Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 8 No. 1, February, pp. 94 – 99
DOI:

Classification of Naive Bayes Algorithm on Dengue Hemorrhagic Fever and Typhoid Fever Based on Hematology Results

Yuni Handayani*[1], **Alvin Rainaldy Hakim**[2]
[1]*Universitas Selamat Sri, Jl. Raya Soekarno-Hatta No. Km 03, Gondoarum, Indonesia*
[2]*Universitas Al Hikmah Jepara, Jl. Raya Mayong-Jepara RW No.2, Mayong, Indonesia*
E-mail : yuni0406handayani@gmail.com[1], alvinrainaldy0@gmail.com[2]*
**Corresponding author*

Abstract - The application of increasing technology developed explicitly in the health field would significantly have an urgent role in guaranteeing quality service. Application deep data mining techniques classifier method, one among them used for classify something possibility, for example for classification disease. Dengue Hemorrhagic Fever is a disease caused by the dengue virus biting the Aedes aegypti mosquito. Meanwhile, Typhoid Fever is a disease caused by the bacterium Salmonella typhi. The second disease could attack all types of circles, fine children or mature ones. The second disease is almost the same symptom, so a proper diagnosis is needed to differentiate it. Study this applies the Naive Bayes algorithm to classify Dengue Hemorrhagic Fever and Typhoid Fever using 250 yield data test blood routine hematology at Tugurejo Hospital. Attributes used in the study, age, type sex, temperature, leukocytes, erythrocytes, hemoglobin, hematocrit, platelets, anti-dengue IgG, anti-dengue IgM, salmonella typhi o and salmonella Typhi h. The Naïve Bayes method is one of the techniques that can be used to perform analysis in determining the diagnostic results from a number of data studied with the aim of producing optimal results. The use of the Naïve Bayes method in this application is due to the probability that the accuracy value of the Nave Bayes method is close to the accuracy value of the experts.[12] The results of testing the Naive Bayes method using a confusion matrix show Recall value is 97.62%, Precision is 93.89%, accuracy is 93.33%, and Error Rate is 6 %. It can be concluded that this method is suitable for classifying Dengue Hemorrhagic Fever and Typhoid Fever and can be applied in studying this.

Keywords – Data Mining, Classifier, Naïve Bayes, Typhoid Fever, Dengue Hemorrhagic Fever

1. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) and Typhoid Fever are health problems frequently in Indonesia. Dengue Hemorrhagic Fever is more known society-wide by the term DHF (Dengue Fever) [1], while Typhoid Fever with term disease typhoid [2]. Sufferer DHF (Dengue Hemorrhagic Fever) and Typhoid Fever are getting worse year the increase, got attack children and adults. For Dengue Hemorrhagic Fever (DHF) and Typhoid Fever (TF) sufferers. No quick given treatment could raise death. Because the effort to limit the death from disease is significant, many people or sometimes doctors wrongly diagnose Dengue Hemorrhagic Fever or Typhoid Fever because the disease has almost the same features causing something so hard for analyze.

One method of prevention is with diagnosis right disease. Dengue Fever or Dengue Haemorhaege Fever (DHF) is one type of disease caused by the dengue virus, which is carried

by the Aedes aegypti mosquito as well marked with bleeding and tendencies, raising patient mortality. Typhoid fever is one type disease infectious infection acute attack small intestine caused by the bacteria salmonella typhi [3]

Research conducted by Shaufiah and Boby Siswanto [4] applies the IST-EFP algorithm to identify Dengue Hemorrhagic Fever (DHF) and Typhoid Fever (TF) disease with the use of results laboratory from hospital. Based on research that has been conducted could conclude that IST-EFP capable algorithm reduces dimensions on the dataset without changing the meaning of that dataset.

In a study by Ulva et al. [5], applying a different algorithm that is deep C4.5 algorithm classification Typhoid fever and dengue hemorrhagic craze produced an accuracy of 91.875%.

The naive Bayes algorithm is applied algorithm in prediction probability. Temporary Therefore, the Bayesian classification is one classification statistic that can predict class from possibility, got assumed that effect from the score given attribute class is independent of attribute other [6] .

