Prediction Of Heart Disease Risk Based on Geographic Variables Using Decision Tree

Bagas Aji Satria
Department of Informatics
Universitas Jenderal Achmad Yani
West Java, Cimahi, Indonesia
bagasajisatria18@if.unjani.ac.id

Faiza Renaldi, Irma Santikarama
Department of Information Systems
Universitas Jenderal Achmad Yani
West Java, Cimahi, Indonesia
faiza.renaldi@unjani.ac.id, irma.santikarama@lecture.unjani.ac.id

Abstract

Heart disease is the leading cause of death in most countries. According to 2018 statistics, about 17.9 million people worldwide died from cardiovascular disease. Early prediction and control can help reduce mortality from heart disease using existing health data. The decision tree method accurately builds a computational model that aims to predict. The usefulness of decision trees in health has made them used to predict cancer, diagnose lung disease, diagnose heart disease, etc. More studies have also shown that geographical variables can affect the prevalence of hypertension, one of the causes of heart disease. Although hypertension is closely related to geographic variables and has a relationship with heart disease, there are still not many studies related to heart disease prevention and have not used the geographic location as a variable. This study predicts the risk of heart disease by combining 15 variables of heart disease patient data obtained from Azra Hospital, Indonesia and geographic data in the form of an area's altitude. We have categorized the elevation data from 367 villages in West Java into 2 categories based on the indicators of plains in Indonesia, highlands (>600m) and lowlands (<600m). Based on the test results, we found that the decision tree method's diagnosis had an accuracy of 93,75%, with the highest level of risk being in people living in the lowlands. This shows that the altitude of an area can affects the risk of heart disease. We suggest optimizing the method to improve the accuracy of the prediction results.

Keywords
Prediction, Heart Disease, Geographic Variables, Decision Tree

1. Introduction

Heart disease is the leading cause of death in most countries(Mamatha Alex and Shaji 2019). According to 2018 statistics, about 17.9 million people worldwide died from cardiovascular disease(Iskandar, Hadi, and Alfridsyah 2017)(Dhar et al. 2018). Early prediction and control can help reduce mortality from heart disease by using existing health data and data mining techniques that can help identify whether a person has a disease or not. Healthcare workers can take quick action to find more patients(Purushottam, Saxena, and Sharma 2016)(Priyanka and Ravikumar 2017).

Efforts continue to be made to predict the possibility of this deadly disease. Data mining techniques can be an advantage(Khennou et al. 2019). In this case, predictions using data mining techniques can provide accurate initial conclusions about this disease(Dhar et al. 2018). Hidden patterns of data and existing relationships can be extracted from significant data sources using data mining techniques. It is also used for the initial automatic diagnosis of patients and to take action more quickly so that more patients can get the drug in a shorter period and can save many lives(Sharma, Yadav, and Gupta 2020).

The Decision tree method is one of the popular methods that can analyze the value of risk and the value of the information contained in alternative problem-solving(Rochmawati et al. 2020). The role of decision trees in the medical world is widely used as a decision support tool. The decision tree method accurately builds a computational model that aims to predict(Kohli and Regression 2020). The usefulness of decision trees in healthcare makes them used to predict cancer(Devi and Devi 2016)(Kohli and Regression 2020), diagnose lung disease(Alfatah, Arifudin, and Muslim 2018), diagnose heart disease(Kaur and Singh 2014)(Mamatha Alex and Shaji 2019), etc.

Further research also shows that geographic variables can affect the prevalence of hypertension which is one of the causes of heart disease(Musadir, Hidayaturahmi, and Juwita 2019). The lowlands tend to experience rapid industrialization and modernization, which leads to an unhealthy lifestyle that increases blood pressure. Therefore, high blood pressure is more likely to occur in the lowlands than in the mountains(Sulistyanto and Madyoratri 2020). Although hypertension is closely related to geographic variables and has a relationship with heart disease, there are still not many studies related to heart disease prevention and have not used geographic location variables. In addition, geographical factors affect the type of food a person consumes(David Israel Garrido And Garrido 2018). The type of food affects cholesterol levels in the blood, affecting the risk of heart disease(Aprillia 2020).

