The 4th International Seminar of Science and Technology
ISST 2024 Vol 4 (2025) 008
Innovations in Science and Technology to Realize Sustainable Development Goals
Faculty of Science and Technology
Universitas Terbuka

Implementation of Decision Tree Algorithm for Activity Recommendations Based on Air Quality Index (AQI) and PM2.5 Pollution in Indonesia

Mayang Anglingsari Putri*, Denisha Trihapningsari, Irpan Kusyadi, and Hasan Basri

Terbuka University, Information System Study Program, South Tangerang, Banten, Indonesia, 15437

Abstract - The increasing air pollution in major cities across Indonesia has raised serious public health concerns. This research aims to develop a recommendation system for daily activities based on the Air Quality Index (AQI) and PM2.5 levels. Using the Decision Tree algorithm, this study categorizes air quality conditions and provides appropriate activity recommendations, such as whether it is safe to exercise outdoors or if it is better to stay indoors. The model utilizes AQI and PM2.5 data collected from various Indonesian cities. The results indicate that the Decision Tree algorithm is effective in providing accurate activity recommendations based on air quality, demonstrating significant accuracy in classifying air conditions. The implementation of this system is expected to aid individuals in making informed decisions about their daily activities, thereby mitigating health risks associated with air pollution exposure. The urgency of this research lies in the need for a more adaptive and personalized system to provide activity recommendations based on real-time air quality data. Amid the rising cases of respiratory illnesses and diseases related to air pollution exposure, this study plays a crucial role in supporting a healthier and safer lifestyle. In addition, the implementation of this system can serve as a foundation for public policy and environmental risk mitigation strategies that are more data-driven and technology-based in the future.

Keywords: decision tree, air quality, AQI, PM2.5, activity recommendation, Indonesia

1 Introduction

Air pollution is one of the largest environmental challenges in Indonesia, particularly in rapidly growing major cities such as Jakarta, Surabaya, and Bandung. The increasing levels of pollution, especially the concentration of Particulate Matter 2.5 (PM2.5), have become a major concern due to their serious impact on public health. PM2.5, fine particles that can penetrate deeply into the respiratory system, contribute to various diseases ranging from respiratory disorders to cardiovascular issues. Poor air quality also affects the overall quality of life, making it crucial to have a system that can guide safe outdoor activities based on current air conditions [1]. By providing timely and personalized recommendations, such a system can help individuals make informed decisions and reduce exposure to harmful pollutants. In the long term, this approach can also raise public awareness and support healthier urban living environments.

* Corresponding author: mayang.anglingsari@ecampus.ut.ac.id

To address this issue effectively, a data-driven air quality recommendation system is essential. This research aims to develop an activity recommendation system that provides advice based on air quality conditions measured by the Air Quality Index (AQI) and PM2.5 concentrations. Using relevant data from various locations in Indonesia, this system is designed to assist the public in making better decisions about their outdoor activities [2]. For example, the system might recommend exercising outdoors if the air quality is good or suggest limiting outdoor activities and using masks when pollution reaches hazardous levels.

The Decision Tree algorithm was chosen as the primary method in this research due to its ability to handle high-complexity data and generate clear, interpretable decisions [3]. Decision Trees are effective in mapping the relationship between input variables, such as AQI and PM2.5, with activity recommendation outputs, allowing for systematic decision-making based on specific conditions. The resulting decision tree structure enables users to better understand the decision-making process and adapt recommendations according to real-time changes in air quality. Implementing this method is expected not only to increase public awareness of the impact of air quality on health but also to facilitate more informed and data-driven decision-making.

The Decision Tree algorithm is applied to analysed AQI and PM2.5 air pollution data in Indonesia to provide appropriate activity recommendations based on air conditions. By building a decision tree structure, this algorithm classifies AQI values and PM2.5 concentrations such as Good, Moderate, Poor, and Very Poor, as well as Low, Medium, High, and Very High, to generate clear and understandable suggestions. The Decision Tree processes these input data to provide precise recommendations, such as whether it is safe to exercise outdoors or if activity should be limited [4]. This method offers clarity in decision-making and can quickly adjust recommendations based on current data, supporting health management and activity planning in environments affected by air pollution.

