ISSN 2723-6471

Identifying Key Factors Causing Flooding Using Machine Learning

Adie Wahyudi Oktavia Gama[1,*], Monalisa Dennatan[2], I Gusti Ngurah Putu Dharmayasa[3], Me Me Maw[4], I Putu Sugiana[5], Irma Suryanti[6]

[1,2]*Department of Information Technology, Faculty of Engineering and Informatics, Universitas Pendidikan Nasional, Bali, Indonesia*

[3,6]*Department of Civil Engineering, Faculty of Engineering and Informatics, Universitas Pendidikan Nasional, Denpasar, Bali, Indonesia*

[4]*Department of Civil and Environmental Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand*

[5]*Environmental Research Center, Universitas Udayana, Denpasar, Bali, Indonesia*

(Received: August 25, 2024; Revised: October 6, 2024; Accepted: November 21, 2024; Available online: December 27, 2024)

Abstract

The impact of flooding extends beyond physical and infrastructural damage, affecting social, economic, and environmental dimensions. This study aims to identify the key factors influencing flooding by developing a decision tree model. The research method applies the C4.5 algorithm to build a decision tree model using flood factors such as rainfall, soil type, elevation, land use, and distance from rivers. The model is then applied to 57 past flood data events to determine key contributors to flooding in Denpasar City, Bali, Indonesia. The analysis showed that land elevation is the most influential factor, with areas below 28 meters above sea level having a 71% likelihood of being flood vulnerability. Additionally, the model reveals unknown patterns contributing to flood vulnerability among the factors considered. These insights give a deeper understanding of how these factors combine to affect flood vulnerability. The model's effectiveness was evaluated using a confusion matrix, resulting in an accuracy rate of 90%, a precision rate of 100%, a sensitivity rate of 90%, a specificity rate of 100%, and a F1 Score rate of 94%, demonstrating its strong predictive power in identifying areas at risk of flood vulnerability. Although this study is limited by the availability of data, the focus on Denpasar City, and the potential omission of other relevant attributes, it advances flood risk assessment by applying machine learning to provide practical insights that could enhance flood management strategies, with potential applications to other urban areas facing similar risks.

Keywords: Flood Disaster, Key Factors, Machine Learning, Decision Tree, C4.5 Algorithm

1. Introduction

Natural disasters are events that are difficult to avoid and accurately predict, often resulting in casualties, social and environmental damage, property loss, and other negative impacts on communities and the environment [1], [2]. One frequently occurring natural disaster is flooding. Flooding occurs when water overflows and submerges areas that are usually dry. Floods can result from the low permeability of soil combined with surface water runoff that exceeds the capacity of the drainage system [3], [4], [5]. The impact of flooding extends beyond physical damage to property and infrastructure, affecting social, economic, and environmental aspects as well. Such events can disrupt human lives, lead to loss of livelihoods, and create social instability. Consequently, the risk of flooding can no longer be ignored, and effective management is urgently required to enhance community resilience against such threats [6], [7], [8].

Almost every region faces flooding issues due to environmental changes or climate change that alter rainfall patterns, including Indonesia, which is in a tropical area with high rainfall intensity, especially in Denpasar City. According to the Bali Regional Disaster Management Agency (BPBD), over the past ten years (2011–2021), Denpasar City has ranked among the top five in terms of flood disasters [9]. During this period, information technology has advanced significantly, providing substantial opportunities to optimize disaster risk management. However, many areas still face significant gaps in utilizing this technology to effectively understand and respond to flood risks [10]. The absence of

***Corresponding author: Adie Wahyudi Oktavia Gama (adiewahyudi@undiknas.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i1.463

 ISSN 2723-6471

an integrated system to monitor and present up-to-date information on potential flood risks to the public results in untimely and inefficient emergency responses.

Considering these challenges, the primary focus of this research is to identify the key factors influencing flooding in Denpasar City. A deep understanding of these significant factors can provide a solid foundation for comprehending the patterns and dynamics associated with flood risk [11], [12]. This study will utilize the C4.5 algorithm to develop a model that uncovers patterns then identify key factors among various factors influencing flood, thereby addressing critical knowledge gaps in flood risk mitigation. The C4.5 algorithm is chosen for its capability to analyze complex data patterns and predict events based on relevant variables [13], [14], [15]. It is particularly advantageous in flood risk assessment due to its high classification accuracy, ease of interpretation, flexibility, and simplicity in implementation, making it a practical and easily adoptable solution for stakeholders in flood risk mitigation efforts in Bali [16], [17], [18]. For the implementation of decision tree visualization, tools such as R can be used. R is a highly flexible and popular programming language and statistical development environment among data researchers. It offers a variety of packages and statistical functions, including those supporting the implementation of the C4.5 algorithm. R is effective for data management and adheres to high standards in data analysis [19].

The aim of this study is to identify the key factors contributing to flooding using a decision tree model. This model will reveal previously unknown patterns and relationships among factors influencing floods, and it will help identify which factors are the most significant. To achieve this goal, the author begins by collecting secondary data on Denpasar City from various government websites. This data will then be compiled and prepared for analysis. Once prepared, the C4.5 classification method will be applied, which involves developing a decision tree by partitioning the data based on the values of the splitting attributes. The resulting model will be evaluated using a confusion matrix. By applying the C4.5 algorithm, this study seeks to uncover previously unknown patterns and dependencies between the attributes and flood occurrences. Consequently, the application of the C4.5 classification method is expected to provide a reliable and effective solution for identifying the key attributes influencing flood risk in Denpasar, Bali, Indonesia. Through this research, the author aims to integrate the findings into an early warning system or provide a decision support tool for authorities.

2. Research Method

2.1. Research Location

First, This research was conducted in Denpasar City, Bali, Indonesia, located at coordinates 8° 35' 56" - 8° 42' 01" S and 115° 10' 23" - 115° 16' 27" E [20]. Denpasar City covers an area of 127.78 km² and is situated on lowland terrain with an elevation ranging from 0 to 75 meters above sea level [21]. Several rivers pass through Denpasar, including the Mati River, Badung River, and Ayung River [22]. With its flat contour and rivers crossing the area, Denpasar City has the potential to frequently experience flooding. The distribution points of flood samples can be seen in figure 1.