This study predicts the risk of heart disease by combining data on heart disease patients with 15 variables obtained from Azra Hospital, Indonesia, and geographic data in the form of the height of an area. The altitude data we use comes from 367 villages in 12 cities in West Java. We divide into two categories based on indicators of plains in Indonesia, namely highlands (more than 600 meters above sea level) and lowlands (less than 600 meters above sea level)(Hasanah 2020). This research has produced a system that can predict a person's risk level for heart disease based on geographical conditions or place of residence using the ID3 decision tree algorithm.

2. Methods

This research was conducted through several stages. The research method in this study is shown in Figure 1.

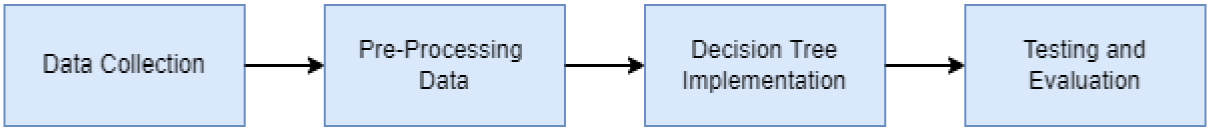


Figure 1. Research methods

The first stage is critical in this research because data mining has original data obtained through data collection or data collection. After the required data has been obtained, the next step is to pre-process the data. This stage is carried out to clean up inappropriate or missing data and convert the data to numeric according to the needs of the method used. Furthermore, implementing the ID3 decision tree technique using python as a tool to create a model, then proceed with testing and evaluation. This needs to be done for the final result of the performance and suitability of the software as product quality. This stage is also the conclusion of the research results that have been carried out from each stage.

2.1 Data Collection

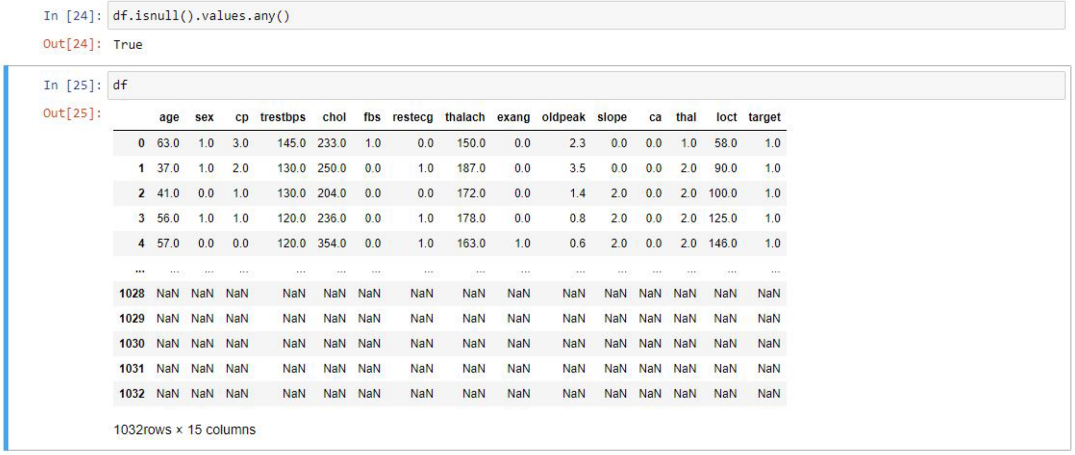
In this study, medical records of heart disease patients were obtained from Azra Hospital. The data used in this study is patient data with the attributes of Age, Sex, cp, rest blood pressure (fbs), Chol, Fbs, Restcg, Thalach, Exang, Oldpeak, Slope, Thal, Ca, and Target. Furthermore, there are regional elevation data from 12 cities in West Java province obtained from the Central Statistics Agency (BPS), West Java, with the location attribute, which is used as an indicator of the height of the patient's residence.

2.2 Pre-Processing Data

At this stage, it functions to change patient data whose data is already available to be processed into data ready to be processed into research objects.(Dhar et al. 2018) Before the data is used in this study, a pre-processing process must be carried out first for changing the patient data that is already available so that it is processed into data that is ready to be processed as an object of research. This study has several stages of pre-processing, namely data cleaning and selection.

