Mathematics and Statistics 9(2): 135-143, 2021
DOI: 10.13189/ms.2021.090207

The Varying Threshold Values of Logistic Regression and Linear Discriminant for Classifying Fraudulent Firm

Samingun Handoyo[1,2,7,*], Ying-Ping Chen[3,4], Gugus Irianto[5], Agus Widodo[6]

[1]Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Malang 65145, Indonesia
[2]Department of EECS-International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[3]Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[4]Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan
[5]Department of Accounting, Faculty of Economics and Business, Brawijaya University, Malang 65145, Indonesia
[6]Department of Mathematics, Faculty of Mathematics and Natural Science, Brawijaya University, Malang 65145, Indonesia
[7]Department of EECS-International Graduate Program, National Chiao Tung University, Hsinchu 30010, Taiwan

Received January 3, 2021; Revised March 5, 2021; Accepted March 23, 2021

Cite This Paper in the following Citation Styles

(a): *[1] Samingun Handoyo, Ying-Ping Chen, Gugus Irianto, Agus Widodo , "The Varying Threshold Values of Logistic Regression and Linear Discriminant for Classifying Fraudulent Firm," Mathematics and Statistics, Vol. 9, No. 2, pp. 135-143, 2021. DOI: 10.13189/ms.2021.090207.*

(b): *Samingun Handoyo, Ying-Ping Chen, Gugus Irianto, Agus Widodo (2021). The Varying Threshold Values of Logistic Regression and Linear Discriminant for Classifying Fraudulent Firm. Mathematics and Statistics, 9(2), 135-143. DOI: 10.13189/ms.2021.090207.*

Abstract The aim of the research is to find the best performance both of logistic regression and linear discriminant which their threshold uses some various values. The performance tools used for evaluating classifier model are confusion matrix, precision-recall, F1 score and receiver operation characteristic (ROC) curve. The Audit-risk data set are used for the implementation of the proposed method. The screening data and dimension reduction by using principal component analysis (PCA) are the first step that must be conducted before the data are divided into the training and testing set. After the training process for obtaining the classifier model parameters has been completed, the calculation of performance measures is done only on the testing set where the various constants are added to the threshold value of both classifier models. The logistic regression classifier has the best performance of 94% on the precision-recall, 91.7% on the F1-score, and 0.906 on the area under curve (AUC) where the threshold values are on the interval between 0.002 and 0.018. On the other hand, the linear discriminant classifier has the best performance when the threshold value is 0.035 and its performance value is respectively the precision-recall of 94%, the F1-score of 91.7%, and the AUC of 0.846.

Keywords Classifier Model, Confusion Matrix, F1-Score, ROC Curve

1. Introduction

Machine Learning (ML) has a central role in processing data to be information or even to be knowledge. The most important characteristic in ML technique is the existence of a score function (objective or target function) that will be optimized by using an optimization method. The model developer has enough space in building model that meets a specification. In supervised learning, the type of response (target) variable will lead to a kind of suitable method specifically. When the response variable has an interval or a ratio measurement scale, so the matching analyses method is called regression technique. On the other hand, if the measurement scale of response variable is a categorical (nominal or ordinal), the suitable analyses method is called classification technique. Recently, implementation of a regression technique based on fuzzy logic for predicting of the time series data have be

conducted by Handoyo and Marji [1], Handoyo et al.[2], Handoyo and Chen [3], Efendy et al.[4]. Kusdarwati and Handoyo [5] also implemented the regression technique based on the hybrid Neural Network and wavelet. The classification of villages suffering dengue fewer uses the classification technique based on the fuzzy logic is also done by Handoyo and Kusdarwati [6]. The performance of the regression models based on the fuzzy logic above are very satisfied, but the performances of the classification models based on the fuzzy logic are a reasonable worse which it is not balance on the trade-off between it complicated in computation and it yielded in performance.

The logistic regression is a popular method in the classical statistics for analyzing categorical data where it is used for building of the model explaining of a causality relationship between the predictor and response variables. In the view point of predictive model, the logistic regression is applied for classification tasks such as works done by Worth and Cronin [7], Zhu and Hastie [8], Khairunnahar et al. [9], Algamal and Lee [10], Halushchak et al. [11] whose applied it in the medical field. Another hand, some researchers are more interested in the performance comparing between the logistic regression and another classifier model such as support vector machine (SVM) [12-13], learning vector quantization (LVQ) [14] where the logistic regression performance is not always poorer than the performance of either SVM or LVQ method.