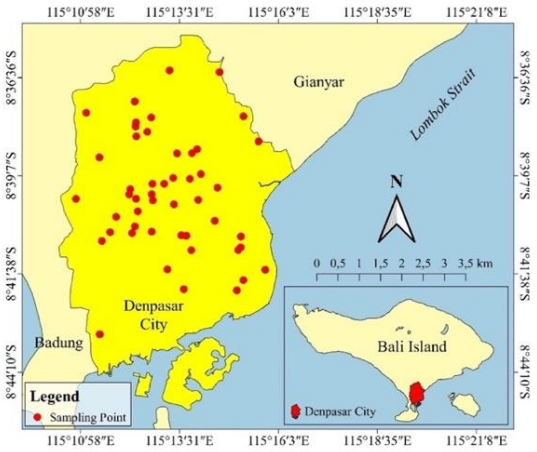

Figure 1. Map of Denpasar City and sampling points for flood events

2.2. Data Collection

The data collection phase begins with gathering data from various sources. Table 1 details the sources and formats of the collected data. Flood events, flood potential maps, soil types, land cover, and river lengths were obtained from the Denpasar City Government. Rainfall data was acquired from NASA POWER [23], a responsive web mapping application that provides meteorological data subsets, and elevation data was gathered using Google Earth [24].

Table 1. Sources And Formats of Collected Data

No	Data Type	Source	Format
1	Flood Event Data	Denpasar Safe City	Point map
2	Rainfall Data	Nasa Power	.csv
3	Soil Type Data	Satu Data Denpasar	.shp
4	Elevation Data	Google Earth	Numeric
5	Land Use Data	Satu Data Denpasar	.shp
6	Distance from River	Satu Data Denpasar	.shp
7	Flood-Vulnerable Area Data	Denpasar City Government Data Center	.pdf

2.3. Data Analysis

In this stage, the data will be filtered to include only the attributes relevant to the research needs. Data cleaning is necessary to remove empty, incomplete, or unclear data, which will be separated from the data to be processed. Subsequently, data integration is carried out to combine multiple datasets into a single database, as data collection may involve multiple databases from various sources. This stage involves transforming the data into a format suitable for processing with the C4.5 algorithm by converting it into categorical data. The categorical transformation for the data attributes is outlined by table 2 below:

Table 2. Categorical conversion for attributes

No	Data Type	Class	Category
1	Rainfall	0,5 – 20 mm/day	Light
		20 – 50 mm/day	Medium
		50 – 100 mm/day	Heavy
2	Soil Type	Latosol	Low
		Regosol	High
3	Elevation	≤ 25 meters above sea level	Low
		26 – 50 meters above sea level	Medium
		≥ 51 meters above sea level	High
4	Land Use	Forest	Low
		Field/Garden	Medium
		Residential	High
5	Distance from River	≤ 100 m	Near
		101 – 250 m	Moderate
		≥ 251 m	Far

2.4. Decision Tree Calculation

In this stage, the total entropy, attribute entropy, and information gain for each attribute are calculated. Once all entropies and information gains for all attributes are obtained, the process continues by calculating the split information, which measures how well the dataset is divided by each attribute. Subsequently, the gain ratio is calculated as the ratio between information gain and split information. After calculating all the gain ratios, the highest gain ratio value is selected to serve as the root node of the decision tree [25], [26], [27]. The data will then be computed using the formulas for Entropy, Information Gain, Split Information, and Gain Ratio as follows:

 ISSN 2723-6471

$$\text{Entropy (S)} = -\sum_{i=1}^{n} \text{Pi} * \text{Log}_2 \text{Pi} \quad (1)$$

Explanation: S = Set of cases; n= Number of values in the variable; Pi = Ratio of the number of samples in class i to the total number of samples in the data set

$$\text{Information Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^{n} \frac{|Si|}{|S|} * \text{Entropy (Si)} \quad (2)$$

Explanation: S = Set of cases; A = Variable; n = Number of attribute partitions A; |Si| = Number of samples for value I; |S| = The number of all data samples

$$\text{Split Information (S, A)} = \sum_{i=1}^{n} \frac{|Si|}{|S|} * \text{Log}_2 \frac{|Si|}{|S|} \quad (3)$$

Explanation: S = Sample space; A = Variable; |Si| = Number of samples for value i

$$\text{Gain Ratio} = \frac{\text{Information Gain (S,A)}}{\text{Split Information (S,A)}} \quad (4)$$

2.5. Model Performance Evaluation

The model evaluation will utilize a confusion matrix to demonstrate the effectiveness of the model on test data. Using functions from the Classification and Regression Training Package, evaluation metrics such as accuracy, precision, sensitivity, specificity and F1 score will be calculated. [28], [29], [30] The process begins with using tools in R Studio, ensuring that all necessary packages for decision tree formation are downloaded beforehand. The dataset containing information about flooding with indicators such as rainfall, soil type, elevation above sea level, population density, and distance from the river is then input and prepared. The dataset will be split into 80% training data and 20% testing data for decision tree modeling and model performance testing. However, for future improvements, we will implement cross-validation techniques, such as k-fold cross-validation, to ensure more robust and consistent performance across different data subsets [31], [32]. Subsequently, the C4.5 decision tree model will be applied to the training data. This process enables the model to identify patterns in the data, including key attributes that influence flood-vulnerable area predictions. In the evaluation stage, the testing data will be used to measure the performance of the decision tree model. To provide a deeper understanding of the model's performance, a confusion matrix with accuracy, precision, and sensitivity metrics will give detailed information about the model's ability to identify flood-vulnerable areas. Clear and comprehensive evaluation results will be produced to present the final outcomes and findings of the model.

Based on the confusion matrix, several evaluation metrics can be calculated to provide further insight into the model's performance, including:

Accuracy represents how often the model makes correct predictions overall. The formula is expressed as:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (5)$$

Precision reflects the proportion of relevant results among all positive predictions made by the model. The formula for precision is:

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

Recall, also known as Sensitivity or True Positive Rate, indicates how well the model can identify actual positive instances.

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

Specificity is a measure of a classification model's ability to correctly identify negative cases.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (8)$$

The F1 Score is the harmonic mean of Precision and Recall, providing a balance between these two metrics. It is particularly useful when both precision and recall are important, especially in datasets with imbalanced class distributions. The F1 Score formula as bellow:

 ISSN 2723-6471

$$\text{F1 Score} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision+Recall}} \quad (9)$$

3. Result and Analysis

3.1. Data Preprocessing

The data obtained through the data collection process is measured and translated into numerical and textual forms, which are then unified during the data preprocessing stage because the data comes from various different sources. After all the data is translated, it goes through the initial stage of data preprocessing, which involves data cleaning techniques such as removing incomplete, duplicate, unclear, and missing data. Out of the 58 flood incident records from the Denpasar Safe City website (https://safecity.denpasarkota.go.id/), one record was deleted due to it being a duplicate, where the incident and location were the same. The small dataset highlights the challenges of collecting comprehensive flood data, including limited historical records. However, It is still appropriate for developing a decision tree model, which can effectively identify patterns even with limited data. The cleaned data, along with the previously translated data, is then integrated into a single table as shown in table 3 for the next processing step.