2.2.1 Data Cleaning

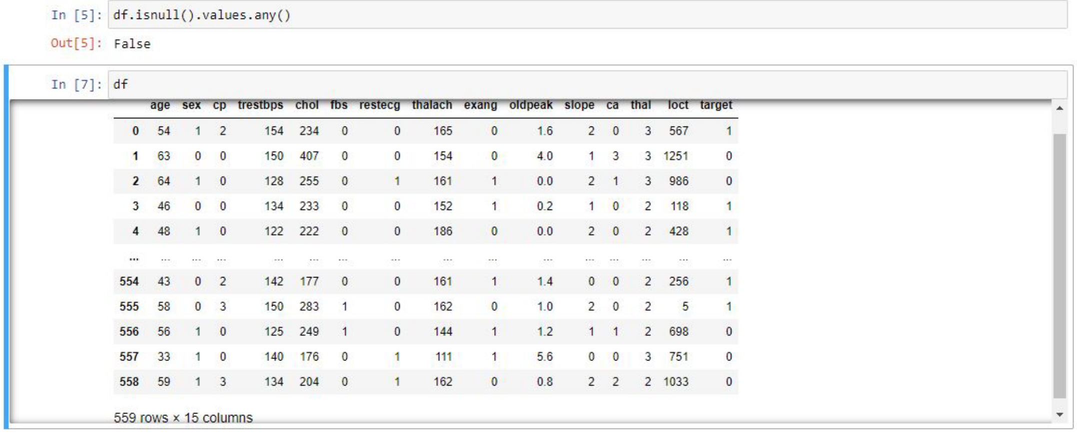
Data cleaning is a process carried out for cleaning patient data that does not have complete or missing medical record data to maintain data quality. the data cleaning process carried out in this study was to correct the data and eliminate data that did not have completeness for large amounts of data as shown in the Figure 2.



```
In [24]: df.isnull().values.any()
Out[24]: True

In [25]: df
Out[25]:
         age  sex   cp  trestbps   chol  fbs  restecg  thalach  exang  oldpeak  slope   ca  thal   loct  target
     0  63.0  1.0  3.0     145.0  233.0  1.0      0.0    150.0    0.0      2.3    0.0  0.0   1.0   58.0     1.0
     1  37.0  1.0  2.0     130.0  250.0  0.0      1.0    187.0    0.0      3.5    0.0  0.0   2.0   90.0     1.0
     2  41.0  0.0  1.0     130.0  204.0  0.0      0.0    172.0    0.0      1.4    2.0  0.0   2.0  100.0     1.0
     3  56.0  1.0  1.0     120.0  236.0  0.0      1.0    178.0    0.0      0.8    2.0  0.0   2.0  125.0     1.0
     4  57.0  0.0  0.0     120.0  354.0  0.0      1.0    163.0    1.0      0.6    2.0  0.0   2.0  146.0     1.0
   ...   ...  ...  ...       ...    ...  ...      ...      ...    ...      ...    ...  ...   ...    ...     ...
  1028   NaN  NaN  NaN       NaN    NaN  NaN      NaN      NaN    NaN      NaN    NaN  NaN   NaN    NaN     NaN
  1029   NaN  NaN  NaN       NaN    NaN  NaN      NaN      NaN    NaN      NaN    NaN  NaN   NaN    NaN     NaN
  1030   NaN  NaN  NaN       NaN    NaN  NaN      NaN      NaN    NaN      NaN    NaN  NaN   NaN    NaN     NaN
  1031   NaN  NaN  NaN       NaN    NaN  NaN      NaN      NaN    NaN      NaN    NaN  NaN   NaN    NaN     NaN
  1032   NaN  NaN  NaN       NaN    NaN  NaN      NaN      NaN    NaN      NaN    NaN  NaN   NaN    NaN     NaN

1032rows × 15 columns
```

Figure 2 Before cleaning data

This data cleaning process resulted in 559 data from 1032 medical records as shown in Figure 3.

