unplug Classification Of Dengue Hemorrhagic Disease Using Decision Tree With Id3 Algorithm

by Mochamad Alfan Rosid

Classification Of Dengue Hemorrhagic Disease Using Decision Tree With Id3 Algorithm

View the article online for updates and enhancements.

Classification Of Dengue Hemorrhagic Disease Using Decision Tree With Id3 Algorithm

M A Rosid*, **A S Fitrani, Y Findawati, S Winata, and V A Firmansyah**

Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Jl. Raya Gelam 250 Candi Sidoarjo, Jawa Timur, Indonesia

*alfanrosid@umsida.ac.id

Abstract. Dengue Hemorrhagic Fever (DHF) is a disease caused by the dengue virus which is transmitted by Aedes Aegypty, and Aedes Albopictus (DHF) mosquitoes is a disease caused by the dengue virus transmitted by mosquitoes Aedes Aegypty and Aedes Albopictus. There are three categories of DHF with different symptoms; the first category is called Dengue Fever (DF), the symptoms of DF are characterized by fever, these symptoms are almost the same as symptoms of typhoid or Typhoid Fever (TF). The second category is Dengue Hemorrhagic Fever (DHF) symptoms of DHF characterized by fever, nausea, vomiting, red spots, and nosebleeds. The last category is dengue shock syndrome (DSS) the advanced category of DHF. To determine the level of DHF disease experienced by patients with a background of various symptoms, the DHF disease classification study was conducted using the ID3 algorithm. It is hoped that this study can help doctors diagnose DHF disease. The achievement of predictions from research using the ID3 algorithm can produce an approach in the form of symptoms that affect the level of DHF disease with an accuracy rate of 82%.

1. Intoduction

Dengue hemorrhagic fever, also known as DHF, is a disease caused by dengue virus through the bite of the Aedes Aegypty and Aedes Albopictus mosquitoes. Dengue hemorrhagic fever is characterized by sudden fever and bleeding either in the skin or elsewhere in the body which can cause shock and death [1]. according to Data from the Directorate for Prevention and Control of Zoonotic and Zoonotic Infection Disease, in Indonesia cases of dengue still occur every year[2], the Ministry of Health (Ministry of Health) released data on the development of dengue fever (DHF) from 2014 to January 2019. Director Vector and Zoonotic Infectious Diseases Ministry of Health Siti Nadia Tarmizi claimed to have received reports on DHF the last five years in a row namely since 2014, 2015, 2016, 2017, 2018, until 2019.

The number of DHF sufferers in 2014 was 100,347 people, then in 2015 it was 129,650, then in 2016 it was 204,171[2]. Then in 2017 as many as 68,407, then in 2018 as many as 53,075, and in 2019 as many as 13,683 people, from these data the number of dengue patients who died in 2014 were 907 people, the following year 1,071 people, then in 2016 as many as 1,598 people, and 2017 as many as 493 Then during 2018 there were 344 people and in 2019 (until 29 January 2019) there were 133 people[3]. Frans Yosep Sitepu et al conducted research on the Epidemiology and Entomology of Extraordinary Dengue Fever in South Nias Regency, from the results of his research An outbreak of DF was confirmed in Teluk Dalam Subdistrict, South Nias District, North Sumatra. Not eliminating mosquito breeding sites

routinely and having the habit of hanging worn clothes are the most potential risk factors associated with the outbreak[4]. Hina Gul et al. Made a study to identify the findings of ultrasound in severe dengue infection and the results of this study will assist doctors in developing treatment options strategies[5]. To predict DHF, the diagnosis of symptoms experienced by patients requires a method. ID3 algorithm is one of the decision making methods in data mining techniques[6][7][8] which can be applied to make it easier to predict DHF because the attributes of the dataset will be managed categorically. Based on the background above, we apply Data Mining for Dengue Hemorrhagic Fever Classification Using the Decision Tree Method With ID3 Algorithm.