Table 3. Data attribute

No.	Region Name	Rainfall (mm/day)	Soil Type	Elevation (meters above sea level)	Land Use	Distance from River (meters)	Flood-Vulnerable Area
1	Kecubung, Sumerta Kaja	68	Latosol	29	Residential	97	Not-vulnerable
2	Moh. Yamin, Panjer	52	Latosol	13	Residential	169	Vulnerable
3	Pulau Bali, Sesetan	38	latosol	14	Residential	43	Vulnerable
4	Danau Tandano, Sanur	60	Regosol	21	Residential	99	Vulnerable
5	Bumi Ayu, Sanur	48	Regosol	5	Residential	549	Vulnerable
6	Ayani Utara Gg Merpati, Peguyangan	68	Latosol	40	Residential	185	Not-vulnerable
7	Kebo Iwa, Padangsambian Kaja	76	Latosol	54	Residential	18	Not-vulnerable
8	Kecubung, Sumerta Kaja	57	Latosol	29	Residential	97	Not-vulnerable
9	Kebo Iwa Selatan, Padangsambian Kaja	65	Latosol	40	Residential	135	Not-vulnerable
10	Sesetan	52	Latosol	7	Residential	208	Vulnerable
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	Nagasari, Penatih	54	Regosol	66	Residential	240	Not-vulnerable

3.2. Data Transformation

The data that has been unified into a single table, as shown in table 3, will then be transformed into categories according to table 2 by changing the data into the appropriate format to facilitate the C4.5 calculation process into categorical data. Below is the content of the data transformation illustrated by table 4:

Table 4. Data attributes transformation

No	Rainfall (mm/day)	Soil Type	Elevation (meters above sea level)	Land Use	Distance from River (meters)	Flood-Vulnerable Area
1	Heavy	Low	Medium	High	Near	Not-vulnerable
2	Heavy	Low	Low	High	Moderate	Vulnerable
3	Medium	Low	Low	High	Near	Vulnerable
4	Heavy	High	Low	High	Near	Vulnerable

 ISSN 2723-6471

5	Medium	High	Low	High	Far	Vulnerable
6	Heavy	Low	Medium	High	Moderate	Not-vulnerable
7	Heavy	Low	High	High	Near	Not-vulnerable
8	Heavy	Low	Medium	High	Near	Not-vulnerable
9	Heavy	Low	Medium	High	Moderate	Not-vulnerable
10	Heavy	Low	Low	High	Moderate	Vulnerable
⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	Heavy	High	High	High	Moderate	Not-vulnerable

3.3. Decision Tree Calculation

3.3.1. Determination of Node 1

Based on the calculation above, determining node 1 or root node is selected from the highest gain ratio value of each attribute. The Entropy, Information Gain, Split Information, and Gain Ratio of each attribute is as follows in table 5.

Table 5. Node 1 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Total		57	39	18	0.899743759			
Rainfall	Light	0	0	0	0	0.008087291	0.899743759	0.008988438
	Medium	18	11	7	0.964078765			
	Heavy	39	28	11	0.858230793			
Soil Type	Low	43	27	16	0.952265625	0.036044797	0.804252236	0.044817777
	High	14	12	2	0.591672779			
Elevation	Low	37	36	1	0.179256067	0.605638103	1.262592202	0.479678318
	Medium	13	3	10	0.779349837			
	High	7	0	7	0			
Land Use	Low	0	0	0	0	0.004643199	0.428810965	0.010828078
	Medium	5	4	1	0.721928095			
	High	52	35	17	0.911751759			
Distance from River	Near	28	18	10	0.940285959	0.007960155	1.454401861	0.005473147
	Moderate	20	14	6	0.881290899			
	Far	9	7	2	0.764204507			

In table 5, the results from the calculation of node 1 show that the highest attribute value is land elevation, with a gain ratio of 0.479678318. This makes land elevation the root node and the key attribute as the primary cause of flooding disasters. Land elevation has three attribute values: low, medium, and high. These can be recalculated to obtain branches from node 1. For the case where the attribute value is high, which has a value of 0, the attribute value calculation will not be continued, and it will be directly concluded that high land elevation is not vulnerable to flooding disasters. The low and medium attribute values will be recalculated to determine the next influential attribute branches.

3.3.2. Calculation of Node 1.1

The next step is to calculate node 1.1 or the branch from the low land elevation attribute. The calculation of node 1.1 uses the same formula as the calculation of node 1, except that the total entropy is replaced with the attribute value of low land elevation. Therefore, the Entropy, Information Gain, Split Information, and Gain Ratio for each attribute, with the low attribute value as the primary entropy, are as shown below in table 6:

 ISSN 2723-6471

Table 6. Node 1.1 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Elevation	Low	37	36	1	0.179256067			
Rainfall	Light	0	0	0	0	0.048594801	0.877962001	0.055349549
	Medium	11	10	1	0.439496987			
	Heavy	26	26	0	0			
Soil Type	Low	26	25	1	0.235193382	0.013985042	0.877962001	0.015928983
	High	11	11	0	0			
Land Use	Low	0	0	0	0	0.003344427	0.405977038	0.008237972
	Medium	3	3	0	0			
	High	34	33	1	0.191433255			
Distance from River	Near	17	16	1	0.322756959	0.030962329	1.500154201	0.020639431
	Moderate	13	13	0	0			
	Far	7	7	0	0			

Based on table 6, the calculation results of node 1.1 show that rainfall becomes the branch node from the low land elevation attribute because it has the highest gain ratio with a value of 0.055349549. Rainfall has three attribute values: light, medium, and heavy. The light and heavy attribute values have an entropy value of 0, so the calculation for these attribute values will stop, and it will be concluded that the light attribute value will not be included in the decision tree node due to the absence of cases. On the other hand, the heavy attribute value will indicate that heavy rainfall is vulnerable to causing flooding disasters. The medium attribute value will be recalculated to determine the next influential attribute branches.

