EPI International Journal of Engineering
Volume 8, Number 2, August 2025, pp. 112-121
DOI: 10.25042/epi-ije.082025.08

*p*ISSN 2615-5109
*e*ISSN 2621-0541

# Predicting Flood Risk in Manado City Using the C4.5 Decision Tree Algorithm: A Data-Driven Approach

Apriandy Angdresey[a], Indah Yessi Kairupan[b,*], Ranodeyansa Rumanjar[c], Andre Gabriel Mongkareng[d], Ignatius Lucky Henokh Tangka[e]

[a]Department of Informatics Engineering, Universitas Katolik De La Salle, Manado, Indonesia. Email:aangdresey@unikadelasalle.ac.id
[b]Department of Information Systems, Institut Agama Kristen Negeri Manado, Indonesia. Email:ikairupan@iaknmanado.ac.id
[c]Department of Informatics Engineering, Universitas Katolik De La Salle, Manado, Indonesia. Email:19013013@unikadelasalle.ac.id
[d]Department of Informatics Engineering, Universitas Katolik De La Salle, Manado, Indonesia. Email:20013004@unikadelasalle.ac.id
[e]Department of Informatics Engineering, Universitas Katolik De La Salle, Manado, Indonesia. Email:20013047@unikadelasalle.ac.id

**Abstract**

Flooding, a natural disaster commonly triggered by heavy rainfall or blocked waterways, poses a persistent threat to Manado City due to its proximity to the sea and major rivers, Tondano and Tikala. This study develops a flood prediction application using the C4.5 Decision Tree algorithm based on historical data (2017–2021) from five rivers. Input variables include rainfall, water discharge, runoff coefficients, and river cross-sectional data. The model supports the Sulawesi River Regional Office I by predicting flood conditions—flood, prone to flooding, or no flooding—to enhance mitigation strategies. Experimental results show robust predictive performance, achieving accuracies of 86.13%-87.07% across different data splits. Future integration with the Internet of Things (IoT) is proposed to enable real-time data acquisition, thereby improving the system's responsiveness and flood risk management effectiveness.

*Keywords:* C4.5 algorithm; decision tree; flooding prediction

## 1. Introduction

Natural phenomena are events in nature that occur beyond human control or influence. Among these phenomena, some confer benefits while others pose risks to human well-being, often referred to as natural disasters [1]. One such disaster is flooding, in which water exceeds its normal boundaries and inundates residential areas, usually triggered by heavy rainfall or the blockage of waterways. The consequences of flooding include infrastructure damage, the spread of diseases, and loss of life [2]. Manado City, the capital of North Sulawesi province in Indonesia, attracts numerous tourists and has a population of around 400,000. Situated in a low-lying area near the sea, the city is prone to flooding due to its proximity to two major rivers, the Tondano and the Tikala [3]. This geographic vulnerability has made Manado City susceptible to recurring flooding. In 2014, a significant flood devastated the city, affecting 59 sub-districts. Subsequently, in late 2017 and early 2023, Manado experienced additional flooding episodes, highlighting the city's vulnerability to extreme weather events.

Prediction involves foreseeing or estimating future events or situations by drawing on current or historical information and analysis. Its primary goal is to help make more informed decisions [4]. While predictions can serve as valuable decision-making tools, they are prone to inaccuracies due to various factors that influence the outcome. This methodology entails analyzing past data trends to forecast the category into which new data will likely fall, a classification technique. It leverages existing patterns to support anticipatory decision-making. For instance, in a study [5], researchers used fuzzy logic to predict floods on the Citarum River, based on rainfall and water discharge. However, the study used only 1 year of data from a single river, potentially limiting the accuracy of the predictions.

The Decision Tree C4.5, a machine learning algorithm, is primarily used for classification tasks but can also be used for predictive modeling. It has evolved from its predecessor, the Decision Tree ID3 (Iterative Dichotomiser 3), to address several inherent shortcomings [6]. Furthermore, this algorithm integrates clustering techniques into its framework, allowing for the creation of a more refined decision tree structure. This is achieved by assigning varying levels of importance to attributes during the decision-making process. In the study [7], the authors applied the C4.5 decision tree algorithm to predict flood damage. However, the study merely presents the results as a decision tree, without developing an automated prediction application. Similarly, in predicting other

*Corresponding author. Tel.: +62-852-4183-0173
*Institut Agama Kristen Negeri Manado*
*Minahasa, Indonesia, 95661*

natural disasters, such as forest fires, the Decision Tree C4.5 algorithm has been used in studies [8], yet no application has been developed for testing and prediction.

This study focuses on predicting floods in Manado City using the C4.5 decision tree algorithm, aiming to develop an application to assist the Manado City government, specifically the Sulawesi River Regional Office I, in obtaining flood prediction information tailored to Manado. The aim is to create a prediction application to reduce the flooding problems that continue to occur in Manado City. By analyzing historical data from five rivers: Bailang, Malalayang, Tikala-Sawangan, Tikala-Paal4, and Tondano [9], the research seeks to provide valuable insights to government agencies involved in flood mitigation [10]. By enhancing predictive capabilities, this study can empower governments to implement adequate preventive measures and deploy appropriate strategies to protect citizens from the adverse effects of flooding.

The remainder of this paper is structured as follows: Section 2 provides an in-depth literature review of pertinent studies and examines the theoretical foundations of this research. Furthermore, Section 3 outlines the methodology utilized in this study. Section 4 presents the research findings and discussions based on the testing results. Finally, Section 5 concludes this study and suggests directions for future work.

## 2. Literature Review

Data mining is a prevalent method to sift through vast data collections to uncover relevant or valuable insights. Various techniques are employed in data mining, including pattern recognition and predictive modeling [11]. These techniques aim to extract meaningful information from the data. Two commonly used techniques are classification and prediction for supervised learning, and clustering for unsupervised learning. For instance, in a study cited in [12], the authors conducted clustering analysis to identify patterns and anomalies in environmental monitoring, focusing specifically on temperature and ammonia levels. Clustering groups data points with similar characteristics, enabling the detection of underlying patterns or anomalies. This approach can provide valuable insights into environmental conditions, helping identify trends or irregularities that may require further investigation or action.

Moreover, supervised learning involves categorizing a dataset with labels or annotations and organizing it into predefined classes. This classification task applies to both structured and unstructured data formats. Initially, the classification process entails predicting the class or category to which each data point belongs. These categories are often denoted as labels, targets, or classes. Additionally, supervised learning algorithms aim to approximate or forecast the function within the mapping domain, allowing them to predict the output variable when provided with input variables [13]. However, it must undergo training before the system can accurately assign labels. As in the study discussed in [14], the author develops and implements an Internet of Things-based soil monitoring system. This system utilizes sensors to collect data, which is then used to classify plant treatment categories using a classification algorithm. This approach enables users to access information about the necessary plant treatments, enhancing agricultural management practices.