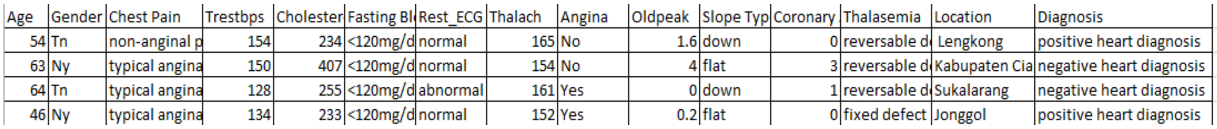


```
In [5]: df.isnull().values.any()
Out[5]: False

In [7]: df
       age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak  slope  ca  thal  loct  target
    0   54    1   2       154   234    0        0      165      0      1.6      2   0     3   567       1
    1   63    0   0       150   407    0        0      154      0      4.0      1   3     3  1251       0
    2   64    1   0       128   255    0        1      161      1      0.0      2   1     3   986       0
    3   46    0   0       134   233    0        0      152      1      0.2      1   0     2   118       1
    4   48    1   0       122   222    0        0      186      0      0.0      2   0     2   428       1
  ...  ...  ...  ...      ...   ...  ...      ...      ...    ...      ...    ...  ...  ...   ...     ...
  554   43    0   2       142   177    0        0      161      1      1.4      0   0     2   256       1
  555   58    0   3       150   283    1        0      162      0      1.0      2   0     2     5       1
  556   56    1   0       125   249    1        0      144      1      1.2      1   1     2   698       0
  557   33    1   0       140   176    0        1      111      1      5.6      0   0     3   751       0
  558   59    1   3       134   204    0        1      162      0      0.8      2   2     2  1033       0

559 rows × 15 columns
```

Figure 3 After cleaning data

2.2.2 Data Selection

Data selection is made to group attributes according to the required information. Data that already has complete information on each attribute(Junaid and Kumar 2020). Data selection is made to group attributes according to the required information. Data selection is made to group attributes according to the required information. Selection data from attributes taken from this process is patient medical record data which includes 14 attributes, namely Age, Sex, cp, resting blood pressure, Chol, Fbs, Restcg, Thalach, Exang, Oldpeak, Slope, Thal, Ca, Target, and altitude data includes the location of the patient as shown in Figure 4.



Age	Gender	Chest Pain	Trestbps	Cholester	Fasting Bl	Rest_ECG	Thalach	Angina	Oldpeak	Slope Typ	Coronary	Thalasemia	Location	Diagnosis
54	Tn	non-anginal p	154	234	<120mg/d	normal	165	No	1.6	down	0	reversable d	Lengkong	positive heart diagnosis
63	Ny	typical angina	150	407	<120mg/d	normal	154	No	4	flat	3	reversable d	Kabupaten Cia	negative heart diagnosis
64	Tn	typical angina	128	255	<120mg/d	abnormal	161	Yes	0	down	1	reversable d	Sukalarang	negative heart diagnosis
46	Ny	typical angina	134	233	<120mg/d	normal	152	Yes	0.2	flat	0	fixed defect	Jonggol	positive heart diagnosis

Figure 4 Attribute Data Selection

2.3 Decision Tree Implementation

The next stage is the decision tree implementation process. At this stage, the data is checked and tries to calculate the accuracy of the data that has passed the pre-processing process using ID3. The steps that will be carried out by the Decision Tree algorithm recursively are as follow(Tyasti, Ispriyanti, and Hoyyi 2015)s:

a. Define one of the features as Root Node in the decision tree
b. Create a branch of each particular feature or attribute value.
c. Repeat the process for each branch until there are no more features to define as Root Nodes or the Decision Tree already has all Leaf Nodes.

To determine a feature as a Root Node, it is based on the highest Information Gain value from the existing features that previously did the Entropy calculation. The following is the formula for calculating the Entropy value and the Gain Information value:

$$H(S) = \sum_{i=1}^{n} p_i \times log_2 p_i \quad (1)$$

Information :

S : Case Collection
n : Number of partitions S
pi : Proportion of Si to S

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

Information:
S : Set
A :Attributes/Features
N : Number of partitions Attribute/Feature A
|Si| : Number of cases on partition i
|S| : Number of cases in

2.4 Testing

This testing stage will be carried out using test data that is entered into the tree model created. The process of testing and making tree models is done using the system.

3. Results and Discussion

The following is data that shows information about the heart disease status of patients at the Azra hospital from January 2017 to September 2021. The data results from data that has gone through preprocessing to produce 559 data. the percentage of positive patients suffering from heart disease can be seen in Table 1

Table 1 heart disease status

Status	Quantity	Percentage
Positive	349	62.43%
Negati	210	37.56%

Table 1 describes the data used. There were 349 cases of heart disease, or equal to 62.43%, and 210 cases without heart disease, or equal to 37.56% of the total 559 data that had been cleaned.