3.3.3. Calculation of Node 2.1

The next step is to calculate node 2.1 or the branch from the medium land elevation attribute. The calculation of node 2.1 uses the same formula as the calculation of node 1.1, except that the low land elevation attribute value is replaced with the medium land elevation attribute value. Therefore, the Entropy, Information Gain, Split Information, and Gain Ratio for each attribute, with the medium attribute value as the primary entropy, are as follows in table 7:

Table 7. Node 2.1 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Elevation	Moderate	13	3	10	0.779349837			
Rainfall	Light	0	0	0	0	0.000661141	0.89049164	0.000742445
	Medium	4	1	3	0.811278124			
	Heavy	9	2	7	0.764204507			
Soil Type	Low	12	2	10	0.650022422	0.17932914	0.391243564	0.458356781
	High	1	1	0	0			
Land Use	Low	0	0	0	0	0.04670193	0.619382195	0.075400828
	Medium	2	1	1	1			
	High	11	2	9	0.684038436			
Distance from River	Near	8	2	6	0.811278124	0.030477722	1.238901257	0.024600607
	Moderate	4	1	3	0.811278124			
	Far	1	0	1	0			

The calculation results of node 2.1, as depicted in table 7, show that soil type is the branch node derived from the medium land elevation attribute, as it has the highest gain ratio with a value of 0.458356781. Soil type has two attribute

 ISSN 2723-6471

values: low and high. The high attribute value has an entropy value of 0, so the calculation for this attribute value will stop, and it will be concluded that the high soil type value is vulnerable to flooding disasters. The low attribute value will be recalculated to determine the next influential attribute branches.

3.3.4. Calculation of Node 1.2

The next step is to calculate node 1.2 or the branch from the medium rainfall attribute. The calculation of node 1.2 uses the same formula as the calculation of node 1.1, except that the low land elevation attribute value is replaced with the medium rainfall attribute value. Therefore, the Entropy, Information Gain, Split Information, and Gain Ratio for each attribute, with the medium rainfall attribute value as the primary entropy, are as follows in table 8:

Table 8. Node 1.2 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Rainfall	Medium	11	10	1	0.439496987			
Soil Type	Low	9	8	1	0.503258335	0.027740168	0.684038436	0.040553522
	High	2	2	0	0			
Land Use	Low	0	0	0	0	0.027740168	0.684038436	0.040553522
	Medium	2	2	0	0			
	High	9	8	1	0.503258335			
Distance from River	Near	7	6	1	0.591672779	0.062977946	1.309296668	0.048100593
	Moderate	2	2	0	0			
	Far	2	2	0	0			

The calculation results for node 1.2, as shown in Table 8, indicate that the water distance, as a branch of node 1.1, is derived from the attribute of moderate rainfall due to having the highest gain ratio, which is 0.048100593. The water distance attribute has three values: near, somewhat near, and far. The attributes somewhat near and far have an entropy value of 0, indicating that calculations for these attributes will cease, concluding that these water distances are vulnerable to flooding. The attribute value near will be recalculated to determine the subsequent influential attribute.

3.3.5. Calculation of Node 2.2

The next step is to calculate node 2.2, a branch from the attribute of moderate rainfall. The calculation for node 2.2 uses the same formula as node 2.1, except that the attribute value for moderate land elevation is replaced with the attribute value for low soil type. Therefore, the Entropy, Information Gain, Split Information, and Gain Ratio for each attribute, with low soil type as the primary entropy, are as follows in table 9:

Table 9. Node 2.2 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Soil Type	Low	12	2	10	0.650022422			
Rainfall	Light	0	0	0	0	0.01722	0.91829583	0.018752
	Medium	4	1	3	0.811278124			
	Heavy	8	1	7	0.543564443			
Land Use	Low	0	0	0	0	0.022987	0.41381685	0.055549
	Medium	1	0	1	0			
	High	11	2	9	0.684038436			
Distance from River	Near	7	1	6	0.591672779	0.034454	1.28067213	0.026903
	Moderate	4	1	3	0.811278124			
	Far	1	0	1	0			

Table 9 illustrates the calculation results for node 2.2 indicate that land use, as a branch of node 2.1 from the low soil type attribute, has the highest gain ratio with a value of 0.055549. The land use attribute has three values: low, moderate, and high. The attributes low and moderate have an entropy value of 0, indicating that calculations for these attributes will cease. The result shows that the low attribute value will not be included in the decision tree node because there are no cases. For the moderate attribute value, the result indicates that moderate land use is not vulnerable to flooding. The high attribute value will be recalculated to determine the subsequent influential attribute.

3.3.6. Calculation of Node 1.3

The next step is to calculate node 1.3, a branch from the attribute of near water distance. The calculation for node 1.3 uses the same formula as node 1.2, except that the attribute value for moderate rainfall is replaced with the attribute value for near water distance. Therefore, the Entropy, Information Gain, Split Information, and Gain Ratio for each attribute, with near water distance as the primary entropy, are as follows in table 10:

Table 10. Node 1.3 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Distance from River	Near	7	6	1	0.591672779			
Soil Type	Low	7	6	1	0.591672779	0	0	0
	High	0	0	0	0			
Land Use	Low	0	0	0	0			
	Medium	1	1	0	0	0.034510703	0.591672779	0.058327346
	High	6	5	1	0.650022422			

The calculation results on table 10 for node 1.3 indicate that land use, as a branch of node 1.2 from the near water distance attribute, has the highest gain ratio with a value of 0.058327346. The land use attribute has three values: low, moderate, and high. The attributes low and moderate have an entropy value of 0, indicating that calculations for these attributes will cease. The result shows that the low attribute value will not be included in the decision tree node because there are no cases. For the moderate attribute value, the result indicates that moderate land use is vulnerable to flooding, whereas the high attribute value can be recalculated. However, since there are no more attributes to calculate, the branch 1.3 will stop, concluding that the high attribute value is vulnerable to flooding.

3.3.7. Calculation of node 2.3

The next step is to calculate node 2.3, a branch from the attribute of near water distance. The calculation for node 2.3 uses the same formula as node 2.2, except that the attribute value for low soil type is replaced with the attribute value for high land use. Therefore, the Entropy, Information Gain, Split Information, and Gain Ratio for each attribute, with high land use as the primary entropy, are as follows in table 11:

Table 11. Node 2.3 calculation results

Attribute	Value	Amount	Vulnerable	Not-vulnerable	Entropy	Info Gain	Split Info	Gain Ratio
Land Use	High	11	2	9	0.684038436			
Rainfall	Light	0	0	0	0			
	Medium	3	1	2	0.918295834	0.038274522	0.845350937	0.045276489
	Heavy	8	1	7	0.543564443			
Distance from River	Near	6	1	5	0.650022422			
	Moderate	4	1	3	0.811278124	0.034470524	1.322179346	0.02607099
	Far	1	0	1	0			

 ISSN 2723-6471

The calculation results for node 2.3 indicate that rainfall, as a branch of node 2.2 from the high land use attribute, has the highest gain ratio with a value of 0.045276489. The rainfall attribute has three values: light, moderate, and heavy. The light attribute value has an entropy value of 0, indicating that calculations for this attribute will cease. The result shows that the light attribute value will not be included in the decision tree node because there are no cases. The moderate and heavy attribute values can be recalculated; however, since there are no more attributes to calculate, Branch 2.3 will stop, concluding that the moderate and heavy attribute values are not vulnerable to flooding, as shown below in figure 2.