Decision Trees are predictive models commonly utilized in supervised learning, celebrated for their wide-ranging applicability, interpretability, and resilience [15]. Like a tree structure, a decision tree consists of branches representing various choices and their potential outcomes, including chance events, resource costs, and utilities. Its utility extends across diverse domains such as agriculture, medicine, education, and environmental management. The primary objective of decision trees is to facilitate optimal decision-making based on available data. This method uses algorithms such as ID3, C4.5, and CART to determine the most favorable path by calculating entropy, information gain, and the Gini index [16]. Decision trees have gained significant traction in data mining due to their simplicity and efficacy in classification. Data Mining, on the other hand, is a process focused on uncovering patterns and insights from large datasets [17].

C4.5, an extension of the ID3 algorithm, was developed by Ross Quinlan in 1986 and has since become a prominent decision tree algorithm [18]. It is adept at constructing decision trees from training data, making it versatile for various datasets, as it can handle both categorical and continuous attribute values. The algorithm selects the best attribute at each node based on the information gain ratio metric, which accounts for the intrinsic information of each attribute. Additionally, C4.5 employs post-pruning to reduce overfitting and generate more generalizable trees. Moreover, it can generate human-readable if-then rules from the decision tree, enhancing interpretability. For instance, in [19], it was used to determine specializations in an informatics engineering program with an impressive accuracy of 93.89%. Furthermore, in [20], C4.5 was employed to assess students' academic abilities, achieving an accuracy of 80.6%. Similarly, in (Prediction Model of Teacher Candidate Student Graduation Status: Decision Tree C4.5, Naive Bayes, and k-NN), it was utilized to predict students' graduation status using three algorithms and to compare their performance. The results showed that the decision tree C4.5 algorithm performed better than the other algorithms, achieving an accuracy of 93.90%.

Additionally, in the study [21], researchers used three types of algorithms—C4.5, naive Bayes, and SVM—to predict lung cancer survival rates; the performance of the C4.5 algorithm notably improved, achieving an accuracy of 82.6%, surpassing naive Bayes (66.1%) and SVM (54.9%). Furthermore, in disaster prediction tasks, the C4.5 algorithm remains a variable option [22]. The authors utilized decision trees to forecast potential floods, achieving an accuracy of 84.0%. This performance was compared with that of a deep learning algorithm, such as an ANN, which achieved an accuracy of 96.65%. Moreover, another study using C4.5 was integrated into a wireless monitoring system for aquariums based on the Internet of Things, achieving an impressive 97.8% accuracy in classifying aquarium conditions [23]. These findings highlight the versatility and efficacy of the C4.5 algorithm across diverse domains, affirming its potential

to solve complex real-world problems in healthcare, disaster prediction, and beyond.

## 3. Methodology

The flowchart in Fig. 1 shows the architecture and processes of our system, emphasizing key components such as inputs, outputs, and decision points. It effectively depicts the direction of data flow and the interactions between modules, enhancing our understanding of the system's operational dynamics and efficiency. The data collection process commences with sourcing information from reliable external entities, specifically BWS Sulawesi 1 and UPTD BAPELITBANGDA Manado. Once the raw data is acquired, the next step is to label it to categorize the information appropriately. This labeling process is vital as it defines the categories for each data point. Following the labeling phase, the next stage is preprocessing, where the labeled data undergoes normalization, ensuring it is standardized and ready for input into the decision tree model.

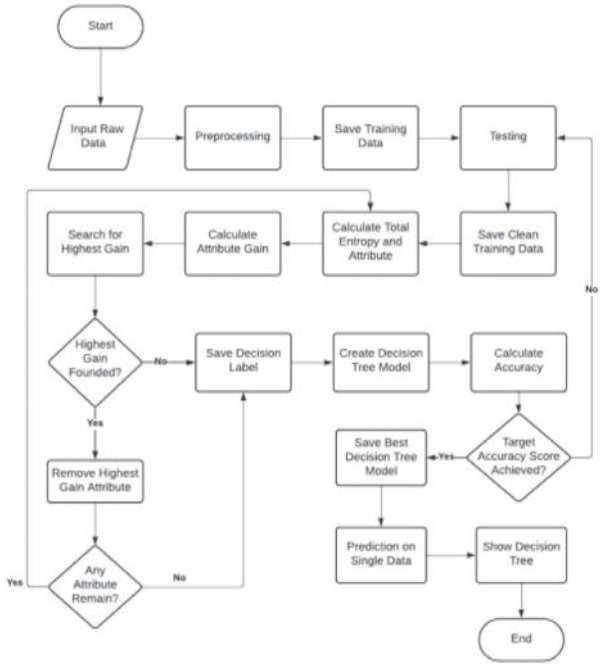


Figure 1. System flowchart

Normalization is essential to ensure uniformity and consistency in the data, thereby enhancing the effectiveness of the decision tree model. The final phase of our system involves the C4.5 decision tree. This section explains how the algorithm processes normalized data to generate outputs, such as predictions. The C4.5 decision tree algorithm uses preprocessed data to construct a decision tree, where each node represents a decision based on specific data attributes. By traversing this tree, the algorithm makes predictions about future outcomes based on the input data. This process is integral to our system's functionality, enabling us to derive valuable insights and predictions from the data we collect.

### *3.1. Data collection*

The data employed in this study are sourced from BWS Sulawesi I and UPTD BAPELITBANGDA Manado. These external organizations provide invaluable data that will serve as the basis for our analysis. The dataset consists of 71 rows, each representing a unique set of attributes. These attributes include rainfall data, water discharge, the runoff coefficient, and river cross-section measurements. A comprehensive overview of these attributes is provided in Table 1.

