Notation	Features Name
X3	CO
X4	O3
X5	NO2
X6	MAX
X7	CRITICAL
Y	Kategori

data used in this study contains data obtained from open government secondarySecondary data is data obtained indirectly from an object of research [13]. dataset obtained from the online data website for Jakarta is shown by https://data.jakarta.go.id/ from 2016 to 2021. dataset has 8 attributes including class attributes.

2.3. Preprocessing

Preprocessing is an important step before classification. the data is processed first before entering the modeling stage. In table 1, column Y is still a category. Column y will be changed to category class label. Labeling is done based on the available categories in the dataset which is then labeled using an encoder label based on the reference [14]. Then the dataset is completed by filling in the blank data with the average of the column. The category refers to [1] where the consideration of the level of air quality on human health is based on the determination of the ISPU. ISPU is determined based on 5 main pollutants, namely: Carbon Monoxide (CO), Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), Surface Ozone (O3), and Dust Particles (PM10).

Table 2. ISPU Category

ISPU	Air Pollution Level
0 - 50	Good
51 - 100	Moderate
101 – 199	Unhealthy
200 – 299	Very Unhealthy
300 – 500	Harmful Dangerous

After the above steps, the existing data is transformed first. One technique of data transformation is min-max normalization. Where this technique is useful for getting new values with a range of minimum and maximum values that exist in a feature [15].formula min-max normalization itself can be seen in the **Error! Reference source not found.** below

$$v' = \frac{v - min}{\max - min} \quad (1)$$

Where *v*' is a new value, *v* is the old value, min is the minimum value, and max is the max value.

2.4. Decision Tree

Decision Tree is a tree where each branch displays a choice among a number of alternative choices, and each leaf displays the chosen decision [16]. Decision trees are often used to obtain information for the purpose of making a decision. Decision Tree is used to study the classification and prediction of patterns from the data and describe the relations and variables of the x attribute and the target variable y in the form of a tree [17]. The main key of the Decision Tree is discrimination and classification which predicts the value obtained from the category of the variable to be predicted which is known as the attribute [18]. For the calculation of the entropy value, it can be seen in the formula **Error! Reference source not found.**

$$Entropy(S) = \sum_{i=1}^{n} -pi * \log_2 pi \quad (2)$$

Where S is a set of cases, A is a feature, N is a number of partitions S and Pi is a proportion of Si in S

2.5. Artificial Neural Network

Artificial Neural Network (ANN) is an algorithm that has the ability to study a set of input-output the subject, then predicts the output for the new sample set at high speed and with a reasonable degree of accuracy [19]. In the ANN classification using multi-layer perceptrons, multi-layer perceptrons are often used because of their flexibility and ability to adapt various non-linear with a high degree of accuracy [20].
ANN is applied to all input for each neuron. The values obtained are added up (dot product) to obtain the output Oj, this process uses the formula **Error! Reference source not found.**

$$O_j = f\left(\sum_{i=1}^{n^{[0]}} w_{ij} X_i + b_j\right) \quad (3)$$

Where N[0] is a number of features for the input pattern, and B is a bias that allows shifting of the linear combinations, first layer of output is calculated by doing the dot product of the output of the previous hidden layer

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

and the actual hidden layer weight added to the actual bias. The result is the input of the activation function f(.) as shown in the formula **Error! Reference source not found.**

$$\hat{Y}_k = f\left(\sum_{j=1}^{n^{[l-1]}} W_{jk} O_j + b_k\right) \quad (4)$$

Where N [l-1] is a number of neuron layers hidden in the previous layer.