	Rainfall	Soil_Type	Elevation	Land_Use	Distance_from_River	Flood_Vulnerable_Area
1	68	Latosol	29	Residential	97	Not-vulnerable
2	52	Latosol	13	Residential	169	Vulnerable
3	38	latosol	14	Residential	43	Vulnerable
4	60	Regosol	21	Residential	99	Vulnerable
5	48	Regosol	5	Residential	549	Vulnerable
6	68	Latosol	40	Residential	185	Not-vulnerable
7	76	Latosol	54	Residential	18	Not-vulnerable
8	57	Latosol	29	Residential	97	Not-vulnerable
9	65	Latosol	40	Residential	135	Not-vulnerable
10	52	Latosol	7	Residential	208	Vulnerable
11	59	Latosol	22	Residential	47	Vulnerable
12	62	Latosol	19	Residential	148	Vulnerable
13	28	latosol	56	Residential	106	Not-vulnerable
14	42	Latosol	20	Residential	14	Vulnerable
15	45	Latosol	19	Residential	62	Vulnerable
16	38	Latosol	38	Field/Garden	25	Not-vulnerable
17	67	Latosol	17	Residential	30	Vulnerable
18	50	Latosol	22	Residential	47	Vulnerable
19	78	Regosol	33	Field/Garden	25	Vulnerable
20	81	Latosol	24	Residential	311	Vulnerable
21	47	Latosol	18	Residential	44	Not-vulnerable

Figure 2. Data analysis using R Studio

To begin the process in R Studio after obtaining all the decision tree calculation results, first to do is download all the necessary library packages used for dataset splitting, decision tree formation, and the confusion matrix. Next, the downloaded packages need to be called back to activate them. The attribute data table in table 3 should then be input into R Studio for processing. The data used in this research is in .xlsx or Excel format. After the data has been input, it is important to check the structure of the input data using the str function. If the code shows chr or character, the data structure needs to be converted to factors to be processed by R Studio, as shown in figure 3 below:

```
> str(data)
tibble [57 × 6] (S3: tbl_df/tbl/data.frame)
 $ Rainfall             : num [1:57] 68 52 38 60 48 68 76 57 65 52 ...
 $ Soil_Type            : chr [1:57] "Latosol" "Latosol" "latosol" "Regosol" ...
 $ Elevation            : num [1:57] 29 13 14 21 5 40 54 29 40 7 ...
 $ Land_Use             : chr [1:57] "Residential" "Residential" "Residential" "Residential" ...
 $ Distance_from_River  : num [1:57] 97 169 43 99 549 185 18 97 135 208 ...
 $ Flood_Vulnerable_Area: chr [1:57] "Not-vulnerable" "Vulnerable" "Vulnerable" "Vulnerable" ...
> data$Soil_Type=as.factor(data$Soil_Type)
> data$Land_Use=as.factor(data$Land_Use)
> data$Flood_Vulnerable_Area=as.factor(data$Flood_Vulnerable_Area)
```

Figure 3. Flood attribute data structure before and after

Once the data structure is appropriate, proceed to split the dataset into 80% training data and 20% testing data. The creation of the decision tree starts with building the decision tree model, followed by displaying the created model. The visualization results of the decision tree for the flood attribute data, showing the most influential key attributes in flood disasters, are as follows in figure 4:

 ISSN 2723-6471

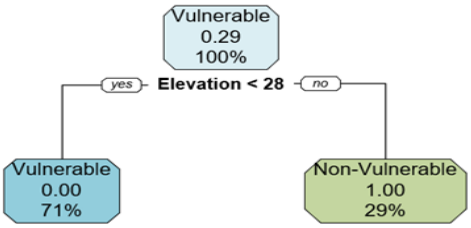

Figure 4. Visualization decision tree using R Studio

Based on the visualization of the decision tree, as shown in Figure 4, the results show that if the land elevation is less than 28 meters above sea level, it is considered vulnerable to flooding with a probability of 71% and a Not-vulnerable probability of 0.00 or 0%. Conversely, if the land elevation is greater than 28 meters above sea level, it is considered not vulnerable to flooding with a probability of 29% and a Not-vulnerable probability of 1.00 or 100%.

The complete visualization of the decision tree, up to the last branch, visualized through Microsoft Visio, is as shown in figure 5:

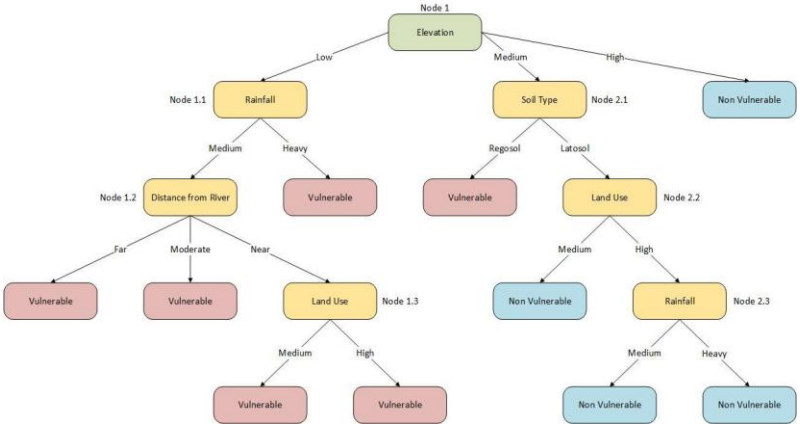

Figure 5. Visualization all decision tree branch using MS Visio.