Table 2 heart disease status by altitude

Height	Positive	Negative	Total
More than 600 above sea level	39	179	218
Less than 600 above sea level	310	31	341
Total	**349**	**210**	**559**

Table 2 describes the composition of heart disease patients based on their altitude. At an altitude of 600 meters above sea level, there were 39 positive and 179 negative cases. In comparison, at an altitude of fewer than 600 meters above sea level, there were 310 positive cases and 31 negative cases.

3.1 Construction of the Decision Tree Model

We divide the data into training data and testing data. In this study, 80% of the data were partitioned for training or 447 data, and 20% for test data or 112 data. The following is a calculation to find the entropy and information gain values at the root node using a training sample with the ID3 Algorithm to construct a decision tree. (Table 3)

1. Calculate the proportion of each class

Table 3 Class Proportion

Status	Quantity	Proportion
Positive	273	0.61
Negative	174	0.39
Total(S)	**447**	**1.00**

2. Calculating the class entropy

   In this study, S is the set of positive and negative classes. Positive class with code 1 and negative class with code 2 so that we get:

$$entropy(S) \ = \sum_{k=1}^{n} - Pi \log_2 Pi$$

$$\text{Entropy } (1,2) = \left(-\left(\frac{273}{447}\right).\log_2\left(\frac{273}{447}\right)\right) + \left(-\left(\frac{174}{447}\right).\log_2\left(\frac{174}{447}\right)\right)$$

$$= 0.964$$

3. Calculating the frequency of categories on the altitude attribute based on the class

Table 4 heart disease status by altitude in testing data

Height	Frequency		Total
	Positive	Negative	
More than equal to600 above sea level	25	149	174
Less than 600 above sea level	248	25	273
Total	**273**	**174**	**447**

Table 4 describes the composition of heart disease patients based on the altitude from the total test data used, which is 477. At an altitude of 600 meters above sea level, there were 25 positive and 149 negative cases. For comparison, at an altitude of fewer than 600 meters above sea level, there were 248 positive cases and 25 negative cases.

4. Calculating the entropy of the altitude attribute

$$\text{Entropy (more than equal to 600 1,2)} = \left(-\left(\frac{25}{174}\right).\log_2\left(\frac{25}{174}\right)\right) + \left(-\left(\frac{149}{174}\right).\log_2\left(\frac{149}{174}\right)\right) = 0.5937$$

$$\text{Entropy (Less than 600 1,2)} = \left(-\left(\frac{248}{273}\right).\log_2\left(\frac{248}{273}\right)\right) + \left(-\left(\frac{25}{273}\right).\log_2\left(\frac{25}{273}\right)\right) = 0.4417$$

5. Calculating the information gain

$$\text{Gain(S, Altitude)} = \text{ Entropy(S)} - \sum_{i=1}^{n} \frac{|\text{Si}|}{|\text{S}|} \times \text{Entropy(Si)}$$

Gain (S, Altitude) = $0.964 - \left(\frac{174}{447}\,.\,0.593\right) + \left(\frac{273}{447}\,.\,0.441\right) = 0.4634$

The following results from the calculation to find the entropy and information gain values of all attributes to determine the best features. information gain data is presented in Table 5.

Table 5 Information Gain

No	Attrubut	Gain
1	Age	0.0476
2	Sex	0.0397
3	Cp	0.1748
4	Tresbps	0.0028
5	Chol	0.0007
6	Fbs	0.0028
7	Restecg	0.0191
8	Thalac	0.1082
9	Exang	0.1630
10	Oldpeak	0.0813
11	Slope	0.1033
12	Thal	0.0320
13	Ca	0.2796
14	**loct**	**0.4634**

Based on Table 5, the location attribute is the attribute with the largest information gain value with a value of 0.4634, then the location attribute becomes the root node.

3.2 Tree Model Analysis

Based on the model construction that has been made, we form a tree model that describes the relationship between attributes. The tree model can be seen in Figure 5.