2.6. Modelling

In this modeling stage, the aim is to create a classification training data model for prediction of air pollution distribution classes. Making several models aims to get the best model. By doing a combination of sharing data and attributes. For example, for the 2017 case, the attributes used are 2016 case data. The inputted data follows the previous few years of the model. The process of dividing the data into training data and test data is carried out in this stage. The percentage of training data and test data using a ratio of 80:20

Table 3. Data Model

Model	Label	Data Attribute
3 Previous Years	3a	2016, 2017, 2018
	3b	2017, 2018, 2019
	3c	2018, 2019, 2020
	3d	2019, 2020, 2021
4 Previous Years	4a	2016, 2017, 2018, 2019
	4b	2017, 2018, 2019, 2020
	4c	2018, 2019, 2020, 2021
5 Previous Years	5a	2016, 2017, 2018, 2019, 2020
	5b	2017, 2018, 2019, 2020, 2021

technique feature selection is used to select which features do have an influence on the target column to be predicted. In addition to this, the selection of relevant features will shorten the time complexity and provide good accuracy for the system [21].best feature selection is done using the SelectKBest. SelectKBest is used to get the best features which can be seen in Table 4. Sample Combination Sample below. After selecting the features, each model will be compared. The method called in the SelectKBest is ANOVA. Where in research [22] the ANOVA method works very well. The workings of the ANOVA method itself is to calculate the value between several features, where the resulting high value indicates the importance of the feature when used.

Table 4. Example of a Combination Sample

No	Number of Attributes	Combination
1	4	X4,x5,x6,x7
2	4	x1,x2,x5,x6
3	4	x1,x4,x5,x8
4	4	x5,x6,x7,x8
…	…	…
n	6	x1,x2,x3,x5,x6,x7

2.7. Evaluation

Evaluation of several models was carried out to determine the performance of the model. Evaluation is done by implementing the k-fold cross validation. This technique will separate the entire data into k folds [23]. Where the data will be trained and tested in every fold. The value of k that will be used in this study is 5. The model that has the best accuracy will be chosen to predict air pollution in 2024 to 2026. The formula **Error! Reference source not found.** for the accuracy itself is:

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{P+N}} \quad (5)$$

Where TP / True Positive is a positive value that is predicted to be true.other hand, TN / True Negative is a negative value that is predicted to be true. While P (positive) and N (negative) is a list of data with positive and negative labels.

2.8. Classification

Prediction The case classification prediction for 2024 to 2026 is implemented using the best model obtained in the previous stage. Later, a comparison of the best models of the two algorithms used in this study will also be carried out.

2.9. Visualization

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

The final stage is to create a classification map from 2024 to 2026. Using 3 maps with the best model of the decision tree algorithm. Then 3 other maps with the best model of artificial neural network algorithm. Making this visualization map is done with the help of ArcMap.

3. RESULTS AND DISCUSSION

In this study, we will use a dataset that has gone through the resampling stage using SMOTE-ENN predict air pollution. The training data taking the 80% of the dataset to be trained with each machine learning method. Then the classification probability making the decision through the process, while testing data is taking the 20% to test the algorithm model that has been built before. In this study, the first scenario is comparing the accuracy of Decision Tree with several model, the second scenario is comparing the accuracy of ANN with several model, the third scenario is choosing the best model to be used in prediction map from 2024 to 2026.

3.1 Model of the previous 3 years

3 year model testing, it was done by comparing the 3a, 3b, 3c, 3d models and the combined models of the four models. The model is given input with the attribute data listed in Table 3. In addition, testing is carried out by comparing the best features from a range of 3 to 18 attributes in each model. The results obtained are the 3c model using Decision Tree to be the best model compared to other models tested. Judging from Figure 2 and Figure 3, the 3c Decision Tree produces high accuracy, which is above 98%. While in the ANN model the best model is in model 3b with an accuracy of 94%.