Based on the results of the decision tree visualization in figure 5, the flow results generated from several branches are shown in table 12:

Table 12. Decision tree flow results

State	Result
1	If the elevation is low, rainfall is medium, and distance from river is far, then it is considered vulnerable
2	If the elevation is low, rainfall is medium, and distance from river is moderate, then it is considered vulnerable
3	If the elevation is low, rainfall is medium, distance from river is near, and land use is medium then it is considered vulnerable
4	If the elevation is low, rainfall is medium, distance from river is near, and land use is high, then it is considered vulnerable
5	If the elevation is medium, and the soil type is regosol, it is considered vulnerable
6	If the elevation is medium, the soil type is latosol, and the land use is medium then it is not considered vulnerable
7	If the elevation is medium, the soil type is latosol, the land use is high, and the rainfall is medium then it is not considered vulnerable
8	If the elevation is medium, the soil type is latosol, the land use is high, and the rainfall is heavy, then it is not considered vulnerable
9	If the elevation is high, then it is not considered vulnerable

The results show that flood vulnerability is influenced by multiple factors, including elevation, rainfall, distance from rivers, land use, and soil type. The C4.5 algorithm successfully developed a decision tree that uncovered previously

 ISSN 2723-6471

unknown patterns contributing to flood vulnerability, as shown in figure 5 and described as table 12. Low elevation consistently appears as a major risk factor, as these areas are more prone to water accumulation, especially when drainage systems are insufficient. Even with moderate rainfall, poor drainage in low-lying areas can increase the likelihood of flooding. The findings also suggest that distance from rivers does not fully eliminate vulnerability, highlighting the importance of localized rainfall and the role of drainage infrastructure in managing flood risks.

Land use also plays a significant role. In low-elevation areas, higher land use increases vulnerability by limiting water absorption and causing more surface runoff. In contrast, in medium-elevation areas, vulnerability decreases with higher land use, suggesting that well-planned urbanization, such as stormwater management systems, can help mitigate flood risks. The interaction between soil type and elevation offers further insights. In medium-elevation areas, regosol soil increases vulnerability due to its poor water retention, while latosol soil reduces flood risk, even with higher land use and heavy rainfall, due to its superior water-holding capacity. High-elevation areas are consistently categorized as not vulnerable, likely because water drains naturally from these areas, and vegetation cover promotes absorption.

These findings demonstrate the power of the C4.5 algorithm in identifying complex patterns and interactions among environmental and human factors that contribute to flood vulnerability. The results underscore the importance of integrated flood management strategies, combining improved infrastructure—such as enhanced drainage systems—with environmental conservation, particularly in low-lying urban areas.

3.4. Model Performance Evaluation

Evaluating the model using a confusion matrix requires the caret library package, which should have been previously downloaded and activated for evaluating the decision tree model. The dataset previously divided into training data and testing data will be used accordingly: the training data to model the decision tree and the testing data to evaluate the decision tree model. Additionally, code is provided to display the most important attributes in constructing the decision tree model, as shown in figure 6.

```
Confusion Matrix and Statistics

          Reference
Prediction     Vulnerable  Not-Vulnerable
  Vulnerable            9               0
  Not-Vulnerable        1               1

               Accuracy : 0.9091
                 95% CI : (0.5872, 0.9977)
    No Information Rate : 0.9091
    P-Value [Acc > NIR] : 0.736

                  Kappa : 0.6207

 Mcnemar's Test P-Value : 1.000

            Sensitivity : 0.9000
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 0.5000
             Prevalence : 0.9091
         Detection Rate : 0.8182
   Detection Prevalence : 0.8182
      Balanced Accuracy : 0.9500

       'Positive' Class : Vulnerable

> #The most important attributes
> model_tree$variable.importance
Elevation  Rainfall
18.488889  2.844444
```

Figure 6. Confusion matrix and the important attribute

Based on the confusion matrix results in figure 6, the classification using the confusion matrix shows that the 'Positive' class represents flood-vulnerable areas. The True Positive (reference: flood-vulnerable, prediction: flood-vulnerable) indicates that the model correctly predicted 9 data points as flood-vulnerable. The False Positive (reference: not flood-vulnerable, prediction: flood-vulnerable) indicates that there were 0 data points where the model incorrectly predicted non-flood-vulnerable areas as flood-vulnerable. The False Negative (reference: flood-vulnerable, prediction: not flood-vulnerable) indicates that the model incorrectly predicted 1 flood-vulnerable data point as not flood-vulnerable. The True Negative (reference: not flood-vulnerable, prediction: not flood-vulnerable) indicates that the model correctly predicted 1 data point as not flood-vulnerable. The accuracy of the model, calculated as (True Positive + True Negative) / Total Data = (9 + 1) / 11, is 0.9091 (90%). The precision, calculated as True Positive / (True Positive + False Positive)

= 9 / (9 + 0), is 1.0000 (100%). The sensitivity (recall), calculated as True Positive / (True Positive + False Negative) = 9 / (9 + 1), is 0.9000 (90%). The specificity, calculated as True Negative / (True Negative + False Positive) = 1 / (1 + 0), is 1.0000 (100%). The F1 Score, calculated as 2 * (Precision x Recall) / (Precision + Recall) = 2 * (1.0000 * 0.9000) / (1.0000 + 0.9000), is 0.9473 (95%). The prediction error rate, represented by the Kappa statistic, is 0.6207 (62%). The most important attributes in building the decision tree model are land elevation and rainfall.

3.5. Discussion

The implementation of the C4.5 algorithm to analyze flooding factors in Denpasar City has provided significant new insights, setting this research apart from previous studies [16]. This model systematically examines their interactions and specific thresholds in a localized context using real data from past flood events, ultimately identifying the key factors that contribute to flooding. The main contribution of this study is the identification of key factors that contribute to flooding, with particular emphasis on land elevation as the most critical factor. The analysis shows that low-elevation areas are especially vulnerable to flooding, associated with other factors such as rainfall, distance from rivers, and land use. In addition, the C4.5 algorithm successfully developed a decision tree that uncovered previously unknown patterns contributing to flood vulnerability, as shown in Figure 5 and described as table 12. These finding challenges traditional flood management methods that often overlook how different factors work together, emphasizing that elevation plays a major role in flood risk. While earlier studies may have looked at these factors separately, this research demonstrates the significant impact of elevation on flood vulnerability, offering new ways to enhance flood management strategies [17], [18].

The evaluation results demonstrate that the model performs well in identifying flood-vulnerable areas, with a high level of accuracy. It successfully classified most flood-vulnerable areas and demonstrated perfect precision, meaning all positive predictions were correct. However, the model did miss one flood-vulnerable area. The prediction error rate suggests moderate agreement between the model's predictions and actual outcomes. Key factors contributing to the model's effectiveness were land elevation and rainfall, which were crucial in determining flood vulnerability. Overall, the model exhibited strong performance with few mistakes.