Table 1. Hydrological and river cross-section dataset

| No | Rainfall Data | Water Discharge | Runoff Coefficient | River Cross-Section |
|---|---|---|---|---|
| 1 | 0.7 | 0.914 | 0.17 | 15 |
| 2 | 5.9 | 0.864 | 0.17 | 15 |
| 3 | 8.5 | 0.722 | 0.17 | 15 |
| 4 | 2.5 | 0.553 | 0.17 | 15 |
| 5 | 1.6 | 0.653 | 0.17 | 15 |
| … | … | ….. | …… | ….. |
| 70 | 75 | 2.188 | 0.18 | 20 |
| 71 | 75 | 15 | 0.18 | 20 |

### *3.2. Labeling*

Based on the collected data, the Hydrograph Method labels it as "prone to flood" or "not prone to flood." This method involves a series of sequential steps:

1) Runoff Volume Calculation: Use Eq. 1, where $V$ represents the runoff volume, $C$ is the runoff coefficient, and rainfall data is denoted by $X_1$.
2) Flood Discharge Calculation: Use Eq. 2 to calculate $Q$ (flood discharge) based on $X_2$ (water discharge), $V$ (runoff volume), and $X_3$ (the river cross-section).
3) River Flood Discharge Threshold: Determine $X_4$ as each river's flood discharge threshold using Eq. 3.

$$V = X_1 * C \quad (1)$$

$$Q = X_2 + (V * X_3) \quad (2)$$

$$X_4 = X_3 * 10 \quad (3)$$

$$m = \frac{X_4}{3} \quad (4)$$

Table 2. Hydrograph method result

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | $V$ | $Q$ |
|---|---|---|---|---|---|---|
| 1 | 0.7 | 0.914 | 0.17 | 15 | 0.119 | 2.699 |
| 2 | 5.9 | 0.864 | 0.17 | 15 | 1.003 | 15.909 |
| 3 | 8.5 | 0.722 | 0.17 | 15 | 1.445 | 22.397 |
| 4 | 2.5 | 0.553 | 0.17 | 15 | 0.425 | 6.928 |
| 5 | 1.6 | 0.653 | 0.17 | 15 | 0.272 | 4.715 |
| … | … | ….. | …… | ….. | …. | ….. |
| 70 | 75 | 2.188 | 0.18 | 20 | 13.5 | 272.188 |
| 71 | 75 | 15 | 0.18 | 20 | 13.5 | 285 |

Based on the calculated flood discharge ($Q$) and river flood discharge threshold $X_4$, the data will be labeled into three categories:

- Category 1: When $Q>X_4$, the data will be labeled "*Flood*."
- Category 2: If $Q > m$, the data will be labeled "*Prone to Flood*," where $m$ is the normal threshold value calculated by using Eq. 4.
- Category 3: "*Not flood*" when $Q < m$.

Table 3. River flood discharge threshold result

| $X_3$ | $X_4$ |
|---|---|
| 15 | 150 |
| 20 | 200 |

The Hydrograph method is demonstrated using previously collected data, specifically from row 1. The first step is to calculate the runoff volume ($V$) as 0.7 × 0.17 = 0.119. After determining the $V$, the $Q$ is calculated using the water discharge, runoff volume, and river cross-section: $Q = 0.914 + (0.119 \times 15) = 2.699$. This process is applied to all 71 data points, with the results summarized in Table 2. Moreover, $X_4$ is derived from the results shown in Table 3. Following this, the labeling process is conducted according to the predefined rules in Table 4.

Table 4. Labeling rules

| $X_4$ | *Not Flood* | *Prone to Flood* | *Flood* |
|---|---|---|---|
| 150 | <50 | 50-150 | >150 |
| 200 | <67 | 67-200 | >200 |

The next phase involves categorizing the data into three predefined groups, as shown in Table 5. Upon completion of the labeling process, the data will be stored in a database in .csv format. This file will later be used as training data for flood prediction models specific to Manado City.

Table 5. Labeling result

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | 0.7 | 0.914 | 0.17 | 15 | Not Flood |
| 2 | 5.9 | 0.864 | 0.17 | 15 | Not Flood |
| 3 | 8.5 | 0.722 | 0.17 | 15 | Not Flood |
| 4 | 2.5 | 0.553 | 0.17 | 15 | Not Flood |
| 5 | 1.6 | 0.653 | 0.17 | 15 | Not Flood |
| … | … | ….. | …… | ….. | |
| 70 | 75 | 2.188 | 0.18 | 20 | Flood |
| 71 | 75 | 15 | 0.18 | 20 | Flood |

### *3.3. Preprocessing*

In this section, the training data attributes, such as rainfall and water discharge, will be normalized into categorical values using rules set in advance to facilitate decision-making. This process will be shown in Table 6. Based on the above rules, the data will be normalized accordingly, as shown in Table 7.

Table 6. Normalization rules

| $X_1$ | $X_2$ |
|---|---|
| <60 | <13 |
| 60-89 | 13-19 |
| ≥90 | ≥20 |

Table 7. Normalization result

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 2 | 60 - 89 | 13 - 19 | 0.17 | 15 | Not Flood |
| 3 | < 60 | < 13 | 0.17 | 15 | Not Flood |
| 4 | < 60 | < 13 | 0.17 | 15 | Not Flood |
| 5 | < 60 | < 13 | 0.17 | 15 | Not Flood |
| … | … | ….. | …… | ….. | … |
| 70 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 71 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |

### *3.4. Decision tree C4.5*

A decision tree is constructed by selecting an attribute as the root node, generating branches for each attribute value, and recursively partitioning the data along these branches. The objective is to split the data into subsets that maximize class-label purity [24]. This process is repeated until the subsets achieve homogeneity or a predefined stopping criterion is encountered. The following equation gives the entropy calculation for an attribute:

$$Entropy(S) = \Sigma_{i=1}^{n} - p_i \times log_2 p_i \quad (5)$$

$$Gain(S, A) = Entropy(S) - \Sigma_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy\,(S_i) \quad (6)$$

Where $S$ represents the dataset, $n$ is the number of partitions, and $p_i$ is the proportion of instances in partition $i$ relative to $S$. The information gained for an attribute is then calculated using Eq. 6. Where $A$ is symbolized for the attribute being evaluated, $|Si|$ is the number of instances in the $i$-th partition, and $|S|$ is the total number of cases in the dataset.

Table 8. Entropy calculation result

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy |
|---|---|---|---|---|---|
| | 71 | 36 | 21 | 14 | 1.478 |
| Rainfall Data | | | | | |
| < 60 | 61 | 36 | 21 | 4 | 1.236 |
| 60 - 89 | 8 | 0 | 0 | 8 | 0 |
| ≥ 90 | 2 | 0 | 0 | 1 | 0 |
| Water Discharge | | | | | |
| < 13 | 62 | 36 | 21 | 5 | 1.277 |
| 13 - 19 | 2 | 0 | 0 | 2 | 0 |
| ≥ 20 | 7 | 0 | 0 | 7 | 0 |
| Runoff Coefficient | | | | | |
| 0.17 | 33 | 18 | 10 | 5 | 1.411 |
| 0.18 | 38 | 18 | 11 | 9 | 1.521 |
| River Cross Section | | | | | |
| 15 | 37 | 18 | 18 | 1 | 1.152 |
| 20 | 34 | 18 | 3 | 13 | 1.325 |