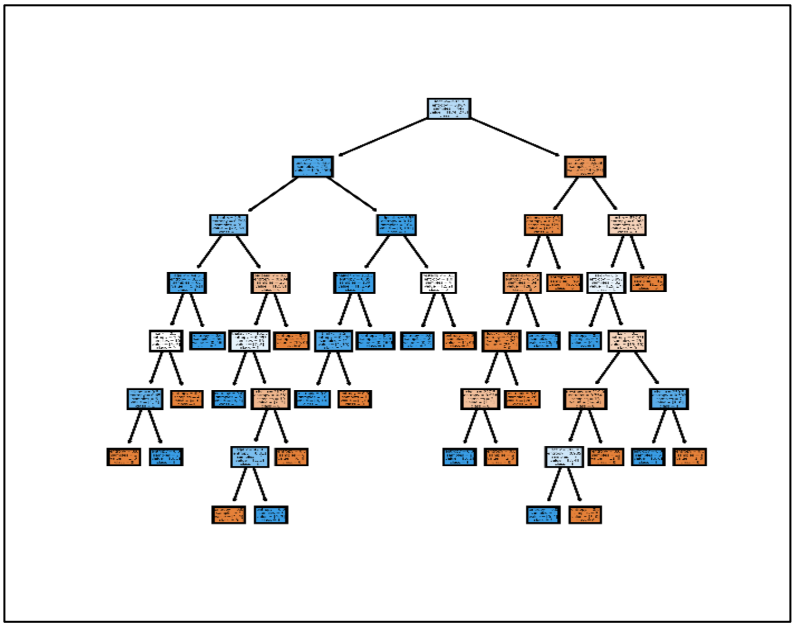


Figure 5 generated tree model results using python

ID3 Algorithm results to identify medical record data of heart disease patients based on geographic variables with attributes Age, Sex, cp, resting blood pressure (fbs), Chol, Fbs, Restcg, Thalach, Exang, Oldpeak, Slope, Thal, Ca, Location and the target obtained from the hospital Azra, the following information can be obtained from the results of making a tree model using the ID3 Algorithm:

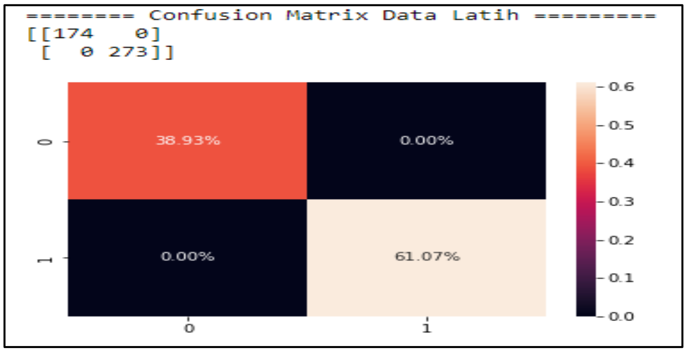
1. In this study, many of the nodes formed were 51 nodes with 26 leaf nodes.
2. The location attribute is selected as the best attribute for the root node based on the largest information gain value.

3.3 Testing

3.3.1. Testing Results based on Training Data

After getting the results from the ID3 algorithm in the form of a decision tree, the next step is to measure the accuracy of the prediction results. The tree model that has been formed is tested by entering the train data into the tree model. The test sample size was 447 cases. The confusion matrix is presented in Figure 6 and Table 6.

Figure 6 confusion matrix data train



```
======== Confusion Matrix Data Latih =========
[[174   0]
 [  0 273]]
```

Table 6 Confusion matrix data train

	Positive	Negative	Total
Positive	174	0	174
Negatif	0	273	273
Total	**174**	**273**	**447**

Based on Table 6, the accuracy value of the ID3 Algorithm in the test sample is as follows:

$$\text{Confusion matrix data train} = \frac{TP + TN}{TP + FP+FN+TN} \text{ x } 100$$

$$= \frac{174 + 273}{174 + 0+0+273} \text{ x } 100$$

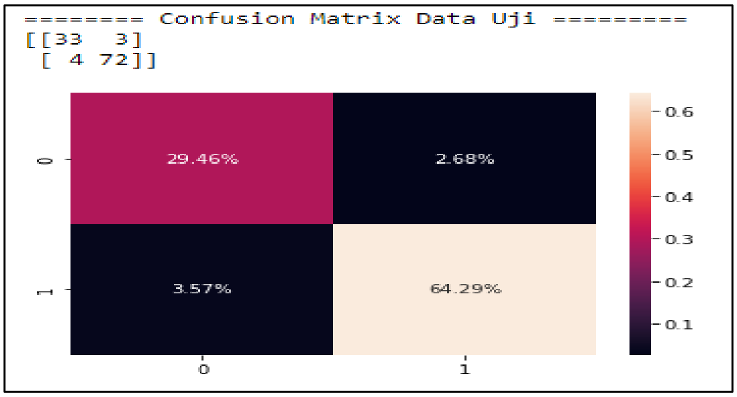
$$= \frac{447}{447} \text{ x } 100$$

$$= 100\%$$

3.3.2. Testing Results Using Testing Data

After getting the results from the ID3 algorithm in the form of a decision tree, the next step is to measure the accuracy of the prediction results. The tree model that has been formed is tested by entering the test data into the tree model. The test sample size was 112 cases. The confusion matrix is presented in Figure 7 and Table 7.