Application with data mining methods as support in giving doctor's decision for diagnosis hit it risk Dengue Hemorrhagic Fever and Typhoid Fever. Data mining is required in the study because [7] data mining is a deep process of collecting and processing purposeful data to extract information necessary to the data.[13] The algorithm used in the study is the Naive Bayes algorithm. The Naive Bayes Algorithm is a classifier method that applies calculation probability [8].

Based on the background back that has been exposed before, this study's goal is to classify types of Dengue Hemorrhagic Fever, and Typhoid Fever with the use naïve Bayes algorithm based on yield data test hematology has been analyzed.

2. RESEARCH METHOD

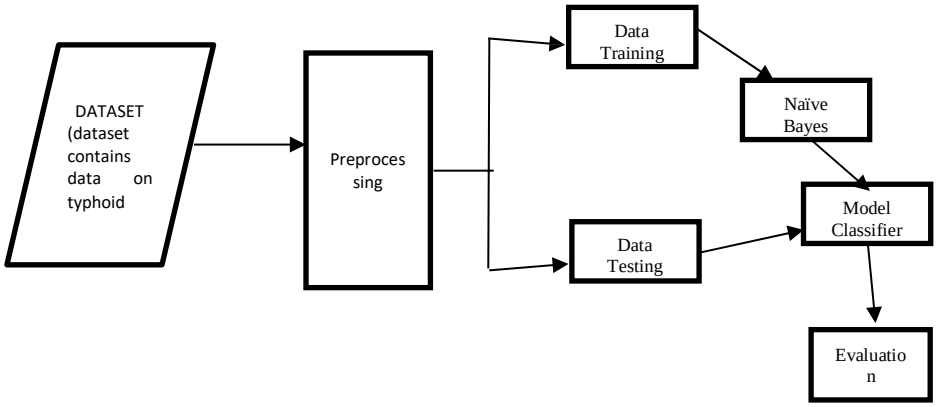


Figure 1. Proposed method

Fig. 1, explains that the first thing to do is data collection from selected Typhoid Fever and Dengue Hemorrhagic Fever datasets to know existing characteristics and features. For the second stage, the processing is carried out by collecting data on Typhoid Fever and Dengue Hemorrhagic Fever. Find and clean data from incorrect data or double. For stages next, the third stage, the data set is divided into a training subset and a test subset which will be applied in arranged classification and evaluation of models.[12] For the last stage, the purpose of type learning from the Naïve Bayes algorithm used was conducted as an evaluation.

2.1. Data Sets

Results data test used writer take is the result data test blood routine hematology taken at the Tugurejo Regional General Hospital Semarang amounted to 250 data, including data with 141 results diagnosis[9] positive disease dengue hemorrhagic fever and 109 positive disease typhoid. Data below this is an example from the data taken by researchers.

2.2. Pre-processing

At stages, this conducted preprocessing step to eliminate duplicate data and non-data filled perfectly. Data that has been preprocessed already corresponds to driven data modeling.

2.3. Naïve Bayes Classifier Model

In stages, a classification algorithm based on the Naïve Bayes Classifier is implemented in classifier models [10]. A reason to study this using based naïve Bayes algorithm because easy and fast for implemented [9] . The preprocessed data set is divided into a training data set and a test set with a ratio of 90: 10. The training set is performed to train the model, and the test set is used to validate the model results. Data training is done as input to the naïve Bayes algorithm. The results of this stage will be formed a training model, which is the stage where machine learning takes place to study the input data set and then correlate the processed features sample output.[14] This training model will later be used to run data tests through an algorithm to determine the accuracy of the model that has been formed.

2.4. Test set Input

For models that have trained, stages Next are validating it. Test datasets will conduct a process with the same steps as the training set for model validation.[15] The results of calculations testing will be compared with training results to calculate model accuracy sourced results from calculations against a backdrop of test inputs. If the prediction results are the same as the observed output, then the effects of classification accuracy increase; otherwise, the resulting error will increase. Count model accuracy with compare results calculation testing with results training.