2 Materials and methods

2.1 Research Method

The research method started with data collection from the IQAir website, which provides up-to-date information on the Air Quality Index (AQI) and PM2.5 levels. Next, the data is processed through cleaning and normalization to ensure accuracy and consistency. A Decision Tree algorithm was applied to build a model that classifies air quality conditions based on the processed data. The model is trained using historical data and evaluated for performance. Based on the classification results, activity recommendations are suggested according to the air quality level. Finally, the research results are analysed, interpreted, and presented through visualizations and comprehensive reports.

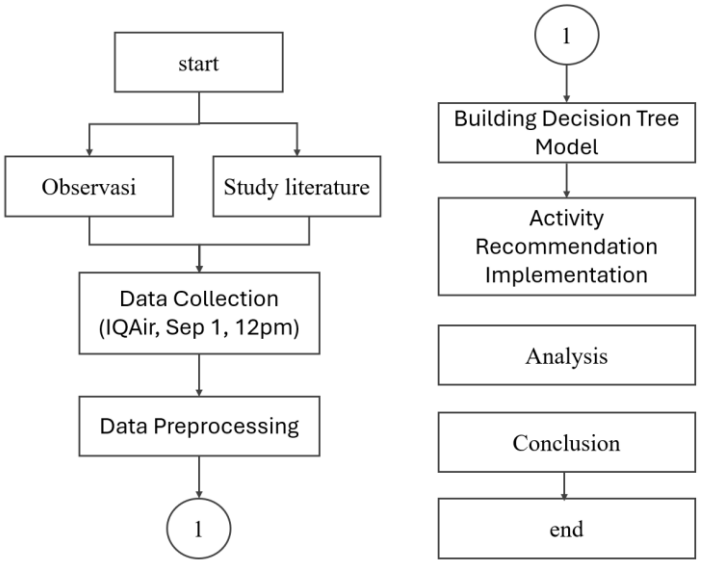

Fig. 1. Method

2.2 The Air Quality Index (AQI)

The Air Quality Index (AQI) is a measure used to assess air pollution levels and their impact on public health, based on concentrations of key pollutants such as PM2.5, PM10, ozone, nitrogen dioxide, and sulfur dioxide. AQI data provides easy-to-understand information on air quality, with categories ranging from “Good” to “Very Hazardous,” helping individuals make informed decisions regarding outdoor activities. This data provides up-to-date information on air quality in various cities in Indonesia, which includes concentrations of pollutants such as PM2.5, PM10, and other air quality indicators. By referring to the real-time data from IQAir (ww.iqair.com), this study analyzes the impact of air quality on health and provides relevant recommendations based on the current conditions reported at that time (see Fig. 2.).