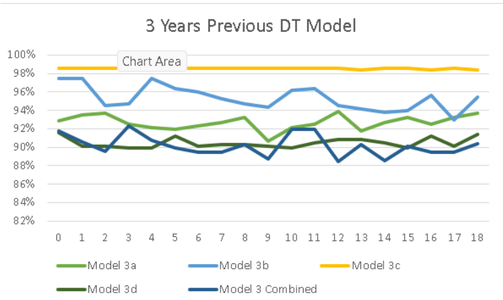

Figure 2. 3 Years Previous DT Model

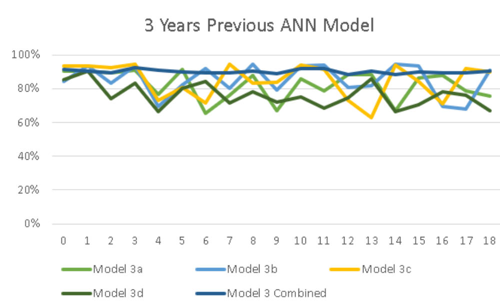

Figure 3. 3 Years Previous ANN

3.2 Model of the previous 4 years

In the previous 4 years model testing, it was done by comparing the 4a, 4b, 4c models and the combined models of the three models. The model is given input with the attribute data listed in Table 3. Tests are carried out with attributes for each model ranging from 3 attributes to 25 attributes. The results obtained are the 4b Decision Tree model being the best model for the overall model being tested. Judging from Figure 4 and Figure 5, the 4b Decision Tree produces a high accuracy of 99%, while the 4b ANN model produces an accuracy of 93%.

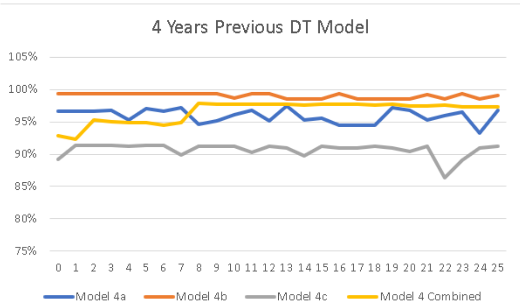

Figure 4. 4 Years Previous DT Model

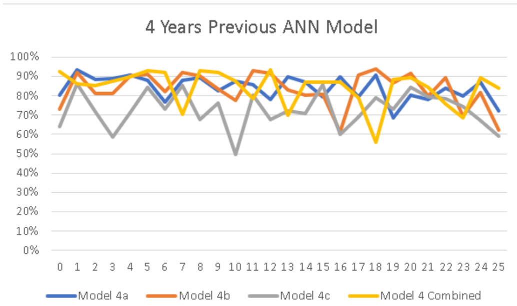

Figure 5. 4 Years Previous ANN Model

3.3 Model of the previous 5 years

In the previous 5 years model testing, it was done by comparing models 5a, 5b and the combined models of the three models. The model is given input with the attribute data listed in Table 3. The results obtained are that model 5a is the best model for the overall model being tested. Judging from Figure 6 and Figure 7, model 5a produces high accuracy, which is above 93%.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

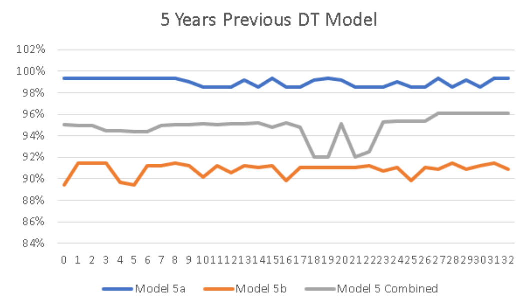

Figure 6. 5 Years Previous DT Model

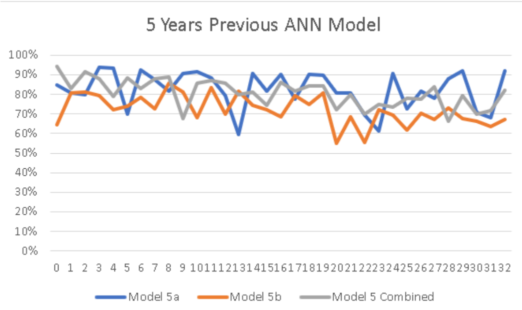

Figure 7. 5 Years Previous ANN Model

Of the three models that have been tested, starting from the 3-years, 4-years and previous 5-years models, the most influential and outgoing attribute patterns are PM10, NO2, SO2, CO, and critical.

3.4 Implementation

Models 3c, 4b, and 5a become classification prediction models for 2024, 2025, 2026. Where based on the previous steps, the three models obtained accuracy results that beat other models. In 2024 it is predicted to use the 3c model, in 2025 it is predicted to use the 4b model, and in 2026 it is predicted to use the 5a model. The results of the accuracy of the predictions obtained from the three models can be seen in the table below.

