Prediction of Days in Hospital Dengue Fever Patients using K-Nearest Neighbor

Dewa Ayu Putri Wulandari[1*], Kadek Ary Budi Permana[2], and Made Sudarma[3]
[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University.
[3]Department of Electrical and Computer Engineering, Udayana University
*Email: dewaayuputriwulandari@gmail.com

***Abstract*— Dengue fever is found more in tropics and subtropics area. The World Health Organization (WHO) noted that Indonesia is as the highest dengue fever cases in South Asian since 1968 till 2009. The treatment of patient dengue fever in hospital had spent rest time or the days in hospital in variation treatment. Days in hospital often becomes questions from the patients' family. To predict of days in hospital is maybe to know the capacity for the long period strategies. In this research, we used machine learning approach to predict the period of dengue fever patients. One of the machines of learning method is K-Nearest Neighbor (K-NN). In this case, we analyze the result of implementation K-NN method that used to predict period of days in hospital patients for dengue fever. This paper used the result of laboratory checking of blood complete with patients as parameter in predict the period of days in hospital dengue fever. The result of examination used K-NN to show the accuracy levels reach 65,67% with *k* optimal is *k*=13.**

Keywords*— Data mining, K-Nearest Neighbor, Prediction of Days in Hospital

I. INTRODUCTION

Dengue fever (DBD) is more found more in tropics and subtropics area. The world data shows that Asian gets the first level in the sufferer of dengue fever for every year. While it could be known since 1968 till 2009 World Health Organization (WHO) noted that Indonesia is as the highest nation in sufferer of dengue fever in the South East Asia [1]. The main factor is *Aedes Aegypti* mosquito in which had spread in tropics and subtropics zone. The spreading is globally which can be by the trading and traveling [2].

The infection is caused by dengue virus by knowing the anamnesis and physic checking. Other hand, it also is needed laboratory checking as supporting as checking blood completely. There are some parameters those can be known in checking blood completely, such as red blood cell, white blood cell, platelet, hemorrhoids and more things as the parameter. The complete blood checking to the patients with the infection of virus dengue in other wise has benefit to diagnosis the patients and it also used to observe the patients [3].

The handling of the dengue fever patients in the hospital had spent the variation period. The days in hospital focuses on the arrival of the patients received by hospital and the date of patient's recovery and leave the hospital [4]. The days in hospital often become big questions from the patient's family. Of course, the family of patients wants to know how long the patients are treated in hospital. This information is as needs generally for the patient's family to prepare the all needs that concern to the cost of patients or the other things in outside by hospital. Without analyzing, the owner of hospital only observed the period of patients commonly. So, it is needed to analyze to predict the days in hospital for the sufferer of dengue fever. To check the days in hospital is also to know the capacity of hospital for long distances strategies [5].

The aim of this research is prediction the days in hospital of dengue patients based on the result of checking in laboratory blood completely. The result of checking in laboratory blood completely that used in this research is the result of checking in the first time accepted by the hospital when the patients come in the hospital. In predicting of days in hospital for the dengue patients is done by looking for the long cases that has correlation with this case.

In this research, we used machine learning approach to predict period days in hospital dengue fever patients. We also used K-Nearest Neighbor (K-NN) classification to run prediction model days in hospital. Other side, we also calculate the accuracy to measure working that is result by classification of K-NN.

The next part is arranged as follow, namely in the part two, we explain shortly about the research that has correlation with this research. In the third part, we describe the data and method which used. In the fourth, we show the result of research. In the fifth part, we made conclusion our research and give guide to the next researchers.

II. RELATED RESEARCH

This research [6] predicted some old patients who getting diabetes and they were treated in the hospital. By using ages categories, gender, rash, expected primary paper, receiving cases, and APR-DRG. This research can be predicted period time from every patient. This research also compares some algorithms machine learning to do prediction. Research [5] compares the three algorithms machine learning that different to predict LOS (Length of Stay) the patients who treated in hospital. The aim in predict of LOS namely to be probably the hospital in schedule receiving more again that focuses on reducing variation in the bed of hospital.

Some other researches also have correlation to the prediction in using K-NN had been done more before. K-NN can be used to do the prediction of cost values [7]. Based on the result of K-NN gotten the small fault ratio so the result of rational and reasonable. Besides, it depends on real data cost value, result of the nearest prediction. The other researches do the ship route prediction based on the classification from K-NN [8]. This algorithm is examined to the real data in which taken from the messages AIS that collected in around Malta. The experiment

showed that algorithm we reached precision 0,794, recall 0,785 and the accuracy 0,931. This research in predict the tending topics can be used K-NN method [9]. This research is done based on the information and conversation in which taken from online mass media and social media that related to government in Makasar City. The result of research showed that the using K-NN method can get the accuracy 81,13%.