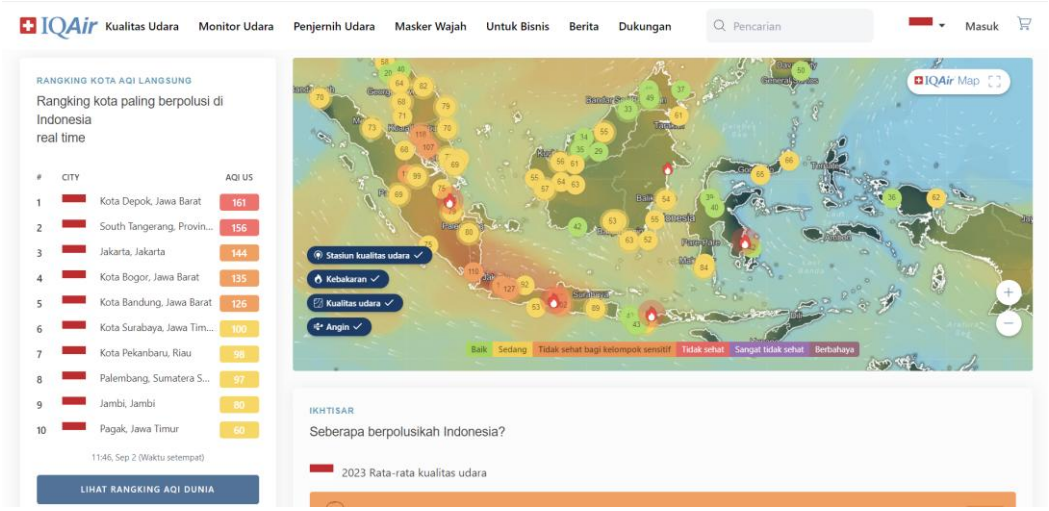

Fig. 2. Air quality in Indonesia IQAir

2.3 Particular Matter 2.5

PM2.5 stands for Particulate Matter with a diameter of less than 2.5 micrometers. PM2.5 consists of fine particles with a diameter smaller than 2.5 micrometers that can penetrate deeply into the respiratory tract and bloodstream, causing various health problems. The main sources of PM2.5 include fossil fuel combustion, industrial processes, biomass burning, and natural processes such as forest fires. Long-term exposure to PM2.5 can lead to respiratory disorders, cardiovascular diseases, and increased health risks for vulnerable groups such as children and the elderly. Monitoring PM2.5 is a crucial part of the Air Quality Index (AQI) system for assessing air quality and protecting public health.

2.4 Decision Tree Algorithm

The decision Tree algorithm is a machine learning method used for classification and regression by dividing data into subsets based on the most relevant features, forming a decision tree structure. This tree consists of nodes that represent features or attributes, branches that indicate decisions or rules, and leaves that indicate the final result or class [5]. The algorithm works by selecting features that optimally separate the data at each step, typically using metrics such as entropy, information gain, or Gini index, to make decisions that are most effective in classifying or predicting outcomes. A decision Tree provides easy-to-understand interpretation due to its structure similar to a flowchart and can handle complex data with different types of features [6].

3 Results and discussion

This research collects data from IQAir to obtain the latest information on the Air Quality Index (AQI) and PM2.5 air pollution concentrations in various cities in Indonesia. This data is crucial for the research as it provides a real-time picture of air quality conditions, which is then used to develop an activity recommendation model based on the Decision Tree algorithm. Below is an example of air quality data from Pekanbaru city, with an AQI of 117 and PM2.5 value of 42 µg/m³.