Figure 7 confusion matrix data test



```
======== Confusion Matrix Data Uji =========
[[33  3]
 [ 4 72]]
```

Table 7 Confusion matrix data test

	Positive	**Negative**	**Total**
Positive	33	3	36
Negatif	4	72	76
Total	**37**	**75**	**112**

Based on Table 7, the accuracy value of the ID3 Algorithm in the test sample is as follows:

$$\text{Confusion matrix data test} = \frac{TP + TN}{TP + FP + FN + TN} \text{ x } 100$$

$$= \frac{33 + 72}{33 + 3 + 4 + 72} \text{ x } 100$$

$$= \frac{105}{112} \text{ x } 100$$

$$= 93{,}75\%$$

Based on the results of testing the tree's accuracy level in predicting the data, an accuracy rate of 93.75% is obtained with a prediction error rate of 6,25%, so tree construction results are good enough to predict possible classes in subsequent cases.

3.4 Heart Disease Prediction Software Implementation

We continue our research and produce web-based software to predict heart disease risk. The display of the software that has been made is shown in Figure 6.

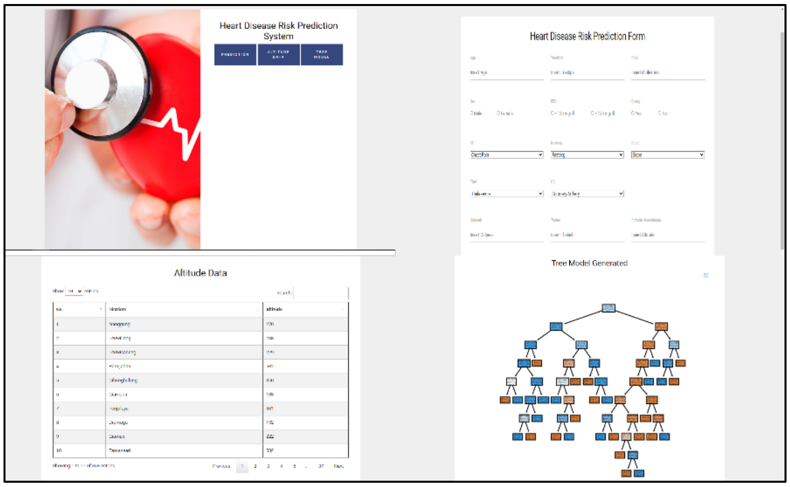

Heart Disease Risk Prediction System
PREDICTION
ALTITUDE DATA
TREE MODEL
Heart Disease Risk Prediction Form
Altitude Data
Tree Model Generated

Figure 6 Software Implementation

There are several menus in the software, including entering test data that can be done through the available forms, viewing the tree model used in the software, and checking altitude location data from 368 districts in 12 cities in West Java which are implemented into the database.

4. Conclusion

This study can conclude that the prediction of the risk of heart disease based on geographic variables using a decision tree gets good results. Based on the test results, it was found that the diagnosis of the decision tree method has an accuracy of 93.75%. This shows that the altitude of an area can affects the risk of heart disease. We suggest optimizing the method and using more geographic variables associated with heart disease to increase the accuracy of the prediction results.

References

Alfatah, Abdul Muis, Riza Arifudin, and Much Aziz Muslim. “Implementation of Decision Tree and Dempster Shafer on Expert System for Lung Disease Diagnosis.” *Scientific Journal of Informatics* 5(1): 57. 2018.

Aprillia, Yessi. “Lifestyle and Diet on the Incidence of Hypertension.” *Scientific Journal of Health Sandi Husada* 12(2): 1044–50. 2020.

David Israel Garrido, And Santiago Moises Garrido. “Cancer Risk Associated with Living at High Altitude in Ecuadorian Population from 2005 to 2014.” : 10. 2018.