2. Dangue Hemorrhagic Fever

Dangue hemorrhagic fever or known as DHF is a disease caused by dengue virus through the bite of Aedes Aegypty and Aedes Albopictus mosquitoes. DHF can be classified according to the severity of the patient. Infants and adults who have been infected for the first time can usually be characterized by fever, skin rashes, respiratory symptoms and digestion. The following picture of the phase of dengue fever:
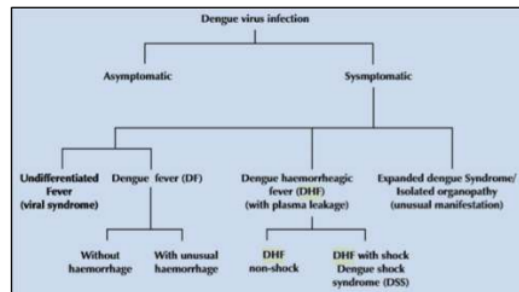

Figure 1. Dengue Fever Phase (World Health Organization , 2011)

At the initial level called Dengue Fever (DF), DF often occurs in children, adolescents and adults. DF can be characterized by acute fever with headaches, muscle aches and rashes on the skin, although DF is not dangerous, in some cases, DF can also cause nosebleeds.

At the second level, it is called Dengue Haemorrhagic Fever (DHF), at the age of under 15 years, it is often infected with DHF. In the initial phase the symptoms are similar to DF symptoms while in the final phase, fever tends to experience hypovolemic shock due to plasma leakage or called Dengue Shock Syndrome (DSS). Before the final phase there are symptoms that can be characterized by nausea, vomiting, abdominal pain, body weakness and nosebleeds. At the final level or DSS, it can involve vital organs such as the heart, kidneys, brain, heart and lungs. Symptoms include unstable breathing, fainting and can cause death[9].

3. ID3 algorithm

ID3 algorithm is a classic algorithm that is often used in data mining technology, which is mainly applied to data mining implementation[8], classification algorithm for data mainly includes decision tree induction classification, rule-based classification methods, Byes, artificial neural networks, K- nearest neighbor , support vector machines and associations rules classification. In this data classification method, the classification process can be concluded as follows: the first is the learning of classification rules, that is, the learning phase of rules is to establish the classification stage; the second stage is the stage of test data, treatment of data classified using class analysis, and treat classification of data to be classified[8][10]. Following are the ID3 algorithm steps in building a decision tree:

1 Select the attribute as the root node from the highest gain value.

2 Create a branch of each variable based on the selected attribute.

3 Calculate each variable until all cases have the same class.

To calculate the gain can be seen from equation 1 as follows:

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

Information:

S	:	case set
A	:	attribute
n	:	Number of attribute attributes A
$|Si|$	:	Number of cases on partitions
$|S|$	:	Number of cases in S

Meanwhile entropy value calculation can be seen from equation 2 below:

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

Information:

S	:	case set
A	:	the features
n	:	number of partitions S
Pi	:	the proportion of Si with respect to S

4. Research methodology

The system design that will be made, based on data mining[11]classification techniques using the Decision Tree method with ID3 algorithm to classify dengue fever (DHF). There are several steps that will be carried out, namely:

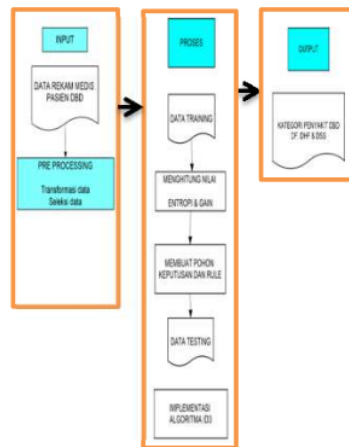a. Classification Diagram ID3

Figure 2. Classification Diagram ID3

The description of the classification diagram is as follows:

1. INPUT Phase, at this stage the data will be pre-processed where the data will be selected and the data will be transformed according to the method to be used.
2. PROCESS Phase, after the input stage is completed, the data will be divided into two namely training data and testing data. at the training data stage, the data will be calculated to find a rule and will be tested with data testing.
3. Output stage, at this stage the results of the data that has been processed include DHF disease categories such as DF, DHF and DSS.

5. Result

Data of DHF patient medical records obtained from the Bhayangkara Pusdik Brimob Watukosek Hospital were 481 data. The data will be tested using the Decision Tree method with ID3 algorithm using WEKA tools and website. Here are some examples of DHF patient medical record data, namely:

Table 1. Dataset

No	Nama	Umur	Demam	Badan_ny	Mual	Diare	Pusing	Nafsu_ma	Badan_ler
1	Didik Hen	lansia	Ya	Ya	Ya	Tidak	Ya	Tidak	Tidak
73	Weni Kus	kanak-kan	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
164	Lintang Na	kanak-kan	Ya	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
10	Kadek Yur	remaja	Ya	Ya					

Mimisan	Gusi_berd	Bintik_me	Trombosit	Diagnosis_utama
Tidak	Tidak	Tidak	pre-syok	DF
Tidak	Tidak	Tidak	pre-syok	DF
Tidak	Tidak	Tidak	ambang-dengue	DF