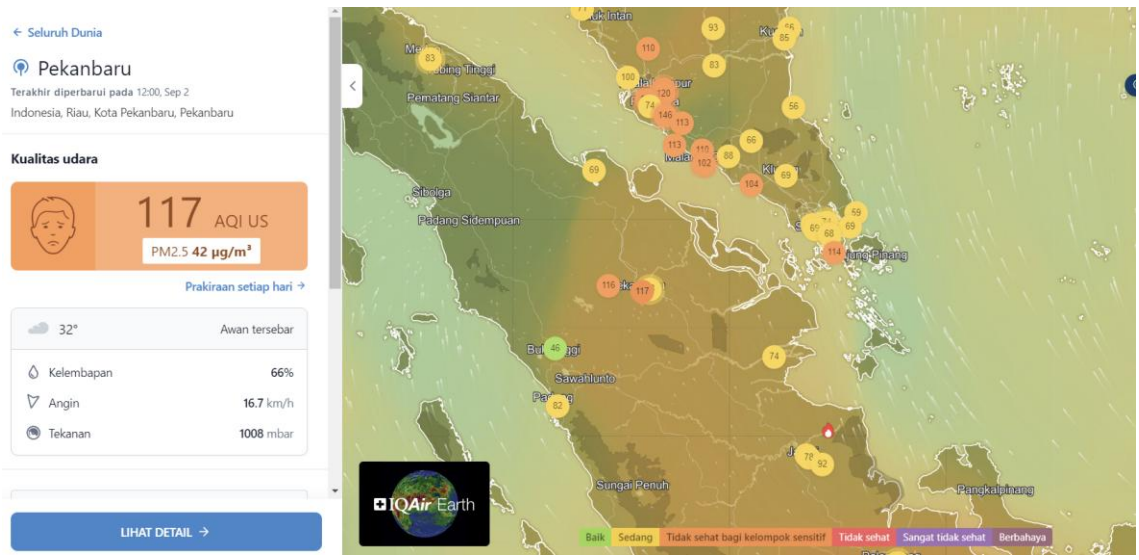

Fig. 3. Air quality in Riau Pekanbaru based on IQAir

The following table provides up-to-date information on air quality in several major cities in Indonesia, which can be used to assess the impact of air pollution on public health and to provide safe

outdoor activity recommendations. This research uses Air Quality Index (AQI) data accessed from the IQAir website on September 1st at 12:00 PM [2].

Table 1. Air Quality Index (AQI) and PM2.5 Air Pollution for sample cities in Indonesia

No	City	AQI	PM2.5 (µg/m³)
1	Kota Bogor, Jawa Barat	127	45.9
2	Tangerang Selatan Provinsi Banten	148	54.5
3	Jakarta	143	52,7
4	Malang, Jawa Timur	72	20.5
5	Denpasar, Bali	28	5
6	Kendari, Sulawesi Tenggara	43	7.8
7	Riau, Pekanbaru	117	42
8	Palangka Raya, Kalimantan Tengah	49	8.9
9	Bandung, Jawa barat	143	52.5
10	Sorong, Papua Barat	41	**7.3**

Information:

AQI: Air Quality Index that indicates the air quality in the city, with higher values indicating worse air conditions.

PM2.5 (µg/m³): Concentration of air particles with a diameter less than or equal to 2.5 micrometers, which is the main indicator of air pollution.

3.1 Application of the Decision Tree Algorithm

In the process of creating a decision tree, we first categorize the data into groups such as Low, Medium, and High based on attributes such as AQI and PM2.5. Next, we calculate Entropy to measure the uncertainty of the overall dataset and after splitting the data based on specific attributes. Gain, which is the difference between the total Entropy and the Entropy of subsets, is used to determine the best attribute for splitting [7]. Based on the highest Gain, we construct the decision tree by splitting the data according to that attribute, and this process is repeated recursively until the data in each subset is homogeneous or a stopping condition is met.

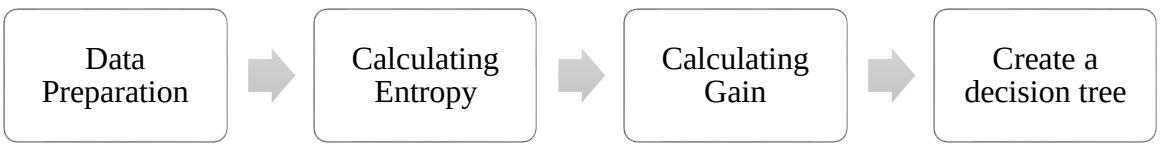

Fig. 4. Steps for implementing the Decision Tree algorithm

Create categories from AQI and PM2.5 with information like the following table.

Table 2. List of Categories

	Low	Medium	Hight
AQi	AQI < 50	50 ≤ AQI < 100	AQI ≥ 100
PM2.5	PM2.5 < 10 µg/m³	10 ≤ PM2.5 < 30 µg/m³	PM2.5 ≥ 30 µg/m³

From the following table, air quality is assessed based on AQI and PM2.5, with quality categorized as High, Medium, and Low, as shown in the table below.