2.5. Evaluation

Evaluation is done in the study to prove the accuracy of the data on the proposed model [11]:

$$\text{a. Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{b. Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{c. Recall} = \frac{TP}{(TP+FN)} \quad (3)$$

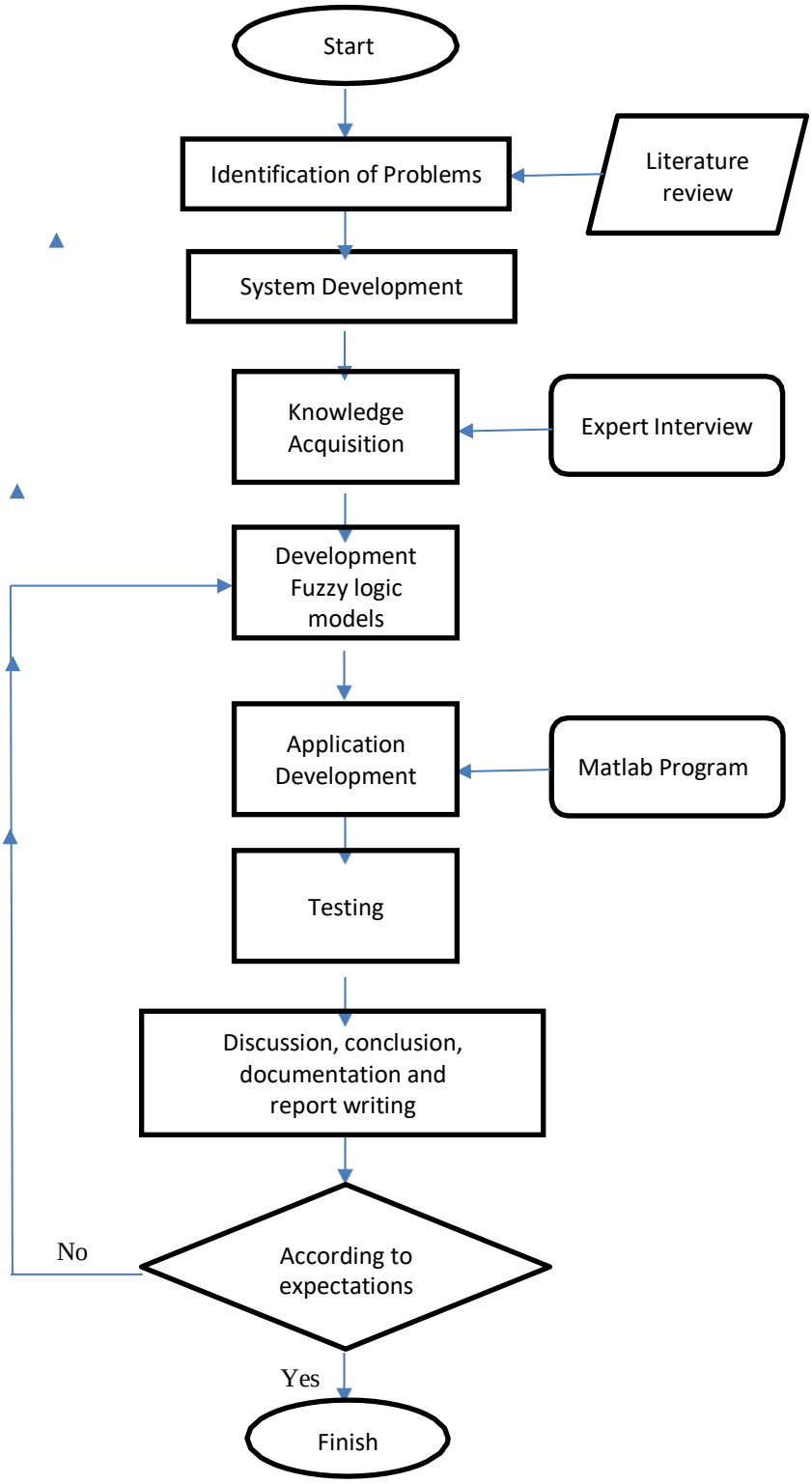
$$\text{d. Error rate} = \frac{(FP+FN)}{(TP+TN+FP+FN)} \quad (4)$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