In summary, this research advances our understanding of flood risk in Denpasar City by highlighting complex interactions between flood factors that have not been fully addressed before. By offering practical insights for improving flood management strategies, this study lays the foundation for more effective predictive modeling and risk reduction in urban areas vulnerability to flooding. However, this model is currently limited to a case study in Denpasar City, and adjustments to topography, climate or other geographic factors may be necessary for application in other areas. In addition, a drawback of using the C4.5 algorithm with small datasets for identifying flood factors is its tendency to overfit, which can reduce accuracy on new data.

4. Conclusion

This study successfully identified the key factors that contribute to flooding in Denpasar City by developing a decision tree model using the C4.5 algorithm. The analysis showed that land elevation is the most influential factor, with areas below 28 meters above sea level having a 71% likelihood of being flood vulnerability, while areas above 28 meters are much less likely to be affected, with only a 29% chance of flooding. The study also reveals unknown patterns contributing to flood vulnerability among the factors considered. These insights give a deeper understanding of how these factors combine to affect flood vulnerability. The decision tree model performed well, showing high accuracy, precision, and sensitivity when evaluated using a confusion matrix, which indicates it can effectively predict flood vulnerability areas. To improve the model's accuracy even further, future research should focus on gathering more data on flood incidents and including additional relevant factors, like slope gradient and population density. Although this study did not employ Geographic Information Systems (GIS) directly, spatial data such as elevation and proximity to rivers were integrated into the decision tree model. Future research could benefit from using GIS to complement machine learning models, enhancing the spatial analysis and providing more detailed flood risk mapping. Additionally, future research could compare the C4.5 algorithm with more advanced models such as Random Forest or Gradient Boosting, which may offer improved accuracy and robustness. Moreover, testing the model on external datasets or simulating future flood events would be a valuable next step to assess the model’s robustness and generalizability in

real-world applications. These improvements would provide a more complete understanding of what causes flooding in Denpasar City and help develop better strategies for managing and mitigating flood risks.

5. Declarations

5.1. Author Contributions

Conceptualization: A.W.O.G., M.D., I.G.N.P.D., M.M.M., and I.P.S.; Methodology: A.W.O.G., and M.D; Software: M.D., and I.P.S.; Validation: A.W.O.G., I.G.N.P.D., and M.M.M.; Formal Analysis: A.W.O.G., I.G.N.P.D., M.M.M., and I.S.; Investigation: A.W.O.G.; Resources: M.D., I.S.; Data Curation: M.D., I.P.S., and I.S.; Writing Original Draft Preparation: A.W.O.G., M.D., and I.G.N.P.D.; Writing Review and Editing: A.W.O.G., I.G.N.P.D.; and I.P.S.; Visualization: A.W.O.G., and M.D.; All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] O. I. Lawanson, D. Proverbs, and R. L. Ibrahim, "The impact of flooding on poor communities in Lagos State, Nigeria: The case of the Makoko urban settlement," *J Flood Risk Manag*, vol. 16, no. 1, pp. 1–16, Mar. 2023, doi: 10.1111/jfr3.12838.

[2] Y. S. Song and M. J. Park, "Development of damage prediction formula for natural disasters considering economic indicators," *Sustainability*, vol. 11, no. 3, pp. 1–22, Feb. 2019, doi: 10.3390/su11030868.

[3] H. Basri, S. Syakur, A. Azmeri, and E. Fatimah, "Floods and their problems: Land uses and soil types perspectives," *IOP Conf Ser Earth Environ Sci*, vol. 951, no. 1, pp. 1–9, Jan. 2022, doi: 10.1088/1755-1315/951/1/012111.

[4] B. Feng, Y. Zhang, and R. Bourke, "Urbanization impacts on flood risks based on urban growth data and coupled flood models," *Natural Hazards*, vol. 106, no. 1, pp. 613–627, Mar. 2021, doi: 10.1007/s11069-020-04480-0.

[5] A. Tariq, J. Yan, B. Ghaffar, S. Qin, B. G. Mousa, A. Sharifi, M. E. Huq, and M. Aslam, "Flash flood susceptibility assessment and zonation by integrating analytic hierarchy process and frequency ratio model with diverse spatial data," *Water (Basel)*, vol. 14, no. 19, pp. 1–22, Sep. 2022, doi: 10.3390/w14193069.

[6] L. M. Erfurth, A. S. Hernandez Bark, C. Molenaar, A. L. Aydin, and R. van Dick, "'If worse comes to worst, my neighbors come first': social identity as a collective resilience factor in areas threatened by sea floods," *SN Social Sciences*, vol. 1, no. 11, pp. 1–24, Nov. 2021, doi: 10.1007/s43545-021-00284-6.

[7] Y. He, S. Thies, P. Avner, and J. Rentschler, "Flood impacts on urban transit and accessibility—A case study of Kinshasa," *Transp Res D Transp Environ*, vol. 96, no. 1, pp. 1–20, Jul. 2021, doi: 10.1016/j.trd.2021.102889.

[8] A. Momčilović Petronijević and P. Petronijević, "Floods and their impact on cultural heritage—A case study of southern and eastern serbia," *Sustainability*, vol. 14, no. 22, pp. 1–25, Nov. 2022, doi: 10.3390/su142214680.

[9] Badan Pusat Statistik Provinsi Bali, "The number of villages/subdistricts affected by natural disasters by regency/city in Bali Province." Accessed: Aug. 25, 2024. [Online]. Available: https://bali.bps.go.id/id/statistics-table/3/YmtNd1RGQkhMelpTV213eFVEUjRZVk4wVmtadGR6MDkjMw==/jumlah-desa-kelurahan-yang-mengalami-bencana-alam-menurut-kabupaten-kota-di-provinsi-bali.html?year=2021

[10] L. S. Awah, J. A. Belle, Y. S. Nyam, and I. R. Orimoloye, "A systematic analysis of systems approach and flood risk management research: Trends, gaps, and opportunities," *International Journal of Disaster Risk Science*, vol. 15, no. 1, pp. 45–57, Feb. 2024, doi: 10.1007/s13753-024-00544-y.

[11] R. Handika, R. A. Pratama, I. M. Ihsan, R. P. Adhi, Sabudin, A. Sundari, I. N. Sulistiawan, and Y. W. Nugraha, "Identifying environmental variables in potential flood hazard areas using machine learning approach at Musi Banyuasin Regency, South Sumatra," *IOP Conf Ser Earth Environ Sci*, vol. 1201, no. 1, pp. 1–10, Jun. 2023, doi: 10.1088/1755-1315/1201/1/012037.