Table 3. City data by category

No	City	Air Quality (AQI)	Air Quality (PM2.5)
1	Kota Bogor, Jawa Barat	High	High
2	Tangerang Selatan Provinsi Banten	High	High
3	Jakarta	High	High
4	Malang, Jawa Timur	Medium	Medium
5	Denpasar, Bali	Low	Low
6	Kendari, Sulawesi Tenggara	Low	Low
7	Riau, Pekanbaru	High	High
8	Palangka Raya, Kalimantan Tengah	Low	Low
9	Bandung, Jawa Barat	High	High
10	Sorong, Papua Barat	Low	Low

3.2 Calculating Total Entropy

In a Decision Tree, the class proportion (or class probability) is calculated by dividing the number of instances of a particular class by the total number of instances in the subset of data being analyzed [8]. The formula for calculating the total entropy is as follows.

$$\text{Class Proportions} = \frac{Number\ of\ Instances\ of\ Class}{Total\ Number\ of\ Instances\ in\ Subset} \quad (1)$$

$$\text{Entropy}_{total} = -(P_{low}\log_2(P_{low}) + P_{medium}\log_2(P_{medium}) + P_{high}\log_2(P_{high})) \quad (2)$$

The results of the class proportion are as follows:

Low: 4 with p Low = 4/10= 0.4

Medium: 1 with p medium = 1/10= 0.1

Height: 5 with p Height = 1/10= 0.5

$$\text{Entropy}_{total} = -(0.4\log_2(0.4) + 0.1\log_2(0.1) + 0.5\log_2(0.5))$$
$$= -\left(0.4 \times (-1.322) + 0.1 \times (-3.322) + 0.5 \times (-1)\right)$$
$$= -(-0.529 - 0.332 - 0.5)) = 1.361$$

The next step is to calculate Entropy based on AQI and PM2.5

Table 4. Category data

AQI Category	PM2.5 Category	Air Quality
Low	low	4 low
medium	medium	1 medium
high	high	5 high

After carrying out the calculations with the data above, the calculations were carried out by calculating the combination of AQI and PM2.5. Because AQI and PM2.5 have the same distribution, in this study we calculated the entropy of only one of them with a value of 0. The entropy becomes

zero when all the data in the subset is in the same category, because there is no uncertainty or variation in the data. In this case, all elements are homogeneous, so no additional information can be obtained from the data. For example, if Low AQI always results in the category "Low," then knowing the AQI is enough to ensure air quality, making entropy, or a measure of uncertainty, zero. This shows that the subset is completely definite and uniform, with no variations that need to be explained further.

$$\text{Entropy}_{AQI\ PM2.5} = \frac{4}{10} \times 0 + \frac{1}{10} \times 0 \frac{5}{10} \times 0 = 0$$

The explanation of the formula above is data $\frac{4}{10}, \frac{1}{10}, \frac{5}{10}$ is the probability of each subset in the data grouped by AQI and PM2.5. For example, $\frac{4}{10}$ indicates that 4 out of 10 events have low AQI and PM2.5. Meanwhile, the number zero can be (×0). This shows the entropy of each subset. Since each subset is homogeneous (for example, all data in the first subset has "Low" air quality), the entropy of that subset is 0. There is no uncertainty in this subset because all values in the subset are the same.

3.4 Calculating Total Gain

Gain is a measure that shows how much uncertainty (entropy) is reduced after data is split based on an attribute. In this context, Gain indicates how much information is gained by separating the data based on AQI and PM2.5.

$$Gain = \text{Entropy}_{total} - \text{Entropy}_{AQI\ PM2.5} \quad (3)$$

$$Gain = 1.361 - 0 = 1.361$$

Splitting the data based on AQI/PM2.5 completely reduces the uncertainty in the data, providing a Gain of 1.361. A Gain of 1.361 indicates that this attribute has a good capability to split the data clearly and assist in making better decisions in the decision tree.