This iterative process continues, selecting the attribute with the highest information gain at each step to build the decision tree. The C4.5 algorithm employs the calculations outlined in Eqs. 5 and 6 on the normalized training data. The results of the entropy calculations are presented in Table 8. For instance, the total entropy is computed as follows:

$$\begin{aligned} Entropy_{total} = & -\left(\frac{36}{71} \times log_2 \times \frac{36}{71} + \frac{21}{71} \times log_2 \times \frac{21}{71}\right. \\ & \left. + \frac{14}{71} \times log_2 \times \frac{14}{71}\right) \\ & - \left(0.45 \times (-0.979) + 0.0507(-1.757)\right. \\ & \left. + 0.197(-2.342)\right) = 1.478 \end{aligned}$$

Once the total entropy has been calculated, the entropy for each attribute is computed. For example, the entropy for the attribute <60 is calculated as follows:

$$\begin{aligned} Entropy_{<60} = & -\left(\frac{36}{61} \times log_2 \times \frac{36}{61} + \frac{21}{61} \times log_2 \times \frac{21}{61}\right. \\ & \left. + \frac{4}{61} \times log_2 \times \frac{4}{61}\right) \\ & - \left(0.590 \times (-0.760) + 0.344(-1.538)\right. \\ & \left. + 0.065\,(-3.930)\right) = 0.416 \end{aligned}$$

Table 9. Gain calculation rules

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy | Gain |
|---|---|---|---|---|---|---|
| | 71 | 36 | 21 | 14 | 1.478 | |
| Rainfall Data | | | | | | 0.416 |
| < 60 | 61 | 36 | 21 | 4 | 1.236 | |
| 60 - 89 | 8 | 0 | 0 | 8 | 0 | |
| ≥ 90 | 2 | 0 | 0 | 2 | 0 | |
| Water Discharge | | | | | | 0.363 |
| < 13 | 62 | 36 | 21 | 5 | 1.277 | |
| 13 - 19 | 2 | 0 | 0 | 2 | 0 | |
| ≥ 20 | 7 | 0 | 0 | 7 | 0 | |
| Runoff Coefficient | | | | | | 0.008 |
| 0.17 | 33 | 18 | 10 | 5 | 1.411 | |
| 0.18 | 38 | 18 | 11 | 9 | 1.521 | |
| River Cross Section | | | | | | 0.243 |
| 15 | 37 | 18 | 18 | 1 | 1.152 | |
| 20 | 34 | 18 | 3 | 13 | 1.325 | |

Table 10. Entropy and gain on runoff coefficient 0.18

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy | Gain |
|---|---|---|---|---|---|---|
| | 38 | 18 | 11 | 9 | 1.521 | |
| Rainfall Data | | | | | | 0.508 |
| < 60 | 31 | 18 | 11 | 2 | 1.241 | |
| 60 - 89 | 6 | 0 | 0 | 6 | 0 | |
| ≥ 90 | 1 | 0 | 0 | 1 | 0 | |
| Water Discharge | | | | | | 0.327 |
| < 13 | 33 | 18 | 11 | 4 | 1.374 | |
| 13 - 19 | 1 | 0 | 0 | 1 | 0 | |
| ≥ 20 | 4 | 0 | 0 | 4 | 0 | |
| River Cross Section | | | | | | 2.241 |
| 15 | 18 | 9 | 9 | 0 | 0 | |
| 20 | 20 | 9 | 2 | 9 | 1.369 | |

After calculating each attribute's total entropy and entropy, the next step is to compute the gain using the previously defined equation. For example, the gain for Rainfall Data is determined as follows:

$$Gain_{RainfallData} = 1.478 - \left(\frac{61}{71} \times 1.236\right) + \left(\frac{8}{71} \times 0\right) + \left(\frac{2}{71} \times 0\right) = 0.416$$

Subsequently, the gain for each attribute is calculated, and the results are summarized in Table 9.

Table 11. Run off coefficient data 0.18

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 2 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 3 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 4 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 5 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 6 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 7 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 8 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 9 | < 60 | < 13 | 0.18 | 15 | Not Flood |
| 10 | 60 - 89 | ≥ 20 | 0.18 | 20 | Not Flood |
| ... | ... | ... | ... | ... | ... |
| 34 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 35 | 60 - 89 | ≥ 20 | 0.18 | 20 | Flood |
| 36 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 37 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 38 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |

The gain values for each attribute were computed to guide the construction of the decision tree. The attribute with the highest information gain—Runoff Coefficient—was selected as the root node, as illustrated in Fig. 2. This selection process is repeated iteratively for each subset until either a homogeneous subset is achieved or a predefined stopping criterion is encountered. After establishing the root node, Node 1 is constructed by generating a new decision table based on the Runoff Coefficient attribute, with an initial value of 0.18, as shown in Table 10. Entropy and information gain values are then recalculated using the data in this table, applying the same formulas previously described. The results of these computations are presented in Table 11. Notably, the Runoff Coefficient is excluded from subsequent calculations, having already been utilized as the root node due to its maximum gain. The attribute with the following highest gain—River Cross-section—is then selected to extend the decision tree. A summary of all computed gain values is provided for reference.

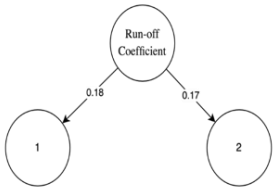


Figure 2. Initial node decision tree

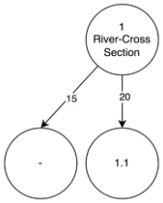


Figure 3. Decision tree node 1

In Fig. 3, when the River Cross-section attribute has a value of 15, the classification becomes determinate, with

entropy reaching 0, indicating complete uniformity in the class distribution. Next, Node 1.1 is computed by generating a new decision table with a River Cross-section of 20, as shown in Table 12.

Table 12. River cross section data 20

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | 15 | < 13 | 0.18 | 20 | Not Flood |
| 2 | 10 | < 13 | 0.18 | 20 | Not Flood |
| 3 | 15 | < 13 | 0.18 | 20 | Not Flood |
| 4 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 5 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 6 | 10 | < 13 | 0.18 | 20 | Not Flood |
| 7 | 15 | < 13 | 0.18 | 20 | Not Flood |
| 8 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 9 | 10 | < 13 | 0.18 | 20 | Not Flood |
| 10 | 20 | < 13 | 0.18 | 20 | Flood Prone |
| ... | ... | ... | ... | ... | ... |
| 16 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 17 | 35 | ≥ 20 | 0.18 | 20 | Flood |
| 18 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 19 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |
| 20 | ≥ 90 | ≥ 20 | 0.18 | 20 | Flood |

The entropy and information gain values are recalculated based on the data in Table 12 using the previously established formula. These recalculations are summarized in Table 13, which shows that the River Cross-section attribute is excluded from further analysis, having already yielded the highest information gain. Consequently, the attribute with the subsequent highest gain—Rainfall—is selected to extend the decision tree. Fig. 4 depicts Node 1.1, demonstrating that Rainfall values of 60–89 and ≥ 90 are classified as flood.