III. METHODOLOGY

A. Data Source

The research is done in the XYZ hospital with the aim to help the owner of hospital in predicting days in hospital dengue fever period. The prediction of days in hospital to the dengue fever patients used the result of checking in laboratory complete blood that used the first time. In prediction the days in hospital dengue fever involve the genders, ages, and result of checking from laboratory in completed blood. In the laboratory completed blood found 20 parameters that consisting WBC, LYM%, MID%, GRA% LYM, MID, GRAN, RBC, HGB, HCT, MCV, MCH, MCHC, RDW%, RDWa, PLT, MPV, PDW, PCT, LPCR. Totally, there are 22 parameters in which used to predict the period days in hospital for the sufferer of dengue fever. The number of training that used for the data is reaches 1881 and testing data 300. The table 1 shows that the examples which used in this research.

TABLE I. THE LIST CHECKING OF LABORATORY BLOOD COMPLETELY

	Case 1	Case 2	Case 3
Genders	Female	Male	Males
Ages	56 years	28 years	6 years
WBC	5.4	3	15.3
LYM%	41.7	22.7	34
MID%	4.3	7.4	8.8
GRA%	5.4	69.9	57.2
LYM	2.2	0.6	5.2
MID	0.3	0.3	1.4
GRAN	2.9	2.1	8.7
RBC	3.91	5.92	6.47
HGB	12	17.1	17.5
HCT	36.6	51.7	51.3
MCV	93.5	87.2	79.1
MCH	30.6	28.9	27
MCHC	32.7	33.1	34.1
RDW%	13.8	13.9	14.3
RDWa	74.7	70.4	63.2
PLT	71	65	49
MPV	8.7	9.1	7.9
PDW	13.2	13.5	12.5
PCT	0.06	0.05	0.03
LPCR	21.5	24.2	12.5
Period Days in Hospital	5 days	6 days	13 days

B. K-Nearest Neighbor

K-Nearest Neighbor (K-NN) is the method to classify of object based on the training with short distance toward the object. K-NN is classified as a supervised learning algorithm which results from new requests are classified by majority of categories in K-NN. The modus class will become algorithm class. Algorithm classify the new object based on attribute of training data. K-NN used the environment classification as prediction value from the new demand examples. These are simple algorithm working in short position from the samples of requirement training to decide the environment [9]. The process of classification is almost similar to the polling because every object nearest give one vote to its class. Then the object is duties to the class with the most voters [10]

The classification of K-NN is supposed as black box that receive the features set $x = (x_1, x_2, ..., x_n)$ as input and back to the class label y as output. Figure 1 showed that as the examples of classification K-NN, that receives two features x_1 and x_2 as inputs, and returns to one of the class as output: green orange, purple. To be given T set that had given training and new input $x = (x_1, x_2)$ (as representative by the red dot in the figure), the classification K-NN takes k the nearest neighbor from S in T and between them choose the often class [8].

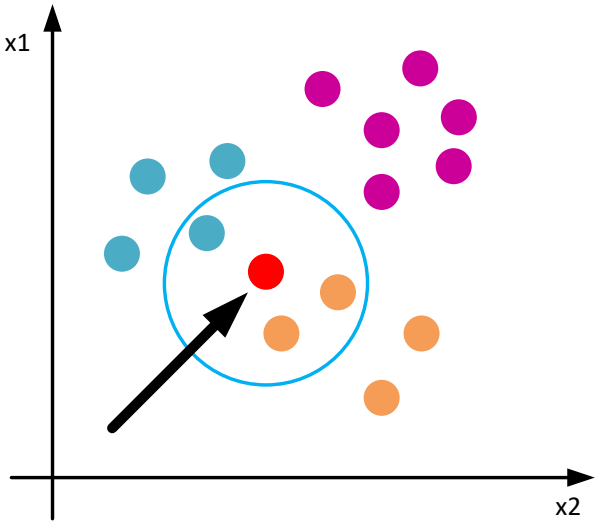

Fig 1. The Classification of K-NN [8]

Euclidean distance is used to calculate the distance between training data with testing data. Euclidean distance between x_i and x_j is written based on the following equations (1):

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2} \quad (1)$$

$d(x_i, x_j)$: *Euclidean Distance*
x_i : *i-record*
x_j : *j-record*
a_r : *r- data*
i, j : *1, 2, 3, ..., n*

The system performance can be calculated using accuracy from the classification as the result. The selection of training data and testing data can influence the result of accuracy that is produced because the pattern in training data is used as a rule in class classification in testing data [11]. The accuracy is calculated using the equations (2).

$$accuracy = \frac{total\ data\ valid}{total\ data\ testing} \times 100\% \quad (2)$$

IV. Result and Discussion

In this step explain about method K-NN steps in doing the prediction for days in hospital period the dengue fever patient and show the result that gotten. The prediction of days in hospital for dengue fever is divided into 3 classes, from this class the period for days in hospital between 0-3 days, 4-9 days and longer than 9 days [2].