4 Create a Decision Tree

A decision tree with activity recommendations is a decision tree model designed to provide suggestions or recommendations about activities or actions that should be taken based on existing conditions or features. Typically, it is used to assist in decision making based on various factors that influence the final outcome.

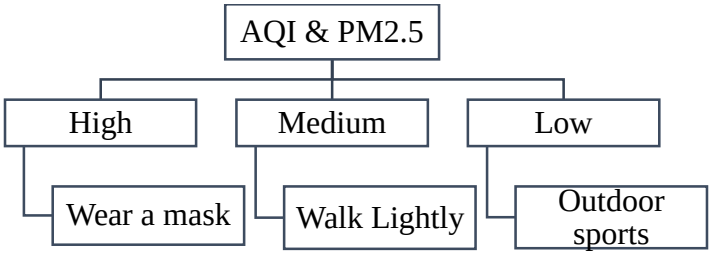

Fig. 5. Decision Tree

This decision tree starts with the main node as AQI (Air Quality Index), which serves as the root of the tree. From this node, the tree branches into three based on AQI values: for High, the recommended activity is to Use a mask; for Medium, it is advised to Take a light walk; and for Low,

the recommended activity is Outdoor exercise. This structure allows for clear, data-driven decision-making about the appropriate activity based on the measured air quality. Below is the implementation in Python. Python is a highly popular language for data analysis, machine learning, and visualization. The following code uses the scikit-learn library to build the decision tree and matplotlib to display its visualization.

```
df = pd.DataFrame(data)

# Label Encoding untuk fitur dan label
le_aqi = LabelEncoder()
le_pm25 = LabelEncoder()
le_rekomendasi = LabelEncoder()

df['AQI'] = le_aqi.fit_transform(df['AQI'])
df['PM2.5'] = le_pm25.fit_transform(df['PM2.5'])
df['Rekomendasi'] = le_rekomendasi.fit_transform(df['Rekomendasi'])

# Fitur dan label
X = df[['AQI', 'PM2.5']]
y = df['Rekomendasi']

# Membangun model pohon keputusan
model = DecisionTreeClassifier()
model.fit(X, y)

# Menampilkan pohon keputusan menggunakan matplotlib
plt.figure(figsize=(12, 8))
plot_tree(
    model,
    feature_names=['AQI', 'PM2.5'],
    class_names=le_rekomendasi.classes_,
    filled=True,
    rounded=True,
    fontsize=10
```

Fig. 6. Implementation in Python

The output of the Python program is a visualization of the decision tree that illustrates how decisions are made based on air quality (AQI) and PM2.5 concentration. In this decision tree, each branch represents the decision made for a specific air quality category. For high AQI and PM2.5 values, the recommended activity is to use a mask; for medium values, it is advised to take a light walk; and for low values, the recommendation is to engage in outdoor exercise. This visualization helps in understanding how the model makes decisions and recommends activities based on the measured air quality conditions.