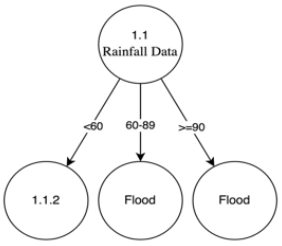


Figure 4. Decision tree node 1.1

Table 13. Entropy and gain of river cross section 20

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy | Gain |
|---|---|---|---|---|---|---|
| | 20 | 9 | 2 | 9 | 1.369 | |
| Rainfall Data | | | | | | 0.59 |
| < 60 | 13 | 9 | 2 | 2 | 1.198 | |
| 60 - 89 | 6 | 0 | 0 | 6 | 0 | |
| ≥ 90 | 1 | 0 | 0 | 1 | 0 | |
| Water Discharge | | | | | | 0.365 |
| < 13 | 15 | 9 | 2 | 4 | 1.338 | |
| 13 - 19 | 1 | 0 | 0 | 1 | 0 | |
| ≥ 20 | 4 | 0 | 0 | 4 | 0 | |

Table 14. Entropy and rainfall data <60

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy | Gain |
|---|---|---|---|---|---|---|
| | 15 | 9 | 2 | 4 | 1.338 | |
| Water Discharge | | | | | | 0 |
| < 13 | 15 | 9 | 2 | 4 | 1.338 | |
| 13 - 19 | 1 | 0 | 0 | 1 | 0 | |
| ≥ 20 | 4 | 0 | 0 | 4 | 0 | |

Table 15. Entropy and gain of runoff coefficient 0.17

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy | Gain |
|---|---|---|---|---|---|---|
| | 33 | 18 | 10 | 5 | 1.411 | |
| Rainfall Data | | | | | | 0.292 |
| < 60 | 30 | 18 | 10 | 2 | 1.231 | |
| 60 - 89 | 2 | 0 | 0 | 2 | 0 | |
| ≥ 90 | 1 | 0 | 0 | 1 | 0 | |
| Water Discharge | | | | | | 0.423 |
| < 13 | 29 | 18 | 10 | 1 | 1.124 | |
| 13 - 19 | 1 | 0 | 0 | 1 | 0 | |
| ≥ 20 | 3 | 0 | 0 | 3 | 0 | |
| River Cross Section | | | | | | 1.919 |
| 15 | 19 | 9 | 9 | 1 | 0 | |
| 20 | 14 | 9 | 1 | 4 | 1.198 | |

Afterward, Node 1.1.2 is computed by constructing a new decision table for Rainfall values below 60, as shown in Table 14. Entropy and gain values are recalculated using the same formula, and the results are provided in Table 15. Rainfall was already utilized in a prior split, which is excluded from further consideration. Table 15 indicates that for Water Discharge values <13, the majority class is "not flooded". For discharge values between 13–19 and ≥ 20, entropy again reaches 0, signifying homogeneous classification and rendering further subdivision unnecessary. Fig. 5 demonstrates Node 1.1.2.

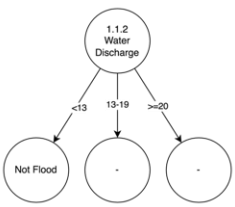


Figure 5. Decision tree node 1.1.2

Table 16. Rainfall data <60

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 2 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 3 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 4 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 5 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 6 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 7 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 8 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 9 | < 60 | < 13 | 0.18 | 20 | Not Flood |
| 10 | < 60 | < 13 | 0.18 | 20 | Flood Prone |
| 11 | < 60 | < 13 | 0.18 | 20 | Flood Prone |
| 12 | 60 - 89 | < 13 | 0.18 | 20 | Flood |
| 13 | 60 - 89 | < 13 | 0.18 | 20 | Flood |
| 14 | 60 - 89 | < 13 | 0.18 | 20 | Flood |
| 15 | 60 - 89 | < 13 | 0.18 | 20 | Flood |

Having completed the decision tree for a Runoff Coefficient of 0.18, the next branch to examine corresponds to a value of 0.17, designated as Node 2. Table 16 presents this case's entropy and gain values, which are calculated using the same methodology. The results, summarized in Table 17, confirm that the Runoff Coefficient is excluded from further splitting, having already served as the root attribute due to its maximal gain. The following most informative attribute—River Cross-section—is selected to continue tree construction.

Table 17. Runoff coefficient data 0.17

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | <60 | <13 | 0.17 | 20 | Not Flood |
| 2 | <60 | <13 | 0.17 | 20 | Not Flood |
| 3 | <60 | <13 | 0.17 | 20 | Not Flood |
| 4 | <60 | <13 | 0.17 | 20 | Not Flood |
| 5 | <60 | <13 | 0.17 | 20 | Not Flood |
| 6 | <60 | <13 | 0.17 | 20 | Not Flood |
| 7 | <60 | <13 | 0.17 | 20 | Not Flood |
| 8 | <60 | <13 | 0.17 | 20 | Not Flood |
| 9 | <60 | <13 | 0.17 | 20 | Not Flood |
| 10 | <60 | <13 | 0.17 | 20 | Flood Prone |
| 11 | <60 | ≥20 | 0.17 | 20 | Flood |
| 12 | 60-89 | ≥20 | 0.17 | 20 | Flood |
| 13 | 60-89 | ≥20 | 0.17 | 20 | Flood |
| 14 | <60 | ≥20 | 0.17 | 20 | Flood |

Table 18. River cross-section data for 20 nodes 2

| # | $X_1$ | $X_2$ | $C$ | $X_3$ | *Label* |
|---|---|---|---|---|---|
| 1 | <60 | <13 | 0.17 | 15 | Not Flood |
| 2 | <60 | <13 | 0.17 | 15 | Not Flood |
| 3 | <60 | <13 | 0.17 | 15 | Not Flood |
| 4 | <60 | <13 | 0.17 | 15 | Not Flood |
| 5 | <60 | <13 | 0.17 | 15 | Not Flood |
| 6 | <60 | <13 | 0.17 | 15 | Not Flood |
| 7 | <60 | <13 | 0.17 | 15 | Not Flood |
| 8 | <60 | <13 | 0.17 | 15 | Not Flood |
| 9 | <60 | <13 | 0.17 | 15 | Not Flood |
| 10 | <60 | <13 | 0.17 | 20 | Not Flood |
| ... | ... | ... | ... | ... | … |
| 29 | ≥90 | <13 | 0.17 | 15 | Flood |
| 30 | <60 | ≥20 | 0.17 | 20 | Flood |
| 31 | 60-89 | ≥20 | 0.17 | 20 | Flood |
| 32 | 60-89 | ≥20 | 0.17 | 20 | Flood |
| 33 | <60 | ≥20 | 0.17 | 20 | Flood |

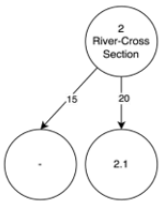


Figure 6. Decision tree node 2

Fig. 6 shows Node 2, where a River Cross-section value of 15 leads to a determinate classification with an entropy of 0, reproducing a completely homogeneous class distribution. To further extend the tree, Node 2.1 is developed by creating a new decision table for instances where the River Cross-section equals 20, as shown in Table 18.