[12] J. Liu, K. Wang, S. Lv, X. Fan, and H. He, "Flood risk assessment based on a cloud model in Sichuan Province, China," *Sustainability*, vol. 15, no. 20, pp. 1–19, Oct. 2023, doi: 10.3390/su152014714.

[13] Y. Bai and J. Guo, "English teaching achievement prediction by big data analysis under deep intervention," *Informatica*, vol. 48, no. 9, pp. 75–88, Jun. 2024, doi: 10.31449/inf.v48i9.4977.

[14] P. Wang, "Teaching management data analysis method based on decision tree algorithm," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, pp. 1–17, Jan. 2024, doi: 10.2478/amns-2024-1842.

[15] Y. Zhang, H. Wang, J. Wu, and X. Chen, "A data mining algorithm-based approach to accounting for enterprise operating costs," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, pp. 1–15, Jan. 2024, doi: 10.2478/amns-2024-0232.

[16] Y. Diao, C. Wang, H. Wang, and Y. Liu, "Construction and application of reservoir flood control operation rules using the decision tree algorithm," *Water (Basel)*, vol. 13, no. 24, pp. 1–14, Dec. 2021, doi: 10.3390/w13243654.

[17] Y. Gu, Y. Chen, H. Sun, and J. Liu, "Remote sensing-supported flood forecasting of urbanized watersheds—a case study in Southern China," *Remote Sens (Basel)*, vol. 14, no. 23, pp. 1–17, Dec. 2022, doi: 10.3390/rs14236129.

[18] V.-H. Nhu, P.-T. T. Ngo, T. D. Pham, J. Dou, X. Song, N.-D. Hoang, D. A. Tran, D. P. Cao, İ. B. Aydilek, M. Amiri, R. Costache, P. V. Hoa, and D. T. Bui, "A new hybrid firefly–PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping," *Remote Sens (Basel)*, vol. 12, no. 17, pp. 1–18, Aug. 2020, doi: 10.3390/rs12172688.

[19] W. Widyananda, M. F. E. Purnomo, M. Aswin, P. Mudjirahardjo, and S. H. Pramono, "Application of data mining and imputation algorithms for missing value handling: A study case car evaluation dataset," *Iraqi Journal of Science*, vol. 64, no. 5, pp. 2481–2491, May 2023, doi: 10.24996/ijs.2023.64.5.32.

[20] Badan Pusat Statistik Kabupaten Jembrana, "The geographic location of regencies/cities in Bali Province." Accessed: Aug. 25, 2024. [Online]. Available: https://jembranakab.bps.go.id/id/statistics-table/1/MTM1IzE=/letak-geografis-kabupaten-kota-di-provinsi-bali.html

[21] Badan Pusat Statistik Kota Denpasar, "The area and elevation of Denpasar City above sea level by district." Accessed: Aug. 25, 2024. [Online]. Available: https://denpasarkota.bps.go.id/id/statistics-table/1/MTU5IzE=/luas-wilayah-kota-denpasar-dan-ketinggiannya-dari-permukaan-laut-menurut-kecamatan--2015.html

[22] Badan Pusat Statistik Provinsi Bali, "Names and lengths of rivers by regency/city in Bali Province." Accessed: Aug. 25, 2024. [Online]. Available: https://bali.bps.go.id/id/statistics-table/1/NTEjMQ==/nama-nama-sungai-dan-panjangnya-menurut-kabupaten-kota-di-provinsi-bali.html

[23] NASA POWER | DAV, "NASA Prediction Of Worldwide Energy Resources (POWER) | Data Access Viewer (DAV)." Accessed: Aug. 25, 2024. [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/

[24] I. G. N. P. Dharmayasa, C. A. Simatupang, and D. M. Sinaga, "NASA Power's: an alternative rainfall data resources for hydrology research and planning activities in Bali Island, Indonesia," *Journal of Infrastructure Planning and Engineering (JIPE)*, vol. 1, no. 1, pp. 1–7, Apr. 2022, doi: 10.22225/jipe.1.1.2022.1-7.

[25] T. M. Lakshmi, A. Martin, R. M. Begum, and V. P. Venkatesan, "An analysis on performance of decision tree algorithms using student's qualitative data," *International Journal of Modern Education and Computer Science*, vol. 5, no. 5, pp. 18–27, Jun. 2013, doi: 10.5815/ijmecs.2013.05.03.

[26] U. Ramasamy and S. Sundar, "An illustration of Rheumatoid Arthritis disease using decision tree algorithm," *Informatica*, vol. 46, no. 1, pp. 109–119, Mar. 2022, doi: 10.31449/inf.v46i1.3269.

[27] A. Munther, M. M. Abualhaj, A. Alalousi, and H. A. Fadhil, "A significant features vector for internet traffic classification based on multi-features selection techniques and ranker, voting filters," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 6, pp. 6958–6968, Dec. 2024, doi: 10.11591/ijece.v14i6.pp6958-6968.

[28] C. Savitha and R. Talari, "Mapping cropland extent using sentinel-2 datasets and machine learning algorithms for an agriculture watershed," *Smart Agricultural Technology*, vol. 4, no. 1, pp. 1–10, Aug. 2023, doi: 10.1016/j.atech.2023.100193.

[29] C. Verma, "Machine learning model for applicability of hybrid learning in practical laboratory," *Procedia Comput Sci*, vol. 235, no. 1, pp. 1600–1607, 2024, doi: 10.1016/j.procs.2024.04.151.

[30] L. Zhao, S. Lee, and S.-P. Jeong, "Decision tree application to classification problems with boosting algorithm," *Electronics (Basel)*, vol. 10, no. 16, pp. 1–13, Aug. 2021, doi: 10.3390/electronics10161903.

[31] N. Gupta, H. K. Gupta, R. Srivastava, C. Saxena, and Surjeet, "Design and implementation of artificial neural network classifiers based on hypertuning parameters for breast cancer diagnosis," *Procedia Comput Sci*, vol. 233, no. 1, pp. 929–938, 2024, doi: 10.1016/j.procs.2024.03.282.

[32] S. Syed, K. Khan, M. Khan, R. U. Khan, and A. Aloraini, "Recognition of inscribed cursive Pashtu numeral through optimized deep learning," *PeerJ Comput Sci*, vol. 10, no. 1, pp. 1–18, Jul. 2024, doi: 10.7717/peerj-cs.2124.