Beside the logistic regression, the linear classification task can also be conducted by using the linear discriminant analysis (LDA) classifier model. Although the LDA is also popular for a dimension reduction task, its performance for a classification task is also very satisfied such as in the works done by Jia et al. [15], and Al-Dulaimi et al. [16]. Even recently, some efforts for improving of the LDA performance by the reformulating of LDA objective function through a regularization technique [17]. Otherwise, some researchers also improved the LDA algorithm through hybrid with a deep learning algorithm [18], and also hybrid with probabilistic mixture model [19]. The efforts increase significantly the LDA classifier performance, but there is a trade-off that must be paid ie the yielding complex and sophisticated model.

The critical tool in the choosing the best classifier models is the accuracy measurement used for the evaluating its performance. There are some measurements that usually use in the choosing best one such as the recall-precision, the AUC and also the F1- score [20-22]. A comprehensive discussion related to a model performance measures can also be found in either Tharwat [23] or Silva and Eugenio [24]. In order to avoid the complicated and sophisticated model which is derived from either the logistic regression or LDA, and also for exploring them for a classification purposes, this research has the aims to give a treatment and to explore a way of the choosing best performance of both model types. The treatment is done by varying threshold values and the exploring choice of the best model performance is done by comparing some performance measures including confusion matrix, ROC, recall-precision, and F1-score. The treatment is inspired by the works of Kusdarwati and Handoyo [25] whose done a threshold modelling in time series data where in the classifier model, a threshold has the main role in the determining decision boundary between both classes.

The next section will be discussed the proposed method theoretically in detail. A brief description of the data set, data pre-processing before the data are divided into training and testing set discuss in the section 3. The software used for the implementation purposes is the Anaconda_3 with the python 3.7 version. The implementation results and discussion are presented in the section 4. Finally, conclusion and remark are given in the section 5.

2. Proposed Method

The binary (bi-class) classification model of a logistic regression has a classification function stated such as eq.1 as follows

$$h(z) = \frac{1}{1+e^{-z}} \text{ , where}$$

$$z = w_0 + w_1 * x_1 + w_2 * x_2 + .. + w_n * x_n \quad (1)$$

It is known that h(z) is a sigmoid function which has the function values in the range from 0 to 1. In the point of view as a classifier model, on the logistic regression, the relationship between a response variable of y and predictor variables of X are stated in the form of y = h(z)). The y variable has 2 possible values that are 0 or 1. Because h(z) is a nonlinear function among the predictor variables, the least squared method cannot be used to obtain the estimate parameters.

For the purpose of estimate parameters, it is defined a cost function stated in the eq.2 as follows

$$J(w) = \frac{1}{m}\sum_{i=1}^{m} cost(h_w(x^i), y^i) \quad (2)$$

Where the cost() value of each class is defined as follows

$$cost(h_w(x), y) = -log\left(h_w(x)\right), when\ y = 1$$

$$cost(h_w(x), y) = -log\left(1 - h_w(x)\right), when\ y = 0 \quad (3)$$

By substituting of the eq.(3) to eq.(2), the objective function can be written in the eq. (4) as follows

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}\left[y^i log\left(h_w(x^i)\right) + (1 - y^i) log\left(1 - h_w(x^i)\right)\right]$$

Which it can be stated in a vector term as follows

$$J(w) = \frac{1}{m}.\left(-y^T log(h) - (1-y)^T log(1-h)\right) \quad (4)$$

The optimal coefficient of w can be obtained by using the gradient descent algorithm [12]. The decision boundary separated between two classes is obtained by taking h(z) in eq.1 equal to the threshold value which it is conventionally equal to 0.5 which means that z is 0.

$$w_1 * x_2 + w_2 * x_2 + .. + w_n * x_n = 0$$

In the case of two features, the decision boundary can be plotted where the both y and x axes as follows

$$y = x_2' = -\frac{w_1 * x_1' + w_0}{w_2} \text{ , and } \quad x = x_1' = \frac{max}{min}(x_1) \pm 2$$