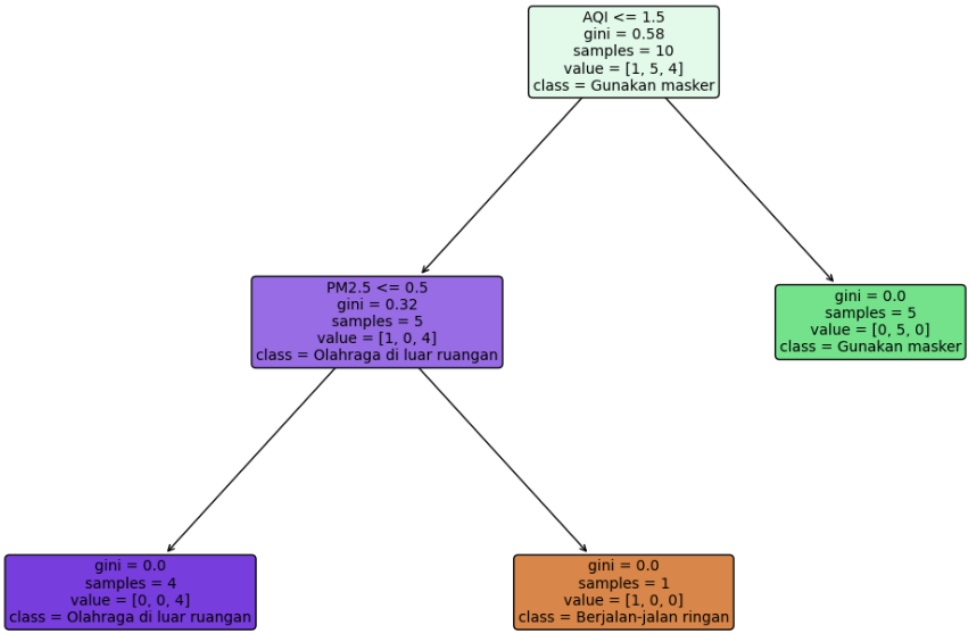

Fig. 7. Implementation in Python

5 Conclusion

The implementation of the Decision Tree algorithm for activity recommendations based on the Air Quality Index (AQI) and PM2.5 air pollution levels in Indonesia has resulted in an effective system for categorizing air quality and suggesting appropriate activities based on these conditions. The model currently provides simple yet actionable recommendations: wearing a mask when the AQI is high, engaging in light walking for medium AQI levels, and performing outdoor exercise when the AQI is low.

To ensure the sustainability and scalability of this research, several steps can be taken. First, the dataset should be expanded to include more geographical locations and additional variables, such as temperature, humidity, and population density. Second, experimenting with alternative machine learning algorithms—such as Random Forests or Gradient Boosting—could enhance the model's accuracy and robustness. Third, incorporating real-time data processing capabilities will allow the system to provide timely and dynamic recommendations.

References

[1] A. Oktaviani and H. Justinawati, "Prediksi rata-rata zat berbahaya di DKI Jakarta berdasarkan indeks standar pencemar udara menggunakan metode Long Short-Term Memory," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 1, pp. 1-10, Apr. 2021.

[2] IQAir, "Air quality monitoring," IQAir. [Online]. Available: https://www.iqair.com/. [Accessed: Sep. 2, 2024].

[3] D. Istiawan and L. Khikmah, “Implementation of C4.5 Algorithm for Critical Land Prediction in Agricultural Cultivation Areas in Pemali Jratun Watershed,” Indones. J. Artif. Intell. Data Min., vol. 2, no. 2, p. 67, 2019, doi: 10.24014/ijaidm.v2i2.7569.

[4] P. Assiroj, "Kajian Perbandingan Teknik Klasifikasi Algoritma C4 . 5 , Naïve Bayes Dan Cart Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Stmik Rosma Karawang) (Studi Kasus : Stmik Rosma Karawang)," Media Inform., vol. 15, no. 2, pp. 1–17, 2018.

[5] J. Eska, "Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C45," JURTEKSI (Jurnal Teknol. dan Sist. Informasi), vol. 2, pp. 9–13, 2016.

[6] Nasrullah, A H. Mplementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris. Jurnal Ilmiah Ilmu Komputer Vol7 Filkom Universitas AL Asyariah Mandar. 2021.

[7] Hariati, M. Wati, and B. Cahyono, "Penerapan Algoritma C4.5 Decision Tree pada Penentuan Penerima Program Bantuan Pemerintah Daerah Kabupaten Kutai Kartanegara," Jurti, vol. 2, no. 1, pp. 27–36, 2018.

[8] S. Bahri and A. Lubis, "Metode klasifikasi decision tree untuk memprediksi juara English Premier League," *Jurnal Sintaksis: Pendidikan Guru Sekolah Dasar, IPA, IPS dan Bahasa Inggris*, vol. 2, no. 1, pp. 1-10, Apr. 2020.