Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 9 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i09.40917

PAPER

Hemodialysis Patient Death Prediction Using Logistic Regression

Dony Novaliendry[1,2](✉), Oktoria[1,2], Cheng-Hong Yang[2], Yenny Desnelita[3], Irwan[3], Roni Sanjaya[3], Gustientiedina[3], Yaslinda Lizar[4], Noper Ardi[5]

[1]Universitas Negeri Padang, Padang, Indonesia

[2]National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[3]Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia

[4]Universitas Islam Negeri Imam Bonjol, Padang, Indonesia

[5]Politeknik Negeri Batam, Batam, Indonesia

dony.novaliendry@ft.unp.ac.id

ABSTRACT

Hemodialysis is a procedure for cleaning the blood from the waste products of the body's metabolism. this is one of modality to treat end stage kidney disease. There are two main classifications of this disease, namely acute kidney failure and chronic kidney failure. Kidney failure occurs when kidney damage is severe enough or lasts a long time so that the disease is generally the final stage of kidney disease. Dialysis is performed on patients with kidney failure, both acute kidney failure and chronic kidney failure. This study is aimed to predict the mortality risk of hemodialysis patients. The Taiwanese hemodialysis center enrolled a total of 665 hemodialysis patients. The prediction is based on Logistic Regression. Compared with K-Nearest Neighbor, linear discriminant, Tree, and ensemble, Logistic Regression performed better. As for related medical variables like parathyroid surgery, urea reduction ratio, etc., they play a much smaller role in mortality risk factors than diabetes and cardiovascular disease.

KEYWORDS

logistic regression, diabetes, hemodialysis, prediction

1 INTRODUCTION

Chronic kidney failure is a condition when kidney function gradually declines due to damage to kidney tissue. Medically, chronic kidney failure is defined as a decrease in the rate of renal filtration for 3 months or more. Consequently, the high-level wastes are accumulated in the blood which may impact to high blood pressure, weak bones, anemia, nerve damage and poor nutritional health. In addition, the risk of cardiovascular diseases is increased [1][2]. Doctors determine the stage of kidney disease using the glomerular filtration rate (GFR), a math formula using a person's age, gender, and their serum creatinine level (identified through a blood test). Creatinine, a waste product that comes from muscle activity, is a key indicator of kidney function. When kidneys are working well they remove creatinine from the blood; but as kidney function slows, blood levels of creatinine rise [3]. If chronic kidney disease is recognized in its early stages, treatment focuses on stopping or

Novaliendry, D., Oktoria, Yang, C.-H., Desnelita, Y., Irwan, Sanjaya, R., Gustientiedina, Lizar, Y., Ardi, N. (2023). Hemodialysis Patient Death Prediction Using Logistic Regression. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(9), pp. 66–80. https://doi.org/10.3991/ijoe.v19i09.40917

Article submitted 2023-03-07. Resubmitted 2023-05-03. Final acceptance 2023-05-03. Final version published as submitted by the authors.

slowing the progression of the disease to prevent further kidney damage [4][5]. If the patient also has diabetes or high blood pressure, it is more important for the patient to receive treatment for both conditions and to effectively relieve the symptoms to prevent further kidney damage. For severe cases of chronic kidney disease where kidney failure has occurred, then one way to overcome it is to do hemodialysis. An important part of treating excess body waste is through hemodialysis (HD) [6].

Biomedical and bio-information, including prediction, are becoming increasingly popular with machine learning. Logistic Regression is one of machine learning algorithm that have many advantages in prediction comparing to other algorithm. Logistic regression does not have normality and heteroscedasticity assumptions on the independent variables used in the model, so there is no need for classical assumption tests even though the independent variables are more than one. The independent variables in logistic regression can be a mixture of continuous, district, and dichotomous variables. Logistic regression does not require the limitations of the independent variables. Logistic regression does not require that the independent variables be in the form of intervals [7].

A Logistic Regression model is presented in this paper for predicting the mortality of hemodialysis patients with diabetes [8][9]. The purpose of this study was to provide a model that would assist doctors/physicians in predicting mortality rates for hemodialysis patients. The rest of this paper is organized as follows. Section 2 presents Methodology used in this research. Section 3 describes the results and analysis. Finally, the conclusion is described in Section 4.

2 MATERIALS & METHOD

2.1 Chronic kidney disease

Kidneys are organs that function to filter waste (waste substances from the body's metabolism) and excess fluid from the blood to be excreted through urine. Each day, the two kidneys filter about 120–150 liters of blood and produce about 1–2 liters of urine. Not only that, the kidneys also have a number of important functions, including: Regulate levels of chemicals in the body so that the heart and muscles can work properly, helps regulate blood pressure, producing substances such as vitamin D to maintain bone health, producing glycoprotein hormones, namely erythropoietin to help stimulate the production of red blood cells [10][1].

Chronic Kidney Failure is one of the diseases experienced by this important organ. This condition is characterized by a gradual decline in kidney function. This kidney damage can be in the form of tissue abnormalities, blood composition, and urine or kidney imaging tests, which are experienced for more than three months [11]. Chronic kidney failure will end up becoming end renal failure (ESRD) if not treated properly. The condition occurs due to the buildup of body waste, fluids, and electrolytes that can harm the body. If it is so, the body requires a dialysis procedure or dialysis. The procedure aims to help the performance of the kidney that has been damaged. Patients are usually not aware of the symptoms that appear, so treatment steps tend to be late. If it is so, a number of symptoms will appear as a sign that the kidneys cannot function properly [12][13].

The National Kidney Foundation (NKF) divided kidney disease into five stages. This helps doctors provide the best care, as each stage calls for different tests and treatments. GFR, or glomerular filtration rate, is a mathematical formula that calculates kidney disease stage based on a person's age, gender, and serum creatinine level. It is important to monitor kidney function by measuring creatinine, a waste product from muscle activity [14]. Creatinine is removed from the blood by the kidneys when they function well, but it is increased when kidney function slows down. There are 5 stages of Chronic Kidney Disease, Stage 1 with normal or high GFR (GFR > 90 mL/min), Stage 2 Mild CKD (GFR = 60–89 mL/min), Stage 3A Moderate CKD (GFR = 45–59 mL/min), Stage 3B Moderate CKD (GFR = 30–44 mL/min), Stage 4 Severe CKD (GFR = 15–29 mL/min), Stage 5 End Stage CKD (GFR <15 mL/min). For kidney disease diagnosed at an early stage, the main treatment options are: Medication, Lifestyle changes, Diet changes. For severe cases of chronic kidney disease where kidney failure is already present, the only available treatment options are: Dialysis (dialysis), Kidney transplant/graft [3][15].