The LDA can be used for handling of two different tasks that are either as dimension reduction or object classification technique [19]. The main difference between PCA and LDA is the availability of the target or response variable when LDA is applied as dimension reduction method but it is not when PCA is applied on the same task. The LDA is popular as a classification method either on binary or on multi classes of classification problem. In this study, it is only discussed a binary classes of classification problem. The LDA classifier has the form as follows

$$g(x) = w^T x + w_0 \quad (5)$$

The decision boundary (a line) separated between two classes is obtained when the classification function is equal to 0. To simplify notation, it is applied a notation changes to incorporate a constant term of $g(x)$:

$$g(x) = \mathbf{w}^T\mathbf{x} \quad \text{with} \quad \mathbf{w} = [w \quad w_0]^T, \ \mathbf{x} = [x \quad 1]^T$$

It considers that there is a set of training data, in the form of inputs x_i and corresponding to the actual output d_i (for binary classes of classification, It will be set that d_i equal to 0 or 1, where x_1^T and x_2^T are feature vectors of the class 0 and class 1 respectively. It can be summarized that both of $x_1^T w = d_1$ and $x_2^T w = d_2$ can be expressed as $Xw = d$. By multiplying with the X invers to the both sides, it will be obtained the formula as follows

$$w = X^{-1}d,$$

where $X = [x_1, x_2, \dots, x_N]^T$ and $d = [d_1, d_2, \dots, d_N]^T$

The computation of the X^{-1} is not directly taken from the feature matrix data because the X matrix is not a squared matrix. Fortunately, the LDA classifier model is a linear model where the mean squared error can be used as a cost function that will be optimized for obtaining the estimate parameters. The following is the elaboration process in the finding of the estimate parameters that can be done by using of the ordinary least square method.

Consider $E(w)$ as a cost function which it will be minimized for obtaining w such as presented in the eq. (6) as follows

$$E(w) = \sum_{i=1}^{N}(d_i - x_i^T w)^2 = \|d - Xw\|^2 \quad (6)$$

The first derivative of eq. 6 with respect to w then it equals to 0 is

$$\frac{\partial E(w)}{\partial w} = 2X^T(Xw - d) = 0 \text{ , and then}$$

$$w = X^+ d = (X^T X)^{-1} X^T d$$

$X^+ \equiv (X^T X)^{-1} X^T$ called the *pseudo-inverse* of X.

The classification decision rule is

if $g_1(x) - g_2(x) \geq 0$ then an object is classified to be the class 0, otherwise, it is classified to be the class 1.

The research purposes a method for finding of the best classifier model through varying threshold by adding a constant value. Consider the classification decision rule of both models as follows

IF$h(x) - 0.5 \geq 0$, THEN an object is classified into the class 1, otherwise it will be classified into the class 0 for a logistic regression classifier, and

IF$g_1(x) - g_0(x) \geq 0$, THEN an object is classified into the class 1, otherwise it will be classified into the class 0 for LDA classifier.

A constant value will be added to the right hand side of the inequality of each classifier model. It means that the treatment is given to the classification decision rule. It will construct a different classifier model with the previous one. In other words, if a classifier model is added a constant value in the way above, it will become a new classifier model.

3. Data and Research Method

3.1. Research Variables

The data set called the Audit risk data are obtained in Hooda et al. [26]. The data are collected through a survey and are taken from the UCI machine learning repository. The firms come from various sectors. The information about the sectors and the counts of firms are listed respectively including Irrigation (114), Public Health (77), Buildings and Roads (82), Forest (70), Corporate (47), Animal Husbandry (95), Communication (1), Electrical (4), Land (5), Science and Technology (3), Tourism (1), Fisheries (41), Industries (37), and Agriculture (200) where the total of observed objects is 777 firms. The response variable is the risk classified as the fraudulent firm (class 1) or non-fraudulent firm (class 0) and the predictor variables consist of 26 features. The complete of the response and predictor variables followed by their data types are as follows