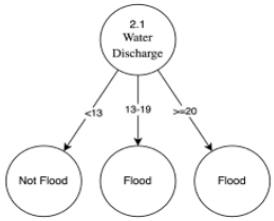


Figure 7. Decision tree node 2.1

The corresponding entropy and information gain values, recalculated using the established formula, are summarized in Table 19. These results reaffirm the exclusion of the River Cross-section attribute from subsequent analysis, as it has already been employed as a splitting criterion. The following highest information gain attribute—Water Discharge—is selected to guide the next decision node, represented in Fig. 7 as Node 2.1. Moreover, the decision tree reveals that all resulting categories exhibit zero entropy, indicating homogeneity and precluding the need for additional nodes. Specifically, Water Discharge values below 13 are classified as "*Not Flooded*", supported by their dominant frequency count of 9. In contrast, discharge values ranging from 13 to 19 and ≥20 are consistently classified as "*Flooded*". The complete structure and final classification results derived from the training dataset are presented in Fig. 8, which encapsulates the entire decision tree developed in this analysis.

Table 19. Entropy and gain of river cross section 15 nodes 2

| | Total Case | *Not Flood* | *Prone to Flood* | *Flood* | Entropy | Gain |
|---|---|---|---|---|---|---|
| | 14 | 9 | 1 | 4 | 1.198 | |
| Rainfall Data | | | | | | 0.305 |
| < 60 | 12 | 9 | 1 | 2 | 1.041 | |
| 60 - 89 | 2 | 0 | 0 | 2 | 0 | |
| ≥ 90 | 0 | 0 | 0 | 0 | 0 | |
| Water Discharge | | | | | | 1.198 |
| < 13 | 10 | 9 | 1 | 0 | 0 | |
| 13 - 19 | 1 | 0 | 0 | 1 | 0 | |
| ≥ 20 | 3 | 0 | 0 | 3 | 0 | |

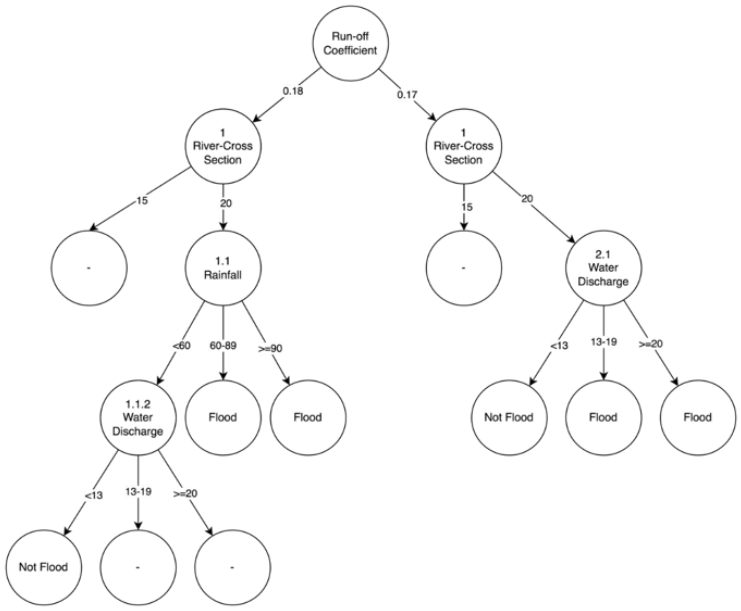


Figure 11. Decision tree

## 4. Results and Discussion

### *4.1. Experimental setup*

The data collection approach entailed obtaining information from two sources: BWS Sulawesi 1 and UPTD BAPELITBANGDA Manado, covering the period from 2017 to 2021. The data, formatted as .csv, was subsequently imported into the database. It constituted the basis for producing training data for the Decision Tree C4.5 algorithm. Before initiating training, the gathered data underwent a cleansing procedure to classify it for further predictive applications. After categorization, the data were reintegrated into the database. A total of 71 entries included rainfall, water discharge, runoff coefficient, and river cross-section measurements. The C4.5 Decision Tree method was subsequently integrated into the application, incorporating functionalities for predicting new data, projecting future floods, and presenting the outcomes of these forecasts, as outlined in the implementation section [22].

### *4.2. Implementation and performance evaluation*

This section demonstrates the application of the C4.5 Decision Tree algorithm to forecast flood disasters in Manado City, located in North Sulawesi Province. The application simulation commences with the cleaned data imported from the database, which functions as the training set. The data is tested to assess accuracy, precision, and recall, utilizing several data partitions for training and testing to identify the most effective division for testing objectives. The incoming data is subsequently partitioned and processed using computation. Upon generating and displaying the findings, the C4.5 decision tree algorithm is used to evaluate individual data.

Afterward, the Decision Tree C4.5 algorithm exclusively tests the data and forecasts flood potential using the provided single data test. This is accomplished by examining the flood prediction training data stored in the database. Following this, a series of tests is conducted using the C4.5 decision tree algorithm. Upon completion of the operations, the results of 10 experiments are presented in Table 20.

Table 20. Testing Results of Data Training

| Time | Accuracy 60:40 | Accuracy 70:30 | Accuracy 80:20 | Precision 60:40 | Precision 70:30 | Precision 80:20 | Recall 60:40 | Recall 70:30 | Recall 80:20 |
|---|---|---|---|---|---|---|---|---|---|
| 1st | 86.91% | 87.26% | 87.46% | 0.81 | 0.81 | 0.82 | 0.59 | 0.60 | 0.60 |
| 2nd | 86.72% | 87.08% | 87.97% | 0.78 | 0.80 | 0.81 | 0.59 | 0.59 | 0.60 |
| 3rd | 85.95% | 86.09% | 86.42% | 0.72 | 0.72 | 0.70 | 0.57 | 0.56 | 0.55 |
| 4th | 85.51% | 85.87% | 86.98% | 0.71 | 0.72 | 0.73 | 0.59 | 0.59 | 0.59 |
| 5th | 85.82% | 85.25% | 86.54% | 0.73 | 0.71 | 0.86 | 0.57 | 0.56 | 0.56 |
| 6th | 85.90% | 85.83% | 86.21% | 0.73 | 0.74 | 0.71 | 0.57 | 0.57 | 0.56 |
| 7th | 86.72% | 87.08% | 87.17% | 0.73 | 0.80 | 0.81 | 0.59 | 0.59 | 0.60 |
| 8th | 85.72% | 85.35% | 86.55% | 0.70 | 0.70 | 0.82 | 0.57 | 0.56 | 0.56 |
| 9th | 86.80% | 87.24% | 87.55% | 0.71 | 0.81 | 0.85 | 0.59 | 0.60 | 0.60 |
| 10th | 85.20% | 85.70% | 87.80% | 0.70 | 0.73 | 0.87 | 0.57 | 0.60 | 0.65 |
| *Average* | *86.13%* | *86.28%* | *87.065%* | *0.73* | *0.74* | *0.82* | *0.58* | *0.59* | *0.60* |