2.2 Logistic regression principle

Logistic Regression is a classification algorithm to find the relationship between discrete/continuous features (input) and the probability of certain discrete output results. Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Types of Logistic Regression:

1. Binary Logistic Regression: This is a Logistic Regression which only has 2 outputs (classifying into 2 different classes). Example: Positive-Negative, Obesity-Not Obesity.
2. Multinomial Logistic Regression: Is a Logistic Regression that has 2 or more outputs (classifies into 2 different classes). For example, Sentiment Analysis class positive, negative, and neutral sentences.
3. Ordinal Logistic Regression: Is a Logistic Regression that has 2 or more outputs with respect to the order. (Classifying into 2 different classes by paying attention to the order). An example is dividing the class of students in the range of Grade Point Average 1.xx, 2.xx, 3.xx, and 4.00.

Logistic Function is a function formed by equating the Y value in Linear Function with Y value in Sigmoid Function. The purpose of the Logistic Function is to represent the data that we have in the form of a Sigmoid function [16].

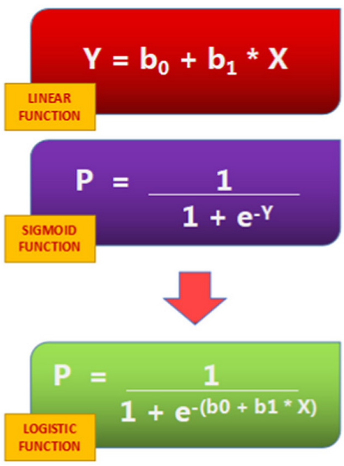

Fig. 1. Logistic function

From Figure 1, we can form a Logistic Function by performing the following steps:

1. Perform the Inverse operation on the Sigmoid Function, so that the sigmoid function changes its shape to $Y = \ln(p/(1 - p))$.
2. Equivalent to the Linear function $Y = b_0 + b_1 * X$ so that we get the equation $\ln(p/(1 - p)) = b_0 + b_1 * X$.

Change the equation $\ln(p/(1-p)) = b_0 + b_1 * X$ into logarithmic form so that we get the equation $P = /(1 + e\text{^}\ b_0 + b_1 * X))$ [17][18].

2.3 Determination of coefficient of logistic function: Maximum likelihood + R-Squared (R^2)

Maximum likelihood estimation is a method that determines values for the parameters of a model. The parameter values are found such that they maximize the likelihood that the process described by the model produced the data that were actually observed. Maximum Likelihood is a way to determine the position of the Sigmoid which is the best model that can be formed from the available data. R-Squared is a method used to determine whether a Logistic Function with a Maximum Likelihood value can represent data well (good if R-squared = 1, Bad if R-squared = 0) [19].

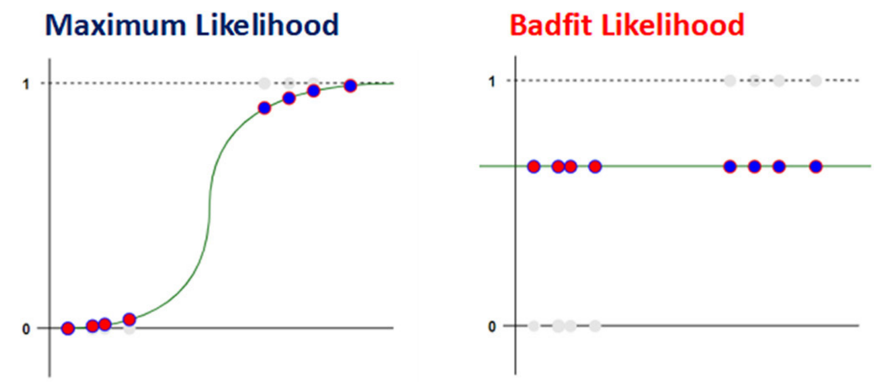

Fig. 2. Logistic function: likelihood

There are 2 important parameters needed in finding the R-Squared value, namely Maximum Likelihood and Badfit Likelihood as seen in Figure 2.

There is another approach to produce a Logistic Function that can classify data well, namely by using the Gradient Descent method. The Gradient Descent method works by updating the weights (b_0 and b_1) by minimizing the Loss value. The smaller the Loss value, the better the Logistic Function in representing the data [15][20][21]. Loss value in Logistic Regression can be known using the following formula:

$$\text{Loss} = -(Y_{aktual} \log(Y_{prediksi}) + (1 - Y_{aktual})\log(1 - Y_{prediksi})) \quad (1)$$

2.4 Dataset

The dataset acquired from Kaohsiung Chang Kung Hospital. Datasets were reviewed in accordance with the Declaration of Helsinki using the approval number 101–1595B of Kaohsiung Chang Kung Hospital in Taiwan.

In total, there are 665 HD patients, of whom 59 suffer from cardiovascular diseases (CVDs) and 159 suffer from diabetes. In accordance with the age of the patients, the data are divided into four sets: Dataset A, Dataset B, Dataset C, and Dataset D.

There are 66 patients in dataset A, all of whom are younger than 56 years of age. 66 patients are included in dataset A, where the average age is under 56 years old. A total of 354 patients in dataset B range in age from 56 to 61 years old. The dataset C consists of 170 patients between the ages of 61 and 75. 75 patients over 75 years old are included in dataset D. 75 patients over 75 years old are included in dataset D. Males and females are included in each dataset.

This paper tests classifier performance using the following terms:

Hemodialysis patient's risk of death is predicted with the aid of this ACC, which measures the accuracy of a classifier in the prediction. An incorrect classification is determined by the error rate [22]. Known also as a recall, sensitivity (Se) measures the accuracy of the classifier when predicting positive proportions. In the classification process, the classifier measures the specificity (Sp) of its predictions. As its name implies, precision, or positive predictive value (PPV), is a measure of how much variability can be predicted from results. NPV refers to the ratio between negative predictions and actual negative values. In health care, prevalence measures the proportion of the population with certain diseases, which may include diabetes, anemia, and CVD [23][24]. Diabetes (DM) was the focus of this paper.

The values of all metrics were derived using the classperf command in Matlab®, as follows:

$$CP = \text{classperf}\ (\mathit{truelabels}, \mathit{classout})$$

As well as these metrics, we also used the following ones. Test/prediction accuracy of a classifier is measured by the F_1 score [25]. A dichotomous predicted result's performance can be determined by informedness (BM), which goes beyond multiclass cases. As the inverse of Youden's J, markedness (MK) describes the degree of markedness.

$$F_1 = 2.\frac{PPV.TPR}{PPV + TPR}$$

$$BM = TPR + TNR = 1$$

$$MK = PPV + NPV$$

3 RESULT AND DISCUSSION

3.1 Prediction results of dataset A

Based on the results of Logistic Regression (LR) testing, it was observed that Dataset A had a correct rate (acc) training of 0.879 for a period of 695.13 msec. In comparison with KNN, simple tree (DT), and linear discriminant (LD), the Logistic Regression execution time is longer. A linear discriminant and ensemble logic algorithm achieved a short training time of 318.69 msec, and a long training time of 695.13 msec for dataset A respectively. Dataset A is listed in Table 1 along with the complete training comparison results.

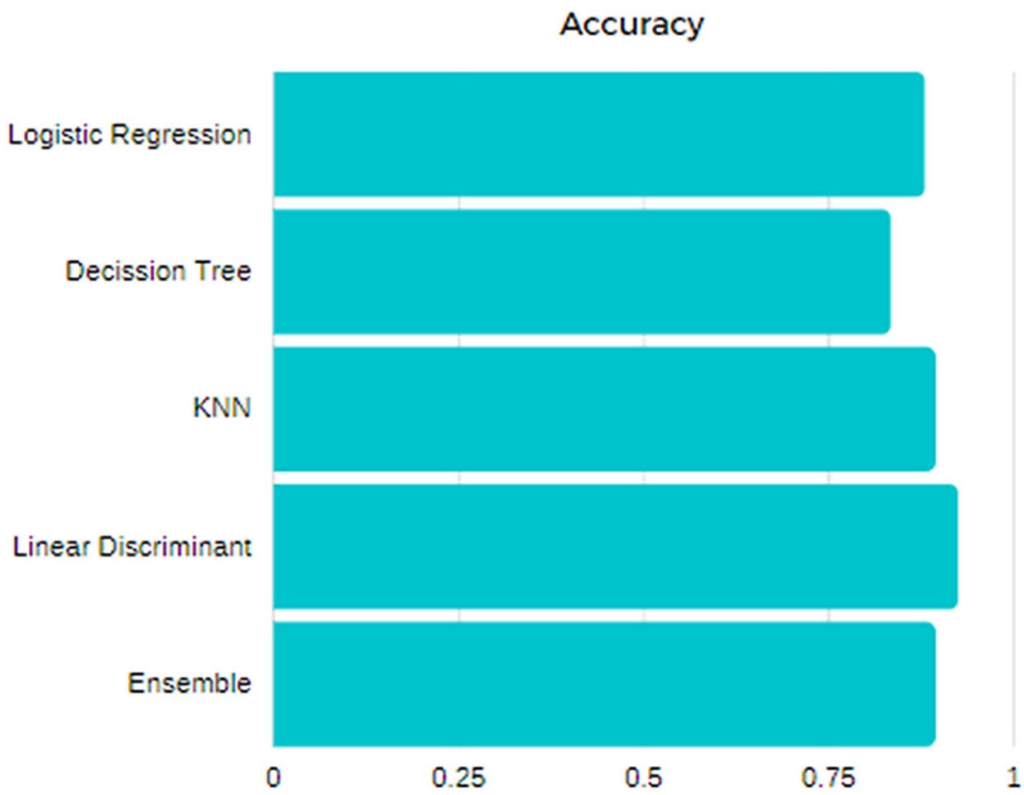

Fig. 3. Training comparison of dataset A -Accuracy

Fig. 4. Training comparison of dataset A – time

From Figures 3 and 4 it can be seen that data analysis using Logistic Regression produces a truth level of 0.879. Logistic Regression has a lower error rate than Decision Tree, yielding 0.04 but higher than Linear Discriminant, Logistic Regression, Ensemble (En), and KNN. Ensemble Tree produces the highest error rate of 0.1061. In Table 2, we present detailed predictions based on dataset A.

Linear discriminant of 0.924 was found to be the most accurate predictor of mortality in Dataset A, followed by Ensemble and KKN of 0.894, and Logistic regression of 0.879.

Table 1. Prediction result comparison of dataset A

Metrics	LR	DT	KNN	LD	En
Correct rate	0.960	0.924	0.893	0.955	0.894
Error rate	0.040	0.076	0.107	0.045	0.106
Sensitivity (TPR)	0.569	0.429	0.514	0.571	0
Specificity (TNR)	1	0.983	0.974	1	1
Positive Predictive Value	1	0.750	1	1	NaN
Negative Predictive Value	0.952	0.936	1	0.952	0.894
PositiveLikelihood	NaN	25.286	NaN	NaN	NaN
NegativeLikelihood	0	0.581	0	0.429	1
Prevalence	0.106	0.106	0.106	0.106	0.106
F1	0.725	0.546	0.679	0.727	NaN
Bookmaker Informedness (BM)	0.569	0.412	0.488	0.571	0
Markedness(MK)	0.952	0.686	1	0.952	NaN

Table 1 shows that the highest correct rate of prediction for Dataset A is 0.960 for Logistic Regression, followed by 0.955 for Linear Discrimination, 0.924 for Decision Tree, 0.894 for Ensemble, and 0.893 for KKN.

3.2 Prediction results of dataset B

Of the total 354 patients Dataset B, Logistic Regression successfully predicted with precision reaching 0.8730 and up to 0.893 respectively for training and test. When compared to the decision tree, Logistic regression is more inferior where Tree provides accuracy 0.927 and 0.944 for training and test dataset B. Other facts show that Logistic regression is still better than other methods, namely KNN and Ensemble, where respectively provides accuracy values well below Logistic Regression and linear discriminant. Table 3 shows the results of the training dataset B, whereas Table 4 is the predicted mortality risk for the B dataset.

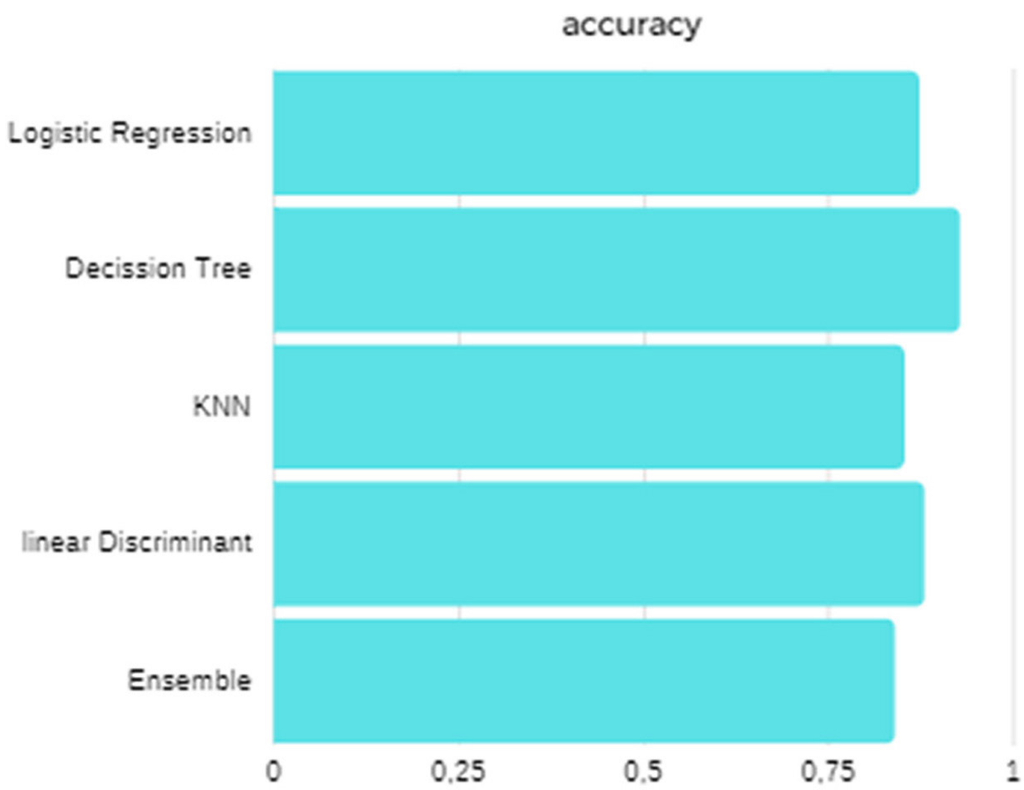

Fig. 5. Training comparison of dataset B – Accuracy

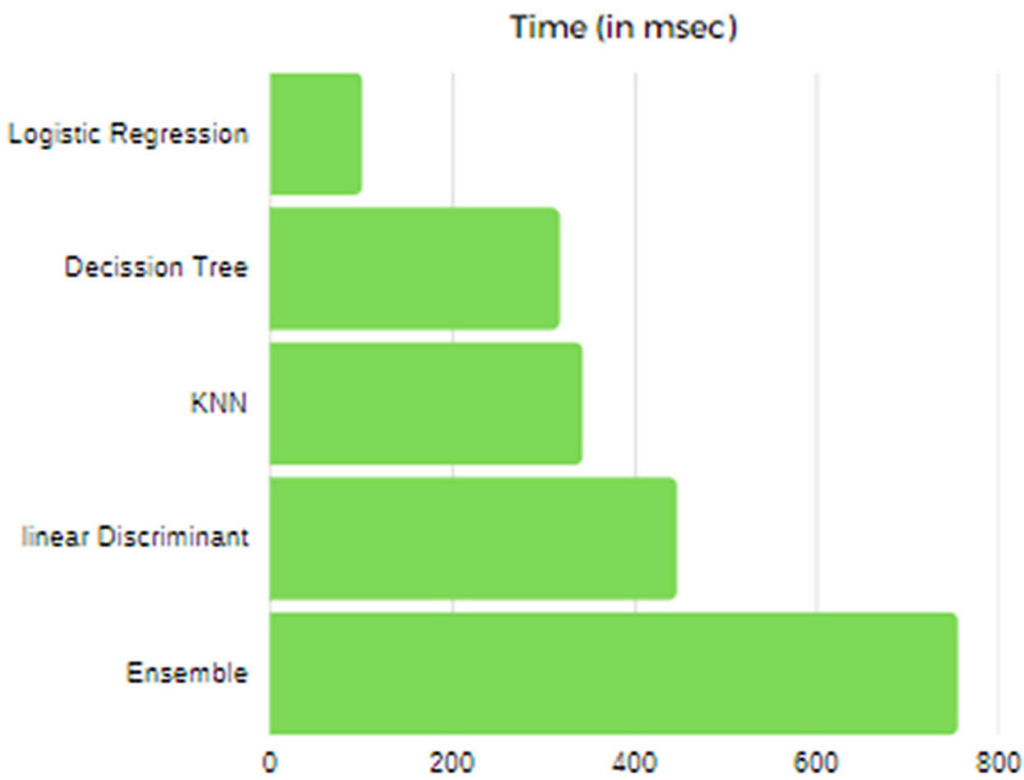

Fig. 6. Training comparison of dataset B – time

Using Dataset B, as shown in Figures 5 and 6, it is evident that the Decision Tree classifier shows the highest accurary in predicting the mortality in HD patients with diabetes, i.e.at the rate of 0.927, followed by Linear Discriminant of 0.879, Logistic Regression of 0.873, KKN of 0.853, and the last one is Ensemble of 0.839.

Table 2. Prediction result comparison of dataset B

Metrics	LR	DT	KNN	LD	En
Correct rate	0.893	0.944	0.880	0.901	0.893
Error rate	0.107	0.057	0.120	0.099	0.107
Sensitivity (TPR)	0.368	0.825	0.544	0.544	0.368
Specificity (TNR)	0.993	0.966	0.970	0.970	0.993
Positive Predictive Value	0.913	0.825	0.775	0.775	0.913
Negative Predictive Value	0.891	0.966	0.917	0.917	0.891
PositiveLikelihood	54.710	24.490	NaN	17.947	54.710
NegativeLikelihood	0.636	0.182	0	0.470	0.636
Prevalence	0.161	0.161	0.161	0.161	0.161
F1	0.693	0.825	0.693	0.693	0.693
Bookmaker Informedness (BM)	0.362	0.791	0.514	0.514	0.362
Markedness(MK)	0.804	0.791	0.692	0.692	0.804

Using Dataset B, as shown in Table 2, it is evident that the Decision Tree classifier shows the highest correct rate in predicting the mortality of hemodialysis patients with diabetes, i.e. at the rate of 0.944, followed by Linear Discriminant of 0.901, Logistic Regression and Ensemble of 0.893, and the last one is KKN of 0.880.

3.3 Prediction results of dataset C

After training on 170 hemodialysis patients in Dataset C, it was found that Logistic regression yielded accuracy of 0.818 achieved for 160.89 msec. Tree only gives accuracy 0.771. KNN and Ensemble each provide accuracy of 0.765, and 0.641, respectively. Figures 7 and 8 shows the training result of Dataset C, whereas Table 3 shows the performances comparison results of Dataset C.

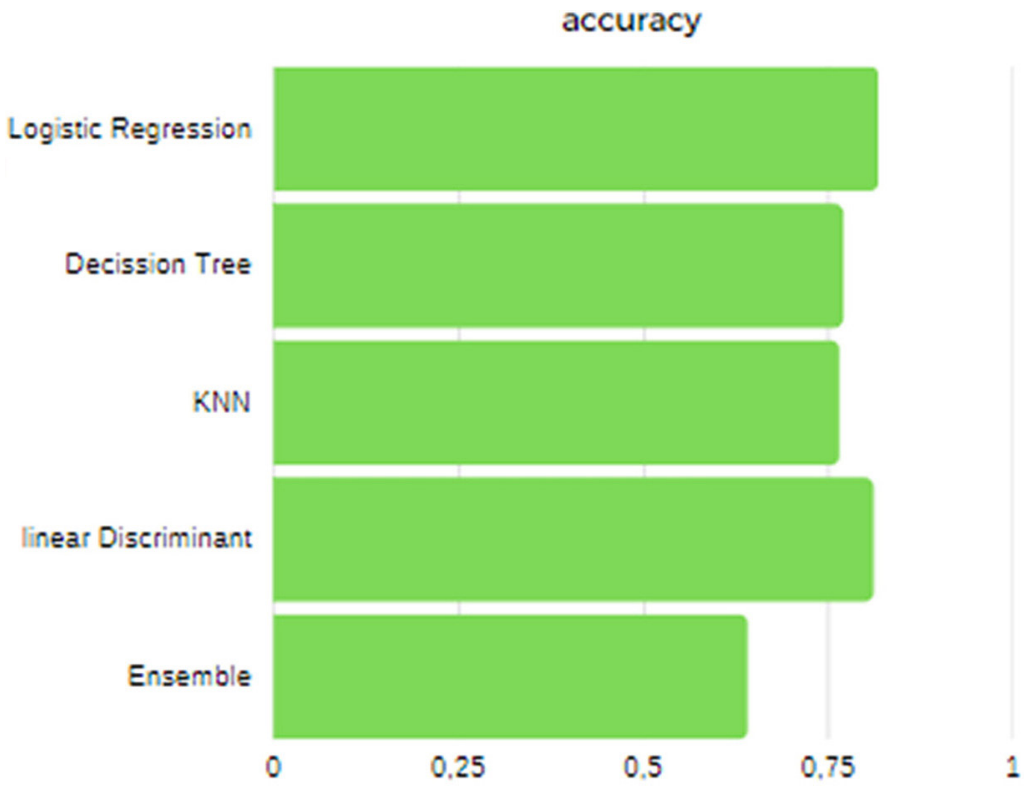

Fig. 7. Training comparison of dataset C – Accuracy

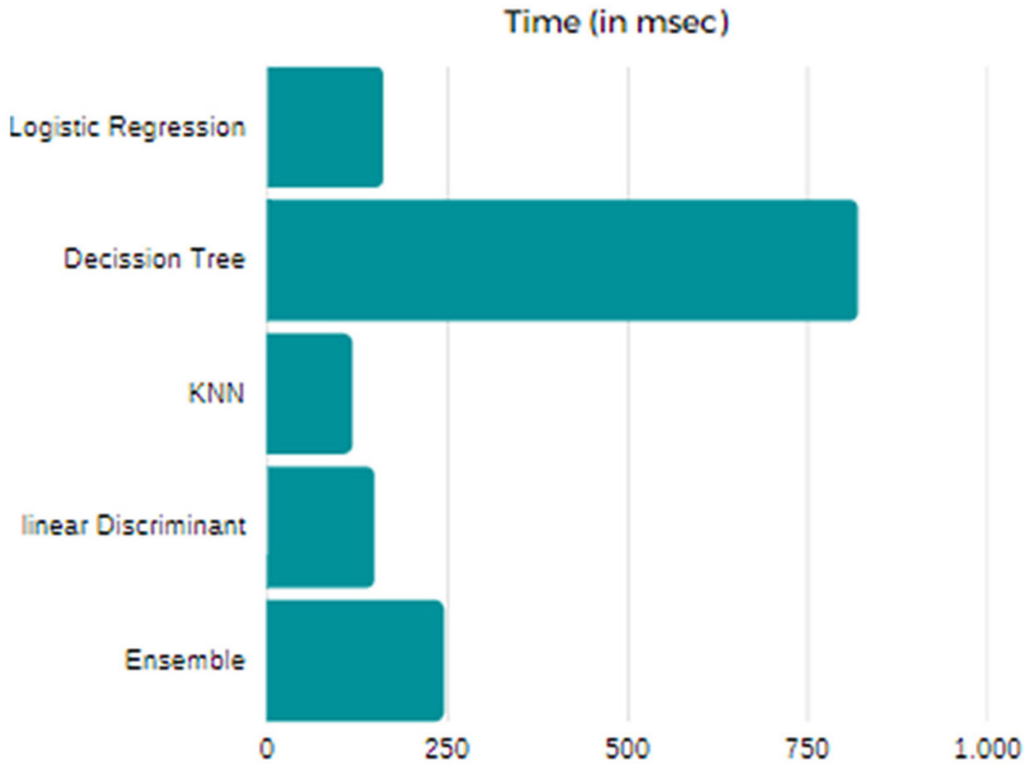

Fig. 8. Training comparison of dataset C – Time

Table 3. Prediction result comparison of dataset C

Metrics	LR	DT	KNN	LD	En
Correct rate	0.882	0.859	0.841	0.841	0.641
Error rate	0.118	0.141	0.159	0.159	0.359
Sensitivity (TPR)	0.803	0.689	0.803	0.721	0
Specificity (TNR)	0.927	0.954	0.927	0.908	1
Positive Predictive Value	0.860	0.894	0.860	0.815	NaN
Negative Predictive Value	0.894	0.846	0.894	0.853	0.641
Positive Likelihood	10.945	15.010	10.945	7.862	NaN
Negative Likelihood	0.212	0.3275	0.212	0.307	1
Prevalence	0.359	0.359	0.359	0.359	0.359
F1	0.831	0.778	0.831	0.765	NaN
Bookmaker Informedness (BM)	0.730	0.643	0.730	0.630	0
Markedness(MK)	0.753	0.739	0.753	0.668	NaN

3.4 Prediction results of dataset D

Figures 9 and 10 show Logistic Regression performance and comparison methods for the D dataset, respectively. After the training and the test turns out Logistic regression is still superior to its competitors except tree. Logistic regression provides training accuracy 0.613 and test accuracy 0.867. The required execution time is 529.88 msec, slower than Tree, KNN and Linear discriminant, but faster than tree and ensemble.

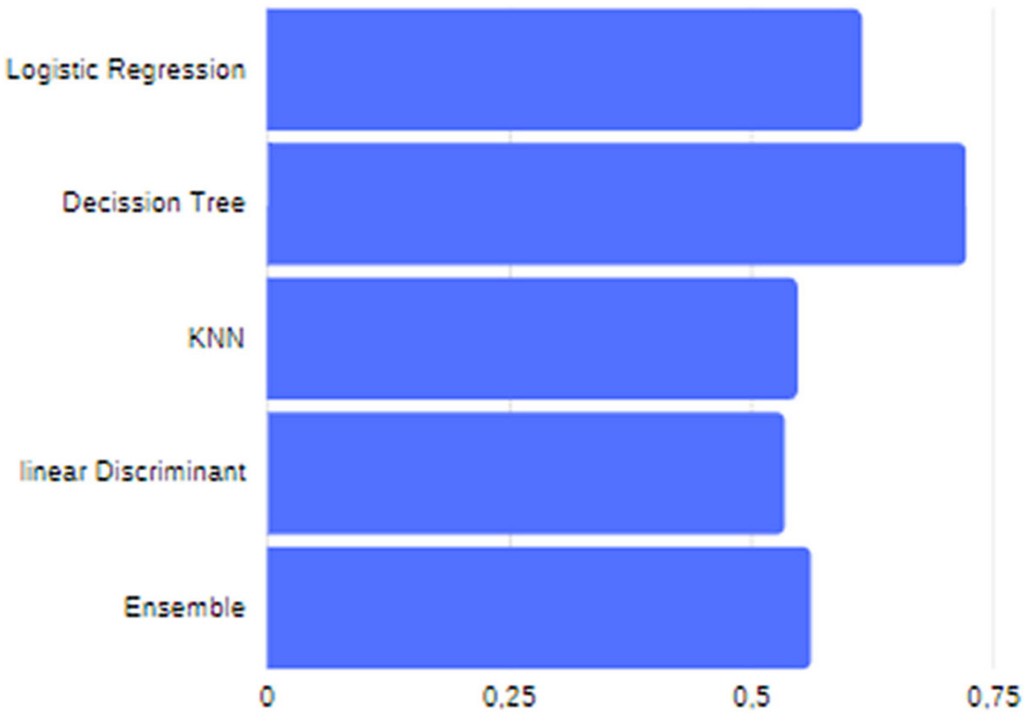

Fig. 9. Training comparison of dataset D – Accuracy

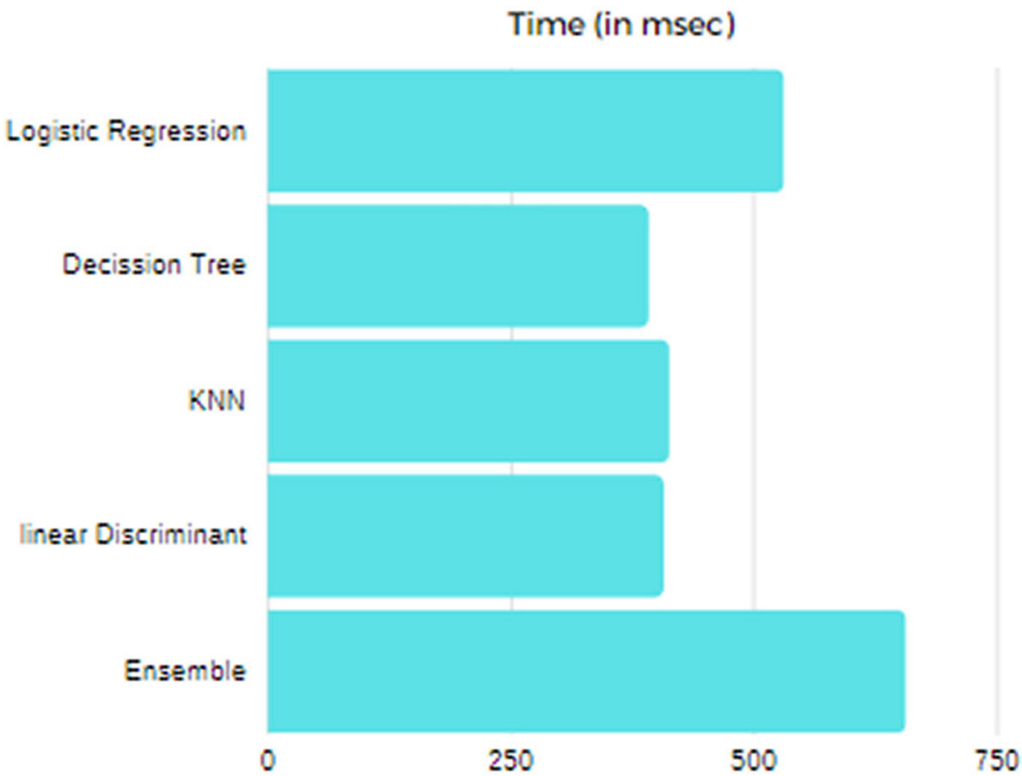

Fig. 10. Training comparison of dataset D – Time

Table 4. Prediction result comparison of dataset D

Metrics	LR	DT	KNN	LD	En
Correct rate	0.867	0.977	0.773	0.708	0.560
Error rate	0.133	0.023	0.227	0.293	0.440
Last Correct Rate	0.867	0.987	0.773	0.708	0.560
Last Error Rate	0.133	0.1330	0.227	0.293	0.440
Inconclusive Rate	0	0	0	0	0
Classified Rate	1	1	1	1	1
Sensitivity (TPR)	0.857	1	0.810	0.691	1
Specificity (TNR)	0.879	0.970	0.727	0.727	0
Positive Predictive Value	0.900	0.970	0.791	0.763	0.560
Negative Predictive Value	0.829	1	0.750	0.649	NaN
PositiveLikelihood	7.071	33	2.968	2.532	1
NegativeLikelihood	0.163	0	0.262	0.426	NaN
Prevalence	0.560	0.560	0.560	0.560	0.560
F1	0.878	0.985	0.800	0.725	NaN
Bookmaker Informedness (BM)	0.736	0.970	0.537	0.418	0
Markedness(MK)	0.729	0.970	0.541	0.412	NaN

From Table 4, it is evident that Logistic regression provides the second highest training accuracy for Dataset D in predicting the mortality in HD patients with diabetes in comparison to other algoritms being used below decision tree, i.e., 0.770, followed by Ensemble of 0.560, KKN of 0.547, and Linear Discriminant of 0.533.

Table 4 shows that Logistic Regression has the second highest level of truth, namely 0.987 under a decision tree of 0.977, the third position is KKN of 0.7733 followed by Linear Discriminant of 0.7076, and lastly is Ensemble of 0.56 as shown in Figure 11.

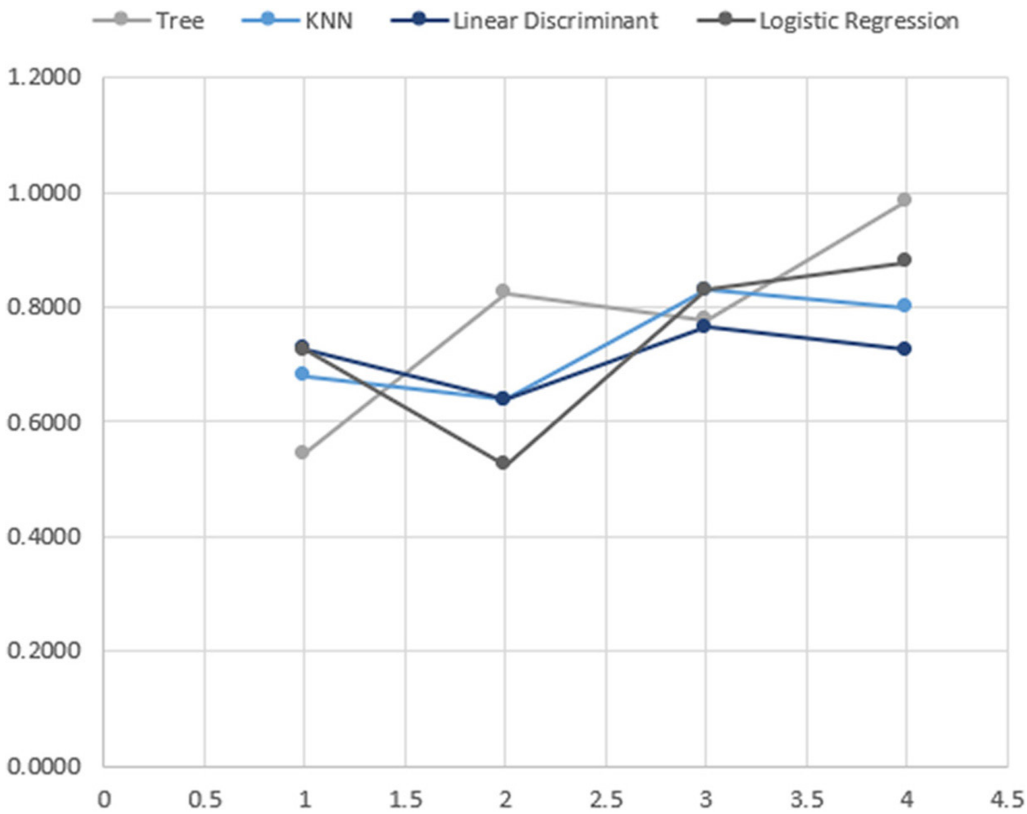

Fig. 11. F1 Score comparison

4 CONCLUSION

The Logistic Regression based mortality prediction in Hemodialysis patients was developed successfully. Compared to KNN, Tree, Ensemble, Linear discriminant and logistic regression, the Logistic Regression performs better in accuracy and small error rate. Proven by F1-score, the Logistic Regression give the highest scores for all test except for dataset B. The mortality in Hemodialysis patients is influenced by diabetes. Based on the results of experiments, it is evident that Logistic Regression has a good performance for prediction of mortality in Hemodialysis patients with diabetes.

5 REFERENCES

[1] B. Jia, D. Li, Z. Pan, and G. Hu, "Two-dimensional extreme learning machine," *Hindawi Math. Probl. Eng.*, vol. 2015, 2015. https://doi.org/10.1155/2015/491587

[2] M. Kumar, "Prediction of chronic kidney disease using random forest machine learning algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, pp. 24–33, 2016.

[3] J. M. Sung *et al.*, "Independent value of cardiac troponin T and left ventricular global longitudinal strain in predicting all-cause mortality among stable hemodialysis patients with preserved left ventricular ejection fraction," *Biomed Res. Int.*, vol. 2014, 2014, https://doi.org/10.1155/2014/217290
[4] S. S. Somji, P. Ruggajo, and S. Moledina, "Adequacy of hemodialysis and its associated factors among patients undergoing chronic hemodialysis in Dar es Salaam, Tanzania," *Hindawi Int. J. Nephrol.*, vol. 2020, 2020. https://doi.org/10.1155/2020/9863065
[5] N. Borisagar, D. Barad, and P. Raval, "Chronic kidney disease prediction using back propagation neural network algorithm," *Adv. Intell. Syst. Comput.*, vol. 508, pp. 295–303, 2017, https://doi.org/10.1007/978-981-10-2750-5_31
[6] T. B. Drüeke *et al.*, "Normalization of hemoglobin level in patients with chronic kidney disease and anemia," *N. Engl. J. Med.*, vol. 355, no. 20, pp. 2071–2084, 2006, https://doi.org/10.1056/NEJMoa062276
[7] X. Liu *et al.*, "A self-representation-based fuzzy SVM model for predicting vascular calcification of hemodialysis patients," *Comput. Math. Methods Med.*, vol. 2021, 2021, https://doi.org/10.1155/2021/2464821
[8] N. Ardi and Isnayanti, "Structural Equation Modelling-Partial Least Square to Determine the Correlation of Factors Affecting Poverty in Indonesian Provinces," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 846, no. 1, pp. 0–13, 2020, https://doi.org/10.1088/1757-899X/846/1/012054
[9] D. Deshwal, P. Sangwan, and D. Kumar, "A language identification system using hybrid features and back-propagation neural network," *Appl. Acoust.*, vol. 164, p. 107289, 2020, https://doi.org/10.1016/j.apacoust.2020.107289
[10] Y. S. Liu *et al.*, "Machine learning analysis of time-dependent features for predicting adverse events during hemodialysis therapy: Model development and validation study," *J. Med. Internet Res.*, vol. 23, no. 9, 2021, https://doi.org/10.2196/27098
[11] D. N. R. S. Ramya, "Diagnosis of chronic kidney disease using machine learning algorithms," *Technology*, vol. 4, no. 2, pp. 2257–2263, 2016.
[12] H. W. Kim, S. J. Heo, J. Y. Kim, A. Kim, C. M. Nam, and B. S. Kim, "Dialysis adequacy predictions using a machine learning method," *Sci. Rep.*, vol. 11, no. 1, 2021, https://doi.org/10.1038/s41598-021-94964-1
[13] D. Novaliendry, "Android-based network services application learning media for vocational high schools," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 20, pp. 83–100, 2021. https://doi.org/10.3991/ijim.v15i20.23745
[14] R. Lacson, "Predicting hemodialysis mortality utilizing blood pressure trends," *AMIA Annu. Symp. Proc.*, pp. 369–373, 2008.
[15] P. Escandell-Montero *et al.*, "Optimization of anemia treatment in hemodialysis patients via reinforcement learning," *Artif. Intell. Med.*, vol. 62, no. 1, pp. 47–60, 2014, https://doi.org/10.1016/j.artmed.2014.07.004
[16] V. Garcia-Montemayor *et al.*, "Predicting mortality in hemodialysis patients using machine learning analysis," *Clin. Kidney J.*, vol. 14, no. 5, pp. 1388–1395, 2021, https://doi.org/10.1093/ckj/sfaa126
[17] S. Patil, "A literature review on prediction of chronic kidney disease using data mining techniques," *Int. J. Sci. Res.*, vol. 6, no. 3, pp. 2015–2018, 2017.
[18] H. Mirinejad and T. Inanc, "RBF method for optimal control of drug administration in the anemia of hemodialysis patients," *2015 41st Annu. Northeast Biomed. Eng. Conf. NEBEC 2015*, 2015, https://doi.org/10.1109/NEBEC.2015.7117180
[19] P. Panwong and N. Iam-On, "Predicting transitional interval of kidney disease stages 3 to 5 using data mining method," *2016 2nd Asian Conf. Def. Technol. ACDT 2016*, pp. 145–150, 2016, https://doi.org/10.1109/ACDT.2016.7437659

[20] D. Novaliendry, Y. Hendriyani, C. H. Yang, and H. Hamimi, "The optimized K-means clustering algorithms to analyzed the budget revenue expenditure in Padang," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2, no. 1, pp. 61–66, 2015, https://doi.org/10.11591/eecsi.v2i1.771

[21] D. Novaliendry, Ahyanuardi, Irma Yulia Basri, Noper Ardi, Nathasyah Utami Hakim, Muhammad Fadilh Putra Pratama, and Nenny Mahyudin, "The development of professional competency certification assessment model for junior mobile programmers," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 8, pp. 181–196, 2023, https://doi.org/10.3991/ijim.v17i08.39213

[22] Dony Novaliendry, Asrul Huda, Latifah Annisa, Resti Rahmi Khairati Costa, Yudhistira, and Fivia Eliza, "The effectiveness of web-based mobile learning for mobile subjects on computers and basic networks in vocational high schools," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 9, pp. 20–30, 2023, https://doi.org/10.3991/ijim.v17i09.39337

[23] Z. Liu, Z. Liu, Z. Zhu, Y. Shen, and J. Dong, "Simulated annealing for a multi-level nurse rostering problem in hemodialysis service," *Appl. Soft Comput. J.*, vol. 64, pp. 148–160, 2018, https://doi.org/10.1016/j.asoc.2017.12.005

[24] J. Y. Yeh, T. H. Wu, and C. W. Tsao, "Using data mining techniques to predict hospitalization of hemodialysis patients," *Decis. Support Syst.*, vol. 50, no. 2, pp. 439–448, 2011, https://doi.org/10.1016/j.dss.2010.11.001

[25] S. P. Potharaju and M. Sreedevi, "Ensembled rule based classification algorithms for predicting imbalanced kidney disease data," *J. Eng. Sci. Technol. Rev.*, vol. 9, no. 5, pp. 201–207, 2016, https://doi.org/10.25103/jestr.095.31

6 AUTHORS

Dony Novaliendry, Universitas Negeri Padang, Padang, Indonesia; National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan.

Oktoria, Universitas Negeri Padang, Padang, Indonesia; National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan.

Cheng-Hong Yang, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan.

Yenny Desnelita, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia.

Irwan, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia.

Roni Sanjaya, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia.

Gustientiedina, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia.

Yaslinda Lizar, Universitas Islam Negeri Imam Bonjol, Padang, Indonesia.

Noper Ardi, Politeknik Negeri Batam, Batam, Indonesia.