Table 1. The feature names and its data types

No	feature and data type		no	feature and data type	
1	sector_score	float64	15	risk_d	float64
2	location_id	object	16	district_loss	int64
3	para_a	float64	17	prob	float64
4	score_a	float64	18	risk_e	float64
5	risk_a	float64	19	history	int64
6	para_b	float64	20	prob	float64
7	score_b	float64	21	risk_f	float64
8	risk_b	float64	22	score	float64
9	total	float64	23	inherent_risk	float64
10	numbers	float64	24	control_risk	float64
11	score_b.1	float64	25	detection_risk	float64
12	risk_c	float64	26	audit_risk	float64
13	money_value	float64	27	risk	int64
14	score_mv	float64			

The features of the number 1 to 26 are the predictor variables, and the last feature (number 27) is the response variable.

3.2. The Stages in Data Analyses

The following is given the steps in the computation process shortly

1) Preparing data in order to ready for model building that includes
   a. data screening
   b. dimension reduction uses PCA.
2) Dividing data set as the training and testing data.
3) Training classifier model uses the training data
   a. logistic regression by using gradient descent method
   b. LDA by using ordinary least square method.
4) Testing classifier model uses the testing data
   a. determining of the various constant values that are added to the threshold
   b. calculating of the confusion matrices
   c. calculating of the true positive rate (TPR) and false negative rate (FNR)
   d. calculating of the accuracy measures of both precision and F1-score.
5) Drawing of the ROC curve of both classifier models.

The data analysis of the research uses Python version 3.7 on the environment of Anaconda 3.

4. Results and Discussion

4.1. The Preparing Data for Building Classifier Model

There are two kinds of problems that can be identified from the Audit risk data set. The second column (the location ID feature) has the object data type (the column must be dropped) and the record that has the NaN observed value (the record also must be removed). PCA is applied to all of the screened input variables and then the transformed predictor variables and the original of its response variable are divided into the training and testing set.

The first five of principal components have explained the variance of 99.98% where their singular values (lambda) equal to *[0.6324, 0.3188, 0.0338, 0.0121, 0.0028]*. It means that the 4^{th} and 5^{th} principal components contribute to the explained variance only about 1.49%. It is reasonable when this study only considers the first 3 of principal components and the explained variance has reached about 98.5%.

The testing set are taken proportionally from each class in the data set by using stratified random sampling based on the data set frequency distribution. In this research, the number of testing set is determined as many 100 records (firms). Because there are 775 records in the data set consisting of the class 0 as many as 470 (61%) records, and the class 1 as many as 305 (0.39) records, the testing set are taken randomly as many as 61 records from the class 0 and as many as 39 records from the class 1. The remaining records of the data set are as the training set.

4.2. The Logistic Regression Classifier Model

In the development of the Logistic regression classifier, there is a primary component called the sigmoid function that has evaluated function values in the ranges of [0,1]. The classification decision rule with respect to an object is as follows.

If its value on the sigmoid function output is greater than 0.5, it is classified into the class 1, otherwise, it will be the member of the class 0. The training process of the logistic regression classifier is to obtain the exponent coefficients called the weights that are calculated through training process using the training set by using the gradient descent algorithm. So the main task of the training process is to obtain the optimal weights via an optimization method such as gradient descent.

The gradient descent method needs some setting parameters including learning rate, epoch number and tolerated error. The tuning parameters of a training model using gradient descent are relatively a hard task. There are two kinds of gradient descent algorithm that are stochastic and mini batch gradient descent. In the research, it is used stochastic gradient descent where the setting parameters of the training model are the learning rate = 00001, the epoch numbers =30, and the tolerated error is 0.00001. After the finishing of training process, it is obtained the optimal weights of [w0, w1, w2, w3] = *[1.15987 0.113263 0.02521 -0.02059]* which the weights are w0 as the constant part, w1 as the first coefficient, w2 as the second coefficient, and w3 as the third coefficient of principal component. The threshold is a value that separates the data to be two classes. In the logistic

regression, the default threshold is 0.5 that it means when the probability value of an object given some its features and the set of weights is lower than 0.5, the object will be the member of the class 0, and otherwise it will be the member of the class 1.