Devi, R Delshi Howsalya, and M Indra Devi. “Outlier Detection Algorithm Combined with Decision Tree Classifier for Early Diagnosis of Breast Cancer.” *International Journal of Advanced Engineering TEchnology* 7(2): 93–98. 2016.

Dhar, Sanchayita et al. “A Hybrid Machine Learning Approach for Prediction of Heart Diseases.” *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*: 1–6. 2018.

Hasanah, Fajri Tsaniati. “Characteristics of Indonesia's Land and Water Areas.” *Geography Journal* 20(13): 1–6. 2020. https://www.researchgate.net/publication/345803603.

Iskandar, Iskandar, Abdul Hadi, and Alfridsyah Alfridsyah. “Risk Factors for Coronary Heart Disease in Meuraxa General Hospital Patients, Banda Aceh.” *AcTion: Aceh Nutrition Journal* 2(1): 32. 2017.

Junaid, Mohammed Jawwad Ali, and Rajeev Kumar. “Data Science and Its Application in Heart Disease Prediction.” *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*: 396–4002020..

Kaur, Beant, and Williamjeet Singh. “Review on Heart Disease Prediction System Using Data Mining Techniques.” *International Journal on Recent and Innovation Trends in Computing and Communication* 2(10): 3003–8. 2014.

Khennou, Fadoua, Charif Fahim, Habiba Chaoui, and Nour El Houda Chaoui. “A Machine Learning Approach: Using Predictive Analytics to Identify and Analyze High Risks Patients with Heart Disease.” *International Journal of Machine Learning and Computing* 9(6): 762–67. 2019.

Kohli, Pahulpreet Singh, and A Logistic Regression. “Application of Machine Learning in Disease Prediction.” *2020*

IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020: 1–4. 2020.

Mamatha Alex, P., and Shaicy P. Shaji. "Prediction and Diagnosis of Heart Disease Patients Using Data Mining Technique." *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*: 848–52. 2019.

Musadir, N, H Hidayaturahmi, and J Juwita. "Effect of Altitude on Stroke." *Jurnal Kedokteran Nanggroe* ... (March 2019). http://jknamed.com/jknamed/article/view/39.

Priyanka, N., and Pushpa Ravikumar. "Usage of Data Mining Techniques in Predicting the Heart Diseases - Naïve Bayes & Decision Tree." *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2017*. 2017.

Purushottam, Kanak Saxena, and Richa Sharma. "Efficient Heart Disease Prediction System." *Procedia Computer Science* 85: 962–69. 2016.

Rochmawati, Naim et al. "Covid Symptom Severity Using Decision Tree." *Proceeding - 2020 3rd International Conference on Vocational Education and Electrical Engineering: Strengthening the framework of Society 5.0 through Innovations in Education, Electrical, Engineering and Informatics Engineering, ICVEE 2020*.

Sharma, Vijeta, Shrinkhala Yadav, and Manjari Gupta. "Heart Disease Prediction Using Machine Learning Techniques." *Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020*: 177–81.

Sulistyanto, Benny Arief, and Mukti Lestari Madyoratri. "Relation between Geographical Location and Compliance with Treatment in Hypertensive Patients in Pekalongan Regency." *Jurnal Ilmiah Kesehatan* 13(1): 39–45. 2020.

Tyasti, Avia Enggar, Dwi Ispriyanti, and Abdul Hoyyi. "Iterative Dichotomiser 3 (ID3) Algorithm to Identify Medical Record Data." *Gaussian* 4(Dm): 237–46. 2015.

Biography

Faiza Renaldi, M.Sc., is a lecturer in the department of Information Systems, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani Indonesia. He received his Master of Business Informatics at Universiteit Utrecht, The Netherlands in 2006. His research interests are related to health informatics, information systems/information technology management, e-government, agile project management, and IT entrepreneurship.

Irma Santikarama received a bachelor's degree in Information System from Universitas Kristen Maranatha and Master's degree in Informatics from Institut Teknologi Bandung. Now, she is a lecture of Information Systems Department, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani. Her interest area related to information system, eGoverment, and agile project managemet.

Bagas Aji Satria is a bachelor's degree student at the Universitas Jenderal Achmad Yani, West Java Indonesia, and joined in informatics in 2018. His research interests are in the fields of information systems, agile project managemen, data mining, UI/UX.