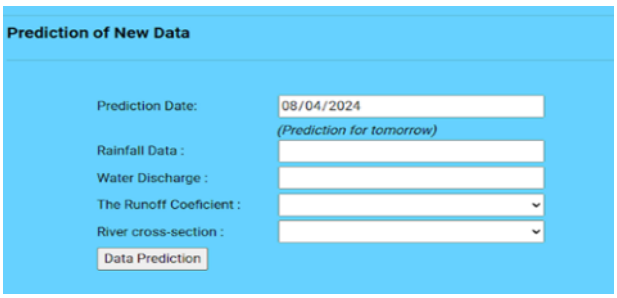


Figure 9. Prediction of new data

Additionally, the system requires users to input specific parameters for flood prediction or forecasting for the following day. These inputs include rainfall, water discharge, runoff coefficient, and river cross-section measurements. These input variables and the prediction flow are illustrated in Fig. 9, which provides a visual representation of the data-driven prediction process. Upon submission, the application displays a pop-up interface showing the C4.5 Decision Tree structure alongside the predicted outcome for the newly entered single test case, as presented in Fig. 1. Within this framework, the River Cross-section values correspond to specific locations: 8 for the Bailang River, 15 for the Malalayang River, 20 for the Tikala–Sawangan and Tikala–Paal 4 Rivers, and 25 for the Tondano River. Moreover, the application includes functionality to simulate future scenarios that may lead to flooding. It can generate flood-free scenario predictions based on new user-defined data. For instance, a test input consisting of a rainfall value of 2, a water discharge of 2, a runoff coefficient of 0.17, and a river cross-section of 8 (Bailang River) is used to simulate a minimal-risk condition. The results of this specific test case are illustrated in Fig. 10.

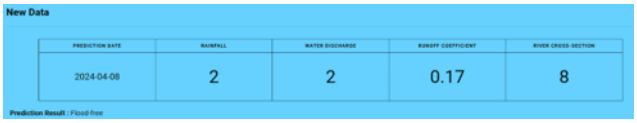


Figure 10. New-flood prediction

Afterward, the application presents the prediction results processed by the C4.5 Decision Tree algorithm using the newly submitted flood forecasting data. These outputs are visually summarized in Fig. 11.

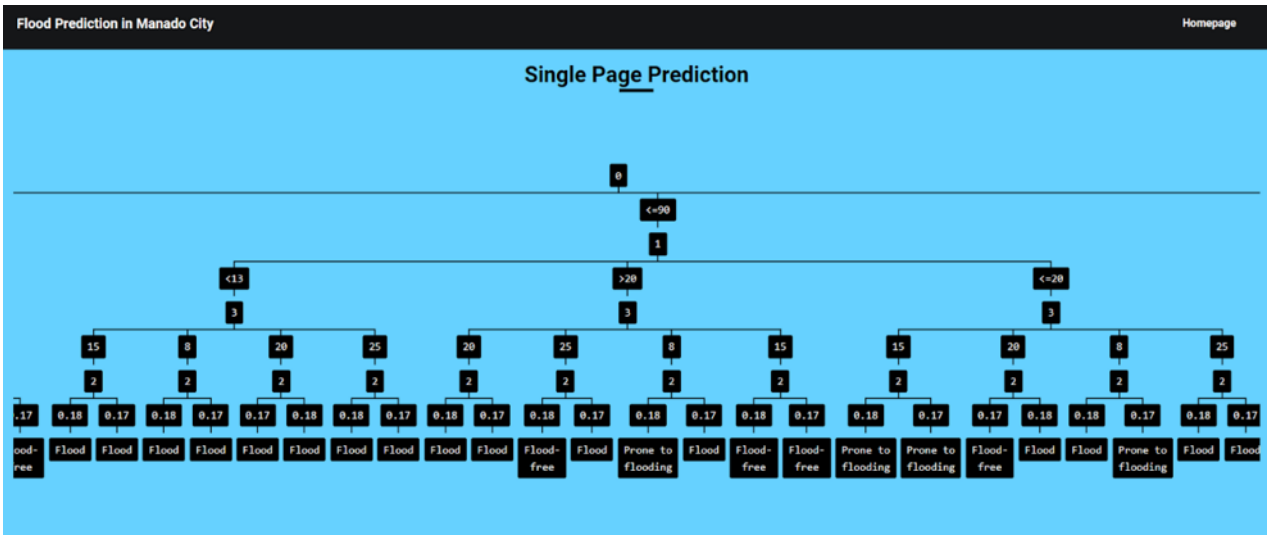


Figure 11. Decision tree C4.5 results of new flood-prediction

The decision tree construction process begins with rainfall as the initial root node, followed sequentially by water discharge and the runoff coefficient as branching nodes. Based on the predictive analysis, the final classification at the terminal node indicates the flood risk category, distinguishing among flood, prone to flooding, or flood-free conditions.

## 5. Conclusion

The flood prediction application developed for Manado City using the C4.5 Decision Tree algorithm has demonstrated strong predictive performance. The model, built on historical river data and complemented by river basin mapping, provides interpretable decision-making support, particularly for the Sulawesi River Regional Office 1 implementing flood prevention strategies. Based on 10 experimental trials, the system consistently achieved high accuracy rates—86.13% with a 60:40 split, 86.28% with a 70:30 split, and 87.07% with an 80:20 split. The highest precision (82%) was observed at the 80:20 ratio, while the best recall was observed with the 70:30 division. Looking forward, integrating Internet of Things (IoT) technology—such as real-time sensors for rainfall, water levels, and flow rates—holds significant promise. This advancement could enable the system to dynamically respond to evolving environmental conditions, further enhancing its accuracy, timeliness, and effectiveness in predicting and mitigating flood events.

## References

[1] N. K. Sutrisnawati, “Dampak Bencana Alam bagi Sektor Pariwisata di Bali,” *J. Ilm. Hosp. Manag.*, vol. 9, no. 1, pp. 57–66, 2018, doi: 10.22334/jihm.v9i1.144.