Table 5. Accuracy of 3C Model

Akurasi Model 3C Decision Tree		Akurasi Model 3C ANN	
3 Atribut	20 Atribut	3 Atribut	20 Atribut
98%	98%	93%	93%

Table 6. Accuracy of 4B Model

Akurasi Model 4B Decision Tree		Akurasi Model 4B ANN	
3 Atribut	25 Atribut	3 Atribut	25 Atribut
99%	98%	93%	60%

Table 7. Accuracy of 5A Model

Akurasi Model 5A Decision Tree		Akurasi Model 5A ANN	
8 Atribut	30 Atribut	8 Atribut	30 Atribut
99%	98%	93%	70%

Based on Table 5, Table 6, and Table 7 prediction result from 3C model with 3 attributes will be used to create a prediction map for 2024. The 4B model with 3 attributes will be used to make a prediction map for 2025. And finally the 5A model with 8 attributes for a prediction map in 2026. The prediction map is displayed through the prediction results from the best model from results by the two algorithms used, namely *Decision Tree* and ANN in the previous stage.

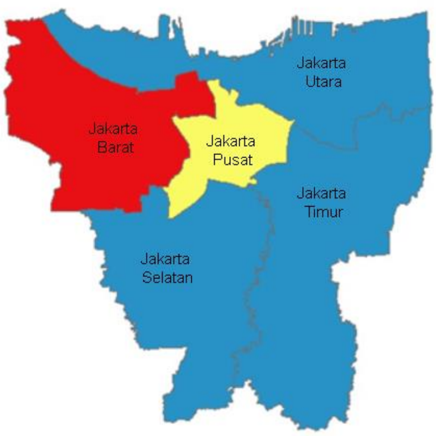

Figure 8. 2024 Prediction Map DT

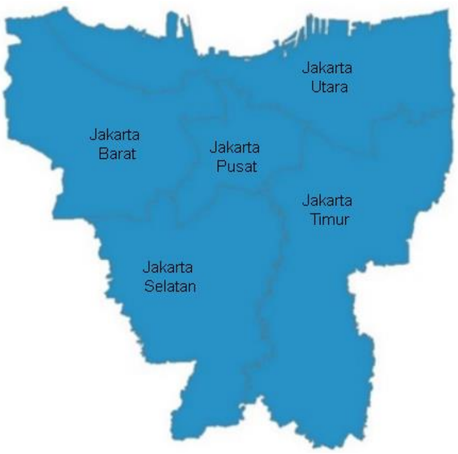

Figure 9. 2024 Prediction Map ANN

It can be seen based on the classification prediction map in 2024 that the performance of using the Decision Tree method is better than ANN with an accuracy of 98% and 94%. When we see the map, the red color means the air pollution is unhealthy at West Jakarta and yellow at Center of Jakarta means moderate then at east Jakarta,

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

North Jakarta, and South Jakarta is Good because the color is blue in Decision Tree method, and the all of Jakarta is blue which means good with ANN.

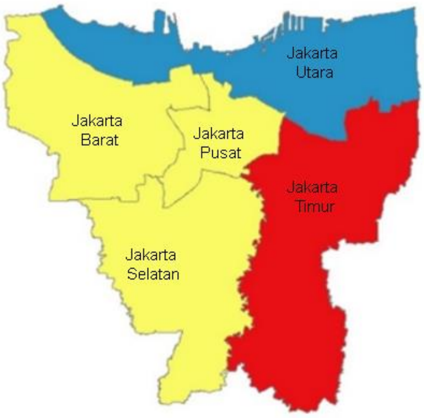

Figure 10.2025 Prediction Map DT

Figure 11. 2025 Prediction Map ANN

then in 2025 predictions the use of the Decision Tree Back method is superior with an accuracy of 99% while the ANN method with a causation of 93 %. Where we can see from the map, West Jakarta, Center of Jakarta, and South Jakarta has a color of yellow which means its moderate air pollution, the east Jakarta is red which means unhealthy then north Jakarta is blue means good in Decision Tree Algorithm. For ANN, west Jakarta and center of Jakarta has a color of yellow which means its moderate, south Jakarta and North Jakarta has a color of red means unhealthy and east Jakarta has a color of blue means good air.