Table 2. Training Data

No	Umur	Demam	Badan_ny	Mual	Diare	Pusing	Nafsu_ma	Badan_ler
1	lansia	Ya	Ya	Ya	Tidak	Ya	Tidak	Tidak
73	kanak-kan	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
164	kanak-kan	Ya	Tidak	Ya	Tidak	Tidak	Tidak	Tidak

Mimisan	Gusi_berd	Bintik_me	Trombosit	Diagnosis_utama
Tidak	Tidak	Tidak	pre-syok	DF
Tidak	Tidak	Tidak	pre-syok	DF
Tidak	Tidak	Tidak	ambang-dengue	DF

Table 3. Testing Data

No	Umur	Demam	Badan_ny	Mual	Diare	Mual	Diare	Pusing	Nafsu_ma	Badan_ler
290	remaja	Ya	Tidak	Ya	Tidak	Ya	Tidak	Ya	Tidak	Tidak
298	remaja	Ya	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
317	kanak-kan	Ya	Tidak	Ya	Tidak	Ya	Tidak	Tidak	Tidak	Tidak

Mimisan	Gusi_berd	Bintik_me	Trombosit	Diagnosis
Tidak	Tidak	Tidak	ambang-d	DF
Tidak	Tidak	Tidak	ambang-d	DF
Tidak	Tidak	Tidak	ambang-d	DF

The dataset used for the calculation process was 481 case data. The dataset is divided into 2 parts, training data and testing data. Training data is used to process a calculation by forming a classifier

model. While testing data is used to measure the extent to which the classifier managed to classify correctly. Variables used for the calculation process are age, fever, body aches, nausea, diarrhea, dizziness, decreased appetite, body weakness, nosebleeds, bleeding gums, platelets and the main diagnosis.

5.1 Preprocessing

Unprocessed data is raw data, so it needs to be prepared in advance so that it can be used in the data mining process. The techniques contained in pre-processing in preparing data are removing data duplication, removing data inconsistency, correcting data errors and transforming data according to the method used. The amount of data after the pre-processing is 339 data that will be divided into two, 65% training data with 220 data and 35% testing data with 119 data.

5.2 Process

The process carried out at this stage is to determine the results of the classification calculation with the ID3 algorithm. The use of Weka tools in implementing classification methods to predict dengue fever. From the results of the process using Weka, it is obtained Re Evaluated Model, from the testing data as many as 5,298 data have been obtained as predicted results as Figure 3.
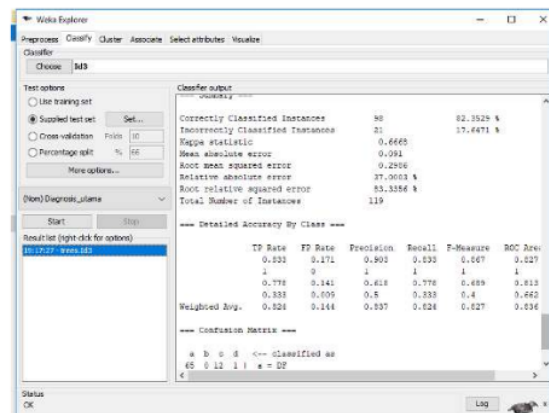

Figure 3. Display of Re-Evaluated Model Results

5.3 Cross Validation

To verify the results of the experiment, cross validation is carried out, the experiment will be carried out 10 times, A 10-fold cross-validation procedure suggests a quite stable selection[12], the dataset of this experiment is divided into two pieces namely training data and testing data, determination of training data and testing data is carried out randomly and evenly so that no grouping of documents occurs[10]. documents from one particular category in an experiment. The training data is taken as much as 277 of the dataset randomly and evenly for each category, while the rest is used as testing data, this data sharing will be done 10 times for ten experiments. The results of the experiments will find the average accuracy value. The results of cross validation from 10 experiments are shown in table 4.

Table 4. Cross Validation Result

Cross Validation		Benar	Salah
Split 1	Hasil	188	32
	%	85,45	14,54
Split 2	Hasil	188	31
	%	85,45	14,09
Split 3	Hasil	193	27
	%	87,72	12,27
Split 4	Hasil	191	27
	%	86,81	12,27
Split 5	Hasil	187	30
	%	85	13,64
Split 6	Hasil	188	30
	%	85,45	13,63
Split 7	Hasil	188	32
	%	85,45	14,54
Split 8	Hasil	187	33
	%	85	15
Split 9	Hasil	189	29
	%	85,9	13,18
Split 10	Hasil	190	29
	%	86,39	13,18
Rata-Rata	Hasil	189	30
	%	85,87	13,71

6. Conclusion

From the research that has been done, it can be concluded:

1. DHF patient data obtained from the Bhayangkara Hospital Pusdik Brimob Watukosek as many as 481 data using 13 variables that have been determined include age, fever, dizziness, nausea, diarrhea, body pain, body weakness, weakness, decreased appetite, red spots, nosebleeds, bleeding gums, platelets and the main diagnosis. Data will pass the pre-processing stage to get rid of data duplication, remove data inconsistencies, correct data errors and transform data according to the method used. The amount of data after the pre-processing is 339 data that will be divided into two, 65% training data with 220 data and 35% testing data with 119 data

2. In evaluating the test using 10-fold cross validation, it produces the highest correct accuracy value (Correctly Classified Instances) of 87.72% in split to 3 with an average accuracy of 85.87%.

3. The accuracy value by using the classification method on the ID3 algorithm produces a truth level of 82% with an error rate of 18%.

References
[1]   Tosepu R, Tantrakarnapa K, Nakhapakorn K and Worakhunpiset S 2018 Climate variability and dengue hemorrhagic fever in Southeast Sulawesi Province, Indonesia *Environ. Sci. Pollut. Res.* **25** 14944–52

[2]   Kementrian Kesehatan RI 2018 InfoDatin Situas Demam Berdarah Dengue

[3]   Damanik C (Kompas. com. 13.683 Kasus DBD di Indonesia dalam Sebulan, 133 Orang Meninggal Dunia

[4]   District S N, Province N S, Sitepu F Y, Nasution H, Supriyadi T and Depari E 2018 Epidemiological and Entomological Investigation of Dengue Fever Outbreak in **11** 8–12

[5]   Tufan Z K, Yigit H, Kacar M, Bulut C, Canpolat G, Hatipoglu C A, Kinikli S, Kosar P and Demiroz A P 2014 Sonographic findings in patients with crimean-congo hemorrhagic fever *J. Ultrasound Med.* **33** 1999–2003

[6]   Kraidech S and Jearanaitanakij K 2018 Improving ID3 Algorithm by Combining Values from

Equally Important Attributes *ICSEC 2017 - 21st Int. Comput. Sci. Eng. Conf. 2017, Proceeding* **6** 102–5

[7]   Wang Y Y, Li Y Bin and Rong X W 2017 Improvement of ID3 algorithm based on simplified information entropy and coordination degree *Proc. - 2017 Chinese Autom. Congr. CAC 2017* **2017-January** 1526–30

[8]   Feng W and Cai X 2016 The application and improvement of ID3 algorithm in WEB log data mining *Lect. Notes Electr. Eng.* **393** 577–83

[9]   WHO 2011 *Comprehensive guidelines for prevention and control of dengue and dengue haemorrhagic fever*

[10]   Rosid M A, Gunawan G and Pramana E 2015 Centroid Based Classifier With TF – IDF – ICF for Classfication of Student's Complaint at Appliation E-Complaint in Muhammadiyah University of Sidoarjo **1**

[11]   Linoff G S and Berry M J A 2011 Data Mining Techniques, Third Edition 847

[12]   Steyerberg E W 2009 Validation of Prediction Models 299–311

unplug Classification Of Dengue Hemorrhagic Disease Using Decision Tree With Id3 Algorithm

9   healthdocbox.com
    Internet Source   1 %

10 Glen Nur Awaludin, Yana Aditia Gerhana, Dian
    Sa'adillah Maylawati, Wahyudin
    Darmalaksana et al. "Comparison of Decision
    Tree C4.5 Algorithm with K-Nearest Neighbor
    (KNN) Algorithm in Hadith Classification",
    2020 6th International Conference on
    Computing Engineering and Design (ICCED),
    2020   1 %
    Publication

11 Frans Yosep Sitepu, Antonius Suprayogi,
    Dibyo Pramono, Harapan Harapan, Mudatsir
    Mudatsir. "Epidemiological investigation of
    chikungunya outbreak, West Kalimantan,
    Indonesia", Clinical Epidemiology and Global
    Health, 2020   1 %
    Publication

12 docplayer.info
    Internet Source   1 %

13 nlistsp.inflibnet.ac.in
    Internet Source   1 %

14 jurnal.iaii.or.id
    Internet Source   1 %