In the research, in order to find the best model, a treatment is given to the model yielded of the training process with adding of a constant value varied on the range of [0.00, 0.45] with a step increasing of 0.02 to the original threshold. The added constant equals to 0.00 that means the original logistic regression model where the threshold equal to 0.5, and the added constant equals to 0.02 that means the new model with the threshold equals to 0.52. The evaluation of model performance on the testing set with all of the new thresholds is presented in the form of confusion matrix or contingency table given in the Table 2. Based on the constant values added to the threshold, each new threshold produces one classifier model so the total of produced classifier model is 23. The best logistic regression classifier model will be chosen according to the highest performance measures on the both of precision value and F1-score criteria.

Table 2 presents the confusion matrices, the pairs of TPR and 1- FNR, and also the added values to the threshold. The model developer always expects to produce a classification model which has a capability to classify with highest accuracy. The classifier model is expected that it is able to place the objects corresponding to their actual classes. In general, the accuracy of classification model is calculated based on the proportion between the precisely classified objects and the total of all objects classified by the classification model. A confusion matrix is an instrument that is able to summarize in a simple way of the outputs of classification model. The elements of the main diagonal of a confusion matrix describes where the number of objects classified correctly, while the elements in other cells are the number of objects that is classified incorrectly. For example, in the case of binary classification (class of 0 or 1), the element (1,1) or above left element of a confusion matrix is the number of objects of the class 0 correctly classified, the element (1,2) or above right element of a confusion matrix is the number of objects which they originally come from the class 0, but they are classified at the class 1, the element (2,1) of a confusion matrix is the number of objects which they originally come from class 1, but they are classified at the class 0, and the element (2,2) is the number of objects from class 1 which they are correctly classified.

Table 2. The Confusion matrix, the pairs of TPR and 1-FNR, and the added threshold values of the Logistic Regression classifier model

No.	Confusion Matrix	Pairs(TPR, 1-FNR)	Added threshold
1	*[array([[61., 0.],[7., 32.]]),*	*[[0.89706 1.]*	*0*
2	*array([[61., 0.],[6., 33.]]),*	*[0.91045 1.]*	*0.002*
3	*array([[61., 0.],[6., 33.]]),*	*[0.91045 1.]*	*0.004*
4	*array([[61., 0.],[6., 33.]]),*	*[0.91045 1.]*	*0.006*
5	*array([[61., 0.], [6., 33.]]),*	*[0.91045 1.]*	*0.008*
6	*array([[61., 0.], [6., 33.]]),*	*[0.91045 1.]*	*0.01*
7	*array([[61., 0.], [6., 33.]]),*	*[0.91045 1.]*	*0.012*
8	***array([[61., 0.],[6., 33.]]),***	***[0.91045 1.]***	***0.014***
9	*array([[61., 0.],[6., 33.]]),*	*[0.91045 1.]*	*0.016*
10	*array([[61., 0.], [6., 33.]]),*	*[0.91045 1.]*	*0.018*
11	*array([[60., 1.], [6., 33.]]),*	*[0.90909 0.97059]*	*0.02*
12	*array([[60., 1.],[6., 33.]]),*	*[0.90909 0.97059]*	*0.022*
13	*array([[59., 2.], [6., 33.]]),*	*[0.90769 0.94286]*	*0.024*
14	*array([[59., 2.], [6., 33.]]),*	*[0.90769 0.94286]*	*0.026*
15	*array([[56., 5.],[6., 33.]]),*	*[0.90323 0.86842]*	*0.028*
16	*array([[54., 7.],[6., 33.]]),*	*[0.9 0.825]*	*0.03*
17	*array([[50., 11.], [6., 33.]]),*	*[0.89286 0.75]*	*0.032*
18	*array([[36., 25.], [5., 34.]]),*	*[0.87805 0.57627]*	*0.034*
19	*array([[3., 58.], [5., 34.]]),*	*[0.375 0.36957]*	*0.036*
20	*array([[0., 61.],[5., 34.]]),*	*[0. 0.35789]*	*0.038*
21	*array([[0., 61.],[4., 35.]]),*	*[0. 0.36458]*	*0.04*
22	*array([[0., 61.],[3., 36.]]),*	*[0. 0.37113]*	*0.042*
23	*array([[0., 61.], [2., 37.]])]*	*[0. 0.37755]]*	*0.044*