2.6 Flowchart Sistem

The initial step of the system is to input test data. Then, the training data will first be calculated the mean and standard deviation. Next, calculate the probability from the training data. After obtaining the mean, standard deviation and probability of the training data, the next step is to calculate the posterior from the test data. Calculation of the posterior test data using the mean, standard deviation and probability of the training data. If the posterior value

has been obtained, the next step is the classification of the test data using naïve Bayes. After the classification of the test data has been obtained, the accuracy of the classification results is calculated. The final stage is storing the test result data, if not stored then it returns to the process posterior count.



3. RESULTS AND DISCUSSION

The study aims to classify the type of Typhoid Fever and Dengue Hemorrhagic Fever based on the parameters and results of hematologists. In this case, 12 parameters are used in the experiment to classify Typhoid Fever and Dengue Hemorrhagic Fever. Based on the results test show that algorithm learning the Naïve Bayes machine can be conducted with good obtained results accuracy up to 93.33%, with 97.62% recall value and Precision of 93.89%. This shows that the Naïve Bayes model perfectly classifies Typhoid Fever and Dengue Hemorrhagic Fever.

This study aims to design a classification model for the diagnosis of Typhoid Fever and Dengue Hemorrhagic Fever. In the research conducted, the classifier model applies the data set. It consists of 250 data, with details of 141 results diagnosed Dengue Hemorrhagic Fever and 109 results diagnosed positive where is typhoid. That data set has a total of 12 parameters. Among these parameters in the study are age, sex, temperature, leukocytes, erythrocytes, hemoglobin, hematocrit, platelets, anti-dengue IgG, anti-dengue IgM, salmonella typhi o, and salmonella typhi h.

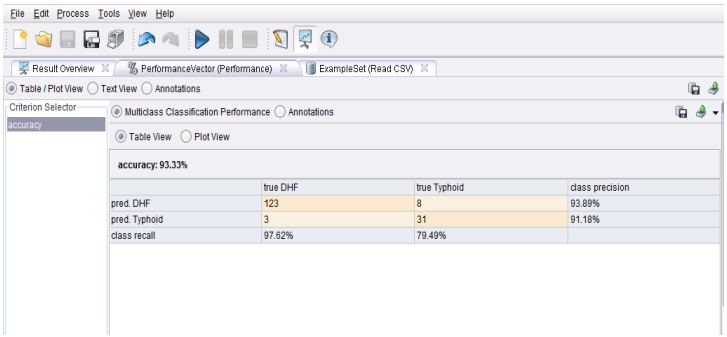


Figure 3. Result display

4. CONCLUSION

The destination of the study is to develop a deep model classification of Typhoid Fever and Dengue Hemorrhagic Fever using the Naïve Bayes algorithm with 12 parameters based on the results of the test blood hematology routine. Based on results testing in research that has been done, found results classification diagnosis Typhoid Fever and Dengue Hemorrhagic Fever using the naïve Bayes algorithm achieves score accuracy of 93.33%, meanwhile value on recall 97.62% while for value at an error rate of 6% and value for the Precision of 93.89%.

Based on results obtained _ showing that the Naïve Bayes Classifier model is excellent in classifying Typhoid Fever and Dengue Hemorrhagic Fever and gives support decisions in diagnosis to the patient.

REFERENCES

[1] F. Tawakal and A. Azkiya, "Diagnosa Penyakit Demam Berdarah Dengue (DBD) menggunakan Metode Learning Vector Quantization (LVQ)," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 4, no. 3, p. 56, 2020, doi: 10.14421/jiska.2020.43-07.

[2] B. Andrianto, Y. K. Suprapto, I. Pratomo, and I. Irawati, "Clinical decision support system for typhoid fever disease using classification techniques," *Proc. - 2019 Int. Semin. Intell. Technol. Its Appl. ISITIA 2019*, pp. 248–252, 2019, doi: 10.1109/ISITIA.2019.8937286.