[2] A. M. Situngkir, “Analisis Data Curah Hujan sebagai Penyebab Banjir di Gedong Tataan Lampung,” *Inov. Pembang. J.*

*Kelitbangan*, vol. 10, no. 01, pp. 99–112, 2022, doi: 10.35450/jip.v10i01.277.

[3] R. Riyandari, "'Water Front City' Mitigasi Bencana Banjir di Kelurahan Dendengan Luar, Kota Manado," *J. Sains dan Teknol. Mitigasi Bencana*, vol. 13, no. 1, pp. 95–108, 2019, doi: 10.29122/jstmb.v13i1.3361.

[4] R. Maiyuriska, "Penerapan Jaringan Syaraf Tiruan dengan Algoritma Backpropagation dalam Memprediksi Hasil Panen Gabah Padi," *J. Inform. Ekon. Bisnis*, vol. 4, no. 1, pp. 28–33, 2022, doi: 10.37034/infeb.v4i1.115.

[5] P. Mauliana, "Prediksi Banjir Sungai Citarum dengan Logika Fuzzy Hasil Algoritma Particle Swarm Optimization," *J. Inform. Fak. Tek. dan Inform. Univ. Bina Sarana Inform.*, vol. 3, no. 2, 2016, doi: 10.31294/ji.v3i2.807.

[6] N. H. Purnomo, B. Pamungkas, and C. Juliane, "Penerapan Algoritma C4.5 untuk Klasifikasi Tren Pelanggaran Kendaraan Angkutan Barang dengan Metode CRISP-DM," *J. Media Inform. Budidarma*, vol. 7, no. 1, 2023, doi: 10.30865/mib.v7i1.5247.

[7] Y. Mendrofa, "Implementasi Algoritma C4.5 untuk Memprediksi Tingkat Kerusakan Akibat Banjir (Studi Kasus: BPBD Prov. Sumut)," *Pelita Inform. Inf. dan Inform.*, vol. 7, no. 4, pp. 584–592, 2019.

[8] A. Primajaya, B. N. Sari, and A. Khusaeri, "Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4.5 Studi Kasus Provinsi Kalimantan Barat," *JEPIN J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, pp. 188–192, 2020, doi: 10.26418/jp.v6i2.37834.

[9] A. K. B. Ginting, M. S. Lydia, and E. M. Zamzami, "Reduksi Atribut Menggunakan Chi Square untuk Optimasi Kinerja Metode Decision Tree C4.5," *JEPIN J. Edukasi dan Penelit. Inform.*, vol. 9, no. 1, pp. 44–49, 2023, doi: 10.26418/jp.v9i1.56542.

[10] K. M. Sudar and P. Deepalakshmi, "A Two Level Security Mechanism to Detect a DDoS Flooding Attack in Software-Defined Networks using Entropy-Based and C4.5 Technique," *J. High Speed Networks*, vol. 26, no. 1, pp. 55–76, 2020, doi: 10.3233/JHS-200630.

[11] S. Križanić, "Educational Data Mining using Cluster Analysis and Decision Tree Technique: A Case Study," *Int. J. Eng. Bus. Manag.*, vol. 12, 2020, doi: 10.1177/1847979020908675.

[12] A. Angdresey and J.-Q. Ooi, "Analyzing Multimodal Sensory Signals Using Unsupervised Machine Learning," in *International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Bandung, 2023, pp. 301–306. doi: 10.1109/IC3INA60834.2023.10285741.

[13] Y.-F. Li, L.-Z. Guo, and Z.-H. Zhou, "Towards Safe Weakly Supervised Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 334–346, 2021, doi: 10.1109/TPAMI.2019.2922396.

[14] A. Angdresey, L. Sitanayah, and T. V. N. Kairupan, "A Soil Monitoring and Recommendation System for Ornamental Plants," in *6th International Conference on New Media Studies (CONMEDIA)*, Tangerang: IEEE, 2021, pp. 40–45. doi: 10.1109/CONMEDIA53104.2021.9617203.

[15] V. G. Costa and C. E. Pedreira, "Recent Advances in Decision Trees: An Updated Survey," *Artif. Intell. Rev.*, vol. 56, pp. 4765–4800, 2023, doi: 10.1007/s10462-022-10275-5.

[16] J. Xu, "Systematic Analysis and Application Prospect of Decision Tree," *Highlights Sci. Eng. Technol.*, vol. 71, pp. 163–170, 2023, doi: 10.54097/hset.v71i.12687.

[17] A. I. Weinberg and M. Last, "Selecting a Representative Decision Tree from an Ensemble of Decision-Tree Models for Fast Big Data Classification," *J. Big Data*, vol. 6, p. 23, 2019, doi: 10.1186/s40537-019-0186-3.

[18] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, pp. 235–240, 1994, doi: 10.1007/BF00993309.

[19] T. Damrongsakmethee and V.-E. Neagoe, "C4.5 Decision Tree Enhanced with AdaBoost Versus Multilayer Perceptron for Credit Scoring Modeling," in *Computational Statistics and Mathematical Modeling Methods in Intelligent Systems*, 2019, pp. 216–226. doi: 10.1007/978-3-030-31362-3_22.

[20] Sulika, R. Kusumawati, and Y. M. Arif, "Classification of Students' Academic Performance Using Neural Network and C4.5 Model," *Int. J. Adv. Data Inf. Syst.*, vol. 5, no. 1, pp. 29–38, 2024, doi: 10.59395/ijadis.v5i1.1311.

[21] J. A. Bartholomai and H. B. Frieboes, "Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques," in *IEEE International Symposium on Signal Processing and Information Technology*, IEEE, 2018, pp. 632–637. doi: 10.1109/ISSPIT.2018.8642753.

[22] I. M. Hayder *et al.*, "An Intelligent Early Flood Forecasting and Prediction Leveraging Machine and Deep Learning Algorithms with Advanced Alert System," *Processes*, vol. 11, no. 2, p. 481, 2023, doi: 10.3390/pr11020481.

[23] A. Angdresey, L. Sitanayah, and T. M. I. Sumajow, "A Real-Time Water Quality and Quantity Monitoring System for Aquarium," in *International Conference on Computer, Control, Informatics and Its Applications*, New York: ACM, 2021. doi: 10.1145/3489088.3489090.

[24] P. F. Opit, I. Y. Kairupan, and F. M. Rusuh, "A MILP Model for Water Level Sensor Placement with Multi-Sensor and Multi-Disaster Areas," *J. Tek. Ind.*, vol. 22, no. 2, pp. 185–195, 2021, doi: 10.22219/JTIUMM.Vol22.No2.185-195.