The steps in applying the method K-NN in this research are:

1) Calculate the distance between the data that evaluated with all training data
2) Arrange the data based on the distance in forming by the nearest distance
3) Decide *k* value as the number of nearest neighbor
4) Look for the nearest number neighbor class then arrange the class into the evaluation data.

Figure 2 show the steps in K-NN method.

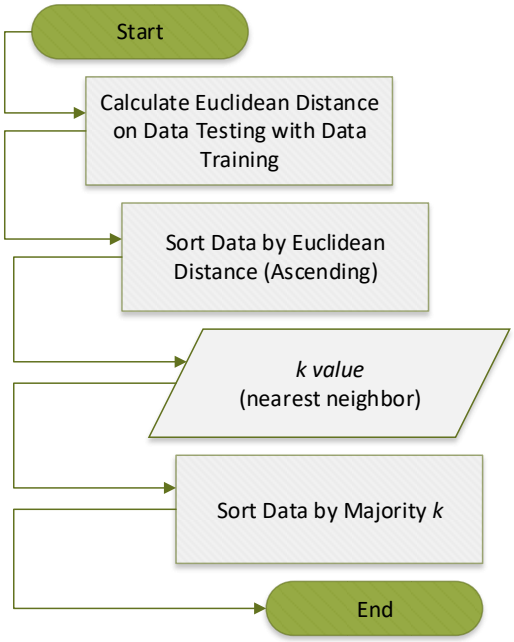

Fig 2. K-NN Method

The examining is done to some *k* value that shows the number of the nearest neighbor. This table shows that the accuracy is in which gotten using K-NN method. In some of *k* value had examined.

TABLE II. Accuracy of K-NN

k	Accuracy
3	62%
5	65%
7	64%
9	62%
11	63,33%
13	65,67%
15	65,33%
17	65,33%

V. Conclusion and Future Work

K-NN had successes in implementation to predict the days in hospital for the dengue fever patients. The *k* value is optimal in predict days in hospital reaching k = 13, because has the accuracy with the biggest presentation namely, 65,67%. The accuracy is gotten in not high. This case happened because of outside checking in laboratory blood completely that also influence the period days in hospital for the dengue fever

For the future in prediction of days in hospital patients are generally also uses the other observing of check as the result in laboratory the patient. It can be used as other parameter such as medical action, medicine and other to more increase the accuracy from the predict result. This research also can be developed to predict the cost of days in hospital that will be charged to the patients.

References

[1] Kementerian Kesehatan RI, "Demam Berdarah Dengue," *Bul. Jendela Epidemiol.*, vol. 2, p. 48, 2010.

[2] S. Pooransingh, S. Teelucksingh, and I. Dialsingh, "Dengue Deaths: Associated Factors and Length of Hospital Stay," vol. 2016, 2016.

[3] N. L. P. P. Sastri and W. A. A. Lestari, "Gambaran Hasil Pemeriksaan Darah Lengkap Pada Pasien Suspect Infeksi Virus Dengue Di Rumah Sakit Surya Husadha Denpasar Februari-Juli 2014," E-Jurnal Med., vol. 5, no. 7, pp. 1–5, 2016.

[4] C. Wang, "Prediction of Days in Hospital for Children Using Random Forest," CISP-BMEI 2017, 2017.

[5] P. C. Pendharkar and H. Khurana, "Machine learning techniques for predicting hospital length of stay in pennsylvania federal and specialty hospitals," *Int. J. Comput. Sci. Appl.*, vol. 11, no. 3, pp. 45–56, 2014.

[6] V. Veena and A. Ravikumar, "Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus," *Int. J. Comput. Appl.*, vol. 95, no. 17, pp. 12–16, 2014.

[7] K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, "Stock price prediction using K-nearest neighbor algorithm," *Int. J. Business, Humanit. Technol.*, vol. 3, no. 3, pp. 32–44, 2013.

[8] A. Lo Duca, C. Bacciu, and A. Marchetti, "A K-nearest neighbor classifier for ship route prediction," *Ocean. 2017 - Aberdeen*, vol. 2017–Octob, pp. 1–6, 2017.

[9] S. Liang, Y. Liu, C. Wang, and L. Jian, "Trending Topic Prediction by Optimizing K-Nearest Neighbor Algorithm," *Architecture*, pp. 291–296, 2009.

[10] J. E. V. Ferreira, C. H. S. da Costa, R. M. de Miranda, and A. F. de Figueiredo, "The use of the k nearest neighbor method to classify the representative elements," Educ. Quim., vol. 26, no. 3, pp. 195–201, 2015.

[11] Bagus, A. Peling, I. N. Arnawan, I. P. A. Arthawan, and I. G. N. Janardana, "Implementation of Data Mining to Predict Period of Students Study Using Naive Bayes Algorithm," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, pp. 53–57, 2017.