[3] I. L. T. Co, M. R. E. Estuar, K. E. Espina, R. J. E. A. Lara, and V. C. D. De Los Reyes,

"Integrating health indices towards the development of a typhoid disease model using STEM," *Proc. 2016 3rd Int. Conf. Inf. Commun. Technol. Disaster Manag. ICT-DM 2016*, 2017, doi: 10.1109/ICT-DM.2016.7857211.

[4] Shaufiah and B. Siswanto, "Association rule mining for identifying Dengue Hemorrhagic Fever (DHF) and Typhoid Fever (TF) disease with IST-EFP algorithm," *2016 4th Int. Conf. Inf. Commun. Technol. ICoICT 2016*, vol. 4, no. c, 2016, doi: 10.1109/ICoICT.2016.7571920.

[5] U. Febriana, M. Tanzil Furqon, and B. Rahayudi, "Klasifikasi Penyakit Typhoid Fever (TF) dan Dengue Haemorhagic Fever (DHF) dengan Menerapkan Algoritma Decision Tree C4.5 (Studi Kasus : Rumah Sakit Wilujeng Kediri)," 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[6] M. A. Jabbar and S. Samreen, "Heart disease prediction system based on hidden naïve bayes classifier," *2016 Int. Conf. Circuits, Control. Commun. Comput. I4C 2016*, 2017, doi: 10.1109/CIMCA.2016.8053261.

[7] C. Raju, E. Philipsy, S. Chacko, L. Padma Suresh, and S. Deepa Rajan, "A Survey on Predicting Heart Disease using Data Mining Techniques," *Proc. IEEE Conf. Emerg. Devices Smart Syst. ICEDSS 2018*, no. March, pp. 253–255, 2018, doi: 10.1109/ICEDSS.2018.8544333.

[8] W. B. Zulfikar, Y. A. Gerhana, and A. F. Rahmania, "An Approach to Classify Eligibility Blood Donors Using Decision Tree and Naive Bayes Classifier," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–5, 2019, doi: 10.1109/CITSM.2018.8674353.

[9] F. J. Yang, "An implementation of naive bayes classifier," *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 301–306, 2018, doi: 10.1109/CSCI46756.2018.00065.

[10] H. T. Zaw, N. Maneerat, and K. Y. Win, "Brain tumor detection based on Naïve Bayes classification," *Proceeding - 5th Int. Conf. Eng. Appl. Sci. Technol. ICEAST 2019*, pp. 1–4, 2019, doi: 10.1109/ICEAST.2019.8802562.

[11] Amrin and H. Saiyar, "Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Naive Bayes," *J. Ris. Komput.*, vol. 5, no. 5, pp. 498–502, 2018, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C498

[12] Guo, Duan Li. "A New Naive Bayes Text Classification Algorithm." *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 2, 2014, https://garuda.kemdikbud.go.id/documents/detail/1498616.

[13] Sutopo, Ristu. "TEXT CLASSIFICATION USING NAIVE BAYES UPDATEABLE ALGORITHM IN SBMPTN TEST QUESTIONS." *Telematika : Jurnal Informatika dan Teknologi Informasi*, vol. 13, 2016, https://garuda.kemdikbud.go.id/documents/detail/1752584.

[14] Suardi, Putri. "The Optimization of Feature Selection Using Genetic Algorithm with Naïve Bayes Classification for Home Improvement Recipients." *International Journal of Engineering and Emerging Technology*, vol. 3, no. 1, 2018, https://garuda.kemdikbud.go.id/documents/detail/1097282.

[15] Endang Sri Palupi. "CLASSIFICATION OF THE POOR IN INDONESIA USING NAIVE BAYES ALGORITHM AND NAIVE BAYES ALGORITHM BASED ON PSO." *Jurnal Riset Informatika*, vol. 4, no. 3, 2022, https://garuda.kemdikbud.go.id/documents/detail/2847406.