Figure 12. 2026 Prediction Map DT

Figure 13.2026 Prediction Map ANN

and the prediction in 2026 has decreased accuracy but is still effective because the accuracy is above 90% and the Decision Tree method is still superior with 94% accuracy and the ANN method with 93% accuracy, and the difference in the results of the classification map shows that areas affected by air pollution are getting worse. spread to south Jakarta over time even though in 2024 the ANN method predicts all DKI Jakarta with moderate air quality, then in 2026 South Jakarta and East Jakarta will be exposed to very unhealthy air pollution, then West Jakarta and Central Jakarta have unhealthy air quality and Jakarta moderate north, then on pre The Decision Tree diction shows quite different predictions, where in 2024 predicts that North Jakarta, East Jakarta and South Jakarta will have moderate air quality, while West Jakarta will be very unhealthy and Central Jakarta will have moderate air quality. In 2026 the Decision Tree method predicts that all of DKI Jakarta will have moderate air quality with an accuracy of 93%.

4 CONCLUSION

From the research that has been done, Decision Tree algorithm works better. This is shown when predicting the 2024 to 2026 case. We can see by comparing the accuracy where the highest accurary is more than 90% for each method in this study, the accuracy of Decision Tree is 98% for 2024, 99% for 2025, and 94% for 2026. The accuracy of ANN is 94% for 2024, 93% for 2025, and 93% for 2026. The most influential attributes/features on accuracy in this study are PM10, NO2, SO2, CO, and critical. Based on field data based on airborne particles and nitrogen can cause a wide spread of air pollution. Based on the predictions made, by looking at the map of the classification results, it can be seen that air pollution is increasing over time from year to year where we can see

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 536−543
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2117

the red color in the map getting spread widely in Jakarta. For further research, predictions can be made by adding data samples using kriging interpolation in spatial so that areas outside the study can be affordable and predictable in the future to avoid increasing air pollution.

REFERENCES

[1] K. Naive *et al.*, "Klasifikasi naive bayes untuk memprediksi kualitas udara berdasarkan indeks standar pencemar udara (ispu)," 2019.
[2] H. Niska, T. Hiltunen, A. Karppinen, J. Ruuskanen, and M. Kolehmainen, "Evolving the neural network model for forecasting air pollution time series," *Eng. Appl. Artif. Intell.*, vol. 17, no. 2, pp. 159–167, 2004, doi: 10.1016/j.engappai.2004.02.002.
[3] R. W. Gore and D. S. Deshpande, "An approach for classification of health risks based on air quality levels," *Proc. - 1st Int. Conf. Intell. Syst. Inf. Manag. ICISIM 2017*, vol. 2017-Janua, pp. 58–61, 2017, doi: 10.1109/ICISIM.2017.8122148.
[4] L. N. Araujo, J. T. Belotti, T. A. Alves, Y. de S. Tadano, and H. Siqueira, "Ensemble method based on Artificial Neural Networks to estimate air pollution health risks," *Environ. Model. Softw.*, vol. 123, no. April 2019, 2020, doi: 10.1016/j.envsoft.2019.104567.
[5] A. Wang, J. Xu, R. Tu, M. Saleh, and M. Hatzopoulou, "Potential of machine learning for prediction of traffic related air pollution," *Transp. Res. Part D Transp. Environ.*, vol. 88, no. October, p. 102599, 2020, doi: 10.1016/j.trd.2020.102599.
[6] Y. Gao, Z. Wang, C. yang Li, T. Zheng, and Z. R. Peng, "Assessing neighborhood variations in ozone and PM2.5 concentrations using decision tree method," *Build. Environ.*, vol. 188, no. June 2020, p. 107479, 2021, doi: 10.1016/j.buildenv.2020.107479.
[7] S. M. Cabaneros, J. K. Calautit, and B. R. Hughes, "A review of artificial neural network models for ambient air pollution prediction," *Environ. Model. Softw.*, vol. 119, no. June 2018, pp. 285–304, 2019, doi: 10.1016/j.envsoft.2019.06.014.
[8] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption," *Energy Build.*, vol. 147, pp. 77–89, 2017, doi: 10.1016/j.enbuild.2017.04.038.
[9] B. Ekiz, O. Baygul, H. Yalcintan, and M. Ozcan, "Comparison of the decision tree, artificial neural network and multiple regression methods for prediction of carcass tissues composition of goat kids," *Meat Sci.*, vol. 161, p. 108011, 2020, doi: 10.1016/j.meatsci.2019.108011.
[10] S. Moni, E. Aziz, A. P. P. Abdul Majeed, and M. Malek, "The prediction of blue water footprint at Semambu water treatment plant by means of Artificial Neural Networks (ANN) and Support Vector Machine (SVM) models," *Phys. Chem. Earth*, vol. 123, no. December 2019, p. 103052, 2021, doi: 10.1016/j.pce.2021.103052.
[11] T. S. Chang, "A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14846–14851, 2011, doi: 10.1016/j.eswa.2011.05.063.
[12] S. Zhou, J. Li, Y. Zhang, X. Liu, and W. Zhang, "Prediction of the ground temperature variations caused by the operation of GSHP system with ANN," *Geothermics*, vol. 95, no. May, p. 102140, 2021, doi: 10.1016/j.geothermics.2021.102140.
[13] T. Akhir, "KLASIFIKASI TINGKAT UDARA DKI JAKARA DENGAN Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung," 2021.
[14] W. Song, Y. Zhang, and S. C. Park, "A novel deep auto-encoder considering energy and label constraints for categorization," *Expert Syst. Appl.*, vol. 176, no. February, p. 114936, 2021, doi: 10.1016/j.eswa.2021.114936.
[15] A. Performance, E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. Moh, "The Effect of Preprocessing Techniques , Applied to Numeric," 2021.
[16] D. Saripurna, A. Calam, Y. Yusnidah, and Z. Lubis, "Sistem Cerdas Pemanggang Jagung Semi Otomatis Berbasis Mikrokontroler Menggunakan Metode PWM (Pulse Width Modulation," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 82, 2019, doi: 10.53513/jis.v18i1.108.
[17] A. I. Sang, E. Sutoyo, and I. Darmawan, "Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 8954–8963, 2021.
[18] F. ezzahra Salamate and J. Zahi, "Supervised learning: classification using decision trees for better practice in epidemiology Case study: The prevalence of tuberculosis," *Procedia Comput. Sci.*, vol. 201, pp. 783–788, 2022, doi: 10.1016/j.procs.2022.03.107.
[19] B. Li, Y. Lee, W. Yao, Y. Lu, and X. Fan, "Development and application of ANN model for property prediction of supercritical kerosene," *Comput. Fluids*, vol. 209, 2020, doi: 10.1016/j.compfluid.2020.104665.
[20] F. Arias del Campo, M. C. Guevara Neri, O. O. Vergara Villegas, V. G. Cruz Sánchez, H. de J. Ochoa Domínguez, and V. García Jiménez, "Auto-adaptive multilayer perceptron for univariate time series classification," *Expert Syst. Appl.*, vol. 181, no. January, 2021, doi: 10.1016/j.eswa.2021.115147.
[21] T. Le Minh, L. Van Tran, and S. V. T. Dao, "A Feature Selection Approach for Fall Detection Using Various Machine Learning Classifiers," *IEEE Access*, vol. 9, pp. 115895–115908, 2021, doi: 10.1109/ACCESS.2021.3105581.
[22] H. B. Chakrapani, S. Chouraisa, A. Saha, and J. N. Swathi, "Predicting Performance Analysis of System Configurations to Contrast Feature Selection Methods," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, 2020, doi: 10.1109/ic-ETITE47903.2020.106.
[23] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016*, no. Cv, pp. 78–83, 2016, doi: 10.1109/IACC.2016.25.