Consider the first row of the Table 2 which the constant value equals to 0, it means that there is no added constant value to the threshold. The confusion matrix has the element (1,1) of 61, and the element (1,2) of 0, it means that all of objects of the testing set with class 0 are predicted by the logistic regression classifier model with 100% correctly. In the second row of the confusion matrix, it can be seen that the element (2,1) of 7 and the element (2,2) of 32 which it means that there are 7 objects from the class 1 predicted incorrectly, and as many as 32 objects from the class 1 predicted correctly by the logistic regression classifier model which it has the type I error of 0%, and the type II error of 18% at the threshold value of 0 (without add a constant value to the original threshold).

The model performance is evaluated with both of precision and F1-score values. Both performance measure values are calculated by using the elements of the confusion matrix. The Table 3 presents the model performance measures on all of the threshold values.

Table 3. The Performance measures of the Logistic Regression on the both of Precision and F1-score

Measure	Values of the performance measures of logistic regression classifier model
Precision	*[0.930, 0.940, 0.940, 0.940, 0.940, 0.940, 0.940, 0.940, 0.940, 0.940, 0.930, 0.930, 0.920, 0.920, 0.890, 0.870, 0.830, 0.700, 0.370, 0.340]*
F1-score	*[0.901, 0.917, 0.917, 0.917, 0.917, 0.917, 0.917, 0.917, 0.917, 0.917, 0.904, 0.904, 0.892, 0.892, 0.857, 0.835, 0.795, 0.694, 0.519, 0.507]*

The accuracy classifier of logistic regression model with the added constant value to the threshold of 0 has reached 93% on the precision and 90.1% on the F1-score performance measures. The best performance of logistic regression model when the added constant value to the threshold is in the interval between 0.001 and 0.019 is respectively the accuracy of 94% on the precision and 91.7% on the F1-score performance measures. The results imply that the logistic regression classifier model can predict accurately 61 objects (all of the members of class 0), but there are 6 objects of the class 1 predicted incorrectly and 33 objects of the class 1 predicted correctly.

The receiver operation characteristic (ROC) curve of the logistic regression classifier model is presented on the Figure 1 and the area under curve (AUC) is very large that is 0.906.

Based on the Figure 1, it can be known that the logistic regression classifier model is not sensitive in the classifying object incorrectly. The ROC curve is very close with the vertical axes (the true positive rate axes) that indicates the classifier model performing well.

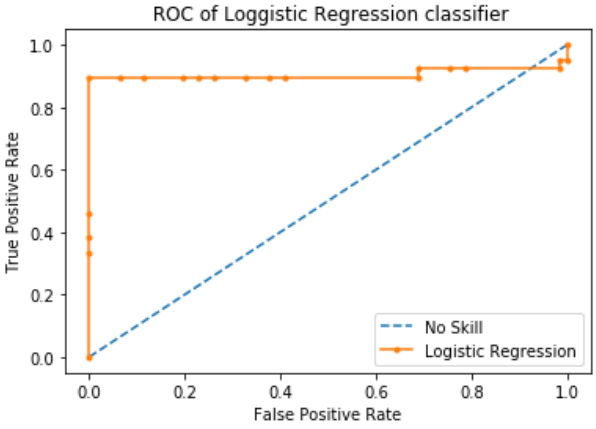

Figure 1. The ROC curve of the Logistic regression classifier model

4.3. The Linear Discriminant Classifier Model

The training of LDA classifier model is simpler than the training of logistic regression model. The coefficients of discriminant function are computed by using the pseudo-inverse matrix. The computation process is similar to the calculating of linear regression coefficients that can be solved analytically without need a numerical algorithm such as gradient descent. The coefficients of LDA function have an important role in the LDA classifier model because the matrix dot products between the coefficients and the testing data features (represented by a principal component matrix) will result the predicted classes of each object on the testing data. The coefficients are saved on a python object called the vector **w** = [w0, w1, w2, w3] = *[0.41140 0.00146 -0.00013 -0.00834]* where w0 is the intercept,, and w1, w2, w3 respectively are the first, second, third of principal component coefficient.

The next step, after the coefficients of LDA function are obtained, the performance of LDA classifier model is evaluated on the testing set. In this research, the added constant value to the threshold has the range from 0 to 0.1 with the increasing step of 0.005 or it can be written as [0.000, 0.100, 0.005]. Some results such as the confusion matrices, the TPR and FNR are produced when running program on the testing data are presented in the Table 4 as follows

Table 4. The Confusion matrices, the TPR and FNR pairs, and the threshold values of linear discriminant classifier model

No.	Confusion Matrix	Pairs(TPR,1-FNR)	Threshold added
1	*[array([[0., 61.],[6., 33.]]),*	*[[0. 0.35106]*	*0.000*
2	*array([[0., 61.],[6., 33.]]),*	*[0. 0.35106]*	*0.005*
3	*array([[11., 50.],[6., 33.]]),*	*[0.64706 0.39759]*	*0.010*
4	*array([[21., 40.],[6., 33.]]),*	*[0.77778 0.45205]*	*0.015*
5	*array([[28., 33.],[6., 33.]]),*	*[0.82353 0.5]*	*0.020*
6	*array([[57., 4.],[6., 33.]]),*	*[0.90476 0.89189]*	*0.025*
7	*array([[60., 1.],[6., 33.]]),*	*[0.90909 0.97059]*	*0.030*
8	***array([[61., 0.],[6., 33.]]),***	***[0.91045 1.]***	***0.035***
9	*array([[61., 0.],[8., 31.]]),*	*[0.88406 1.]*	*0.040*
10	*array([[61., 0.],[9., 30.]]),*	*[0.87143 1.]*	*0.045*
11	*array([[61., 0.],[9., 30.]]),*	*[0.87143 1.]*	*0.050*
12	*array([[61., 0.],[9., 30.]]),*	*[0.87143 1.]*	*0.055*
13	*array([[61., 0.],[9., 30.]]),*	*[0.87143 1.]*	*0.060*
14	*array([[61., 0.],[9., 30.]]),*	*[0.87143 1.]*	*0.065*
15	*array([[61., 0.],[10., 29.]]),*	*[0.85915 1.]*	*0.070*
16	*array([[61., 0.],[10., 29.]]),*	*[0.85915 1.]*	*0.075*
17	*array([[61., 0.],[11., 28.]]),*	*[0.84722 1.]*	*0.080*
18	*array([[61., 0.],[13., 26.]]),*	*[0.82432 1.]*	*0.085*
19	*array([[61., 0.],[14., 25.]]),*	*[0.81333 1.]*	*0.090*
20	*array([[61., 0.],[15., 24.]])]*	*[0.80263 1.]]*	*0.095*

Based on Table 4, when the added constant value to threshold equals to 0, the performance of classifier model is very poor which all of objects of the class 0 are classified incorrectly that is shown by the confusion matrix at the first row on the Table 4. By adding constant value with increasing step of 0.005, the performance of classifier also increases to be better than the previous one. The classifier performance is the best when the added constant value to the threshold is 0.035 at the row number 8 which its confusion matrix indicates that all of objects of the class 0 are classified correctly. This condition is 100% of contradiction when it is compared to the confusion matrix at the first row on the Table 4.

The linear discriminant classifier performance is easier to see by using the performance measures both of precision and F1- score values that are presented in the Table 5 as follows

Table 5. The Performance measures of LDA classifier on the both of precision and F1-score

Measure	The values of accuracy measure of LDA model
Precision	*[0.330, 0.330, 0.440, 0.540, 0.610, 0.900, 0.930, 0.940, 0.920, 0.910, 0.910, 0.910, 0.910, 0.910, 0.900, 0.900, 0.890, 0.870, 0.860, 0.850]*
F1 score	*[0.496, 0.496, 0.541, 0.589, 0.629, 0.868, 0.904, 0.917, 0.886, 0.870, 0.870, 0.870, 0.870, 0.870, 0.853, 0.853, 0.836, 0.800, 0.781, 0.762]*

In the Table 5, it is shown that both of precision and F1-score have respectively the highest value of 94% and 91.7% which they only accur in one time. It means that only one classifier model has the best performance when the added constant value to the threshold is 0.035. This condition is significantly different if it is compared to the previous one classifier model (logistic regression classifier) that has some added constant values ranging from 0.002 to 0.018 which its performance is optimal.

The performance of LDA classifier also is described by the ROC curve presented in the Figure 2 as follows

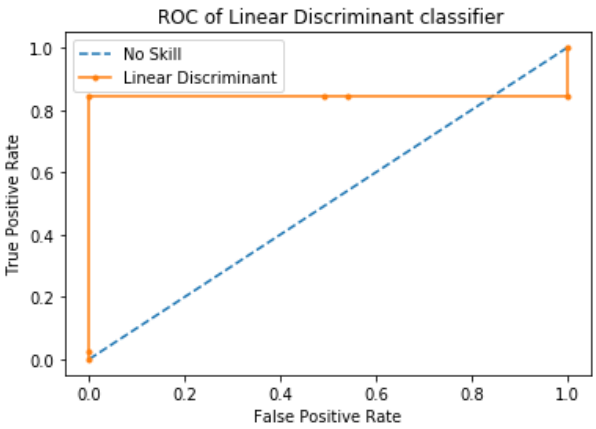

Figure 2. The ROC curve of LDA lassifier model

The LDA classifier has the AUC of 0.846. The ROC is also close to the TPR axes which it means the LDA

classifier has strong decision when it predicts the objects classified correctly. The probability of an object of positive class (the class 1) is classified correctly by the classifier model is high, in other words, the probability of a decision making has the type I error is low.

4.4. Discussion

The performances of both logistic regression and LDA have presented in the previous sub section. Both classifiers have the same as the best performance of 94% and 91.7% of the precision and the F1-score respectively. Nevertheless, the AUC of logistic regression classifier of 0.906 is higher than the AUC of LDA classifier of 0.846. The difference of their AUC shows that the logistic regression classifier has not improved significantly on its performance by adding a constant value to the threshold (varying threshold) where its performance increases only 1% (from 93% to 94%) in precision measures. On the other hand, the LDA classifier increases significantly (from 33% to 94%) in the precision measures. It is a reasonable result as a trade off between both classifier models. The training of logistic regression is harder than the training of LDA. The training logistic regression needs the setting parameters such as the tuning value of learning rate, epoch numbers, mini batch sizes and tolerated error. Specially, the setting of learning is done by trial and erros where a different problem needs a difference of learning rate. The logistic regression classifier is almost optimal classifier model without the treatment of varying its threshold value. The LDA classifier is easy or simple in the training process that is only involves multiplication and inversion of matrix, but it has a poor performance when the varying its threshold value is not done. The shape of ROC curve gives intuitive insight of the classifier model probability for making decision of the type I error. In the case of the data set used in this study, both of the classifier models have the low of type I error which it means that the probability of classifier model classifies the fraudulent firm incorrectly is low.

5. Conclusions

The implementation of logistic regression and linear discriminant classifier models is easy and is produced a simple model. The training process of logistic regression is harder than linear discriminant because it must involve the optimization technique such as gradient descent. On the other hand, the training process of linear discriminant just involves an inverse matrix multiplication. The logistic regression classifier has reached almost an optimal in performance without varying its threshold value, but the linear discriminant classifier for obtaining the best performance is affected by varying its threshold value significantly. In their best performance both models have the same as an accuracy measures on the precision-recall and the F1-score. The ROC-AUC of the logistic regression classifier is larger than the linear discriminant classifier. In the future research, it is an interesting work for investigating the effect of an adding regularization on the cost function to the performance of both classifiers.

REFERENCES

[1] S. Handoyo, Marji, The Fuzzy Inference System with Least Square Optimization for Time Series Forecasting, Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), Vol. 7, No. 3, pp. 1015-1026, 2018.

[2] S. Handoyo, Marji, I. N. Purwanto, F. Jie, The Fuzzy Inference System with Rule Bases Generated by using the Fuzzy C-Means to Predict Regional Minimum Wage in Indonesia, *International J. of Opers. and Quant. Management (IJOQM)*, Vol. 24, No. 4, pp.277-292, 2018.

[3] S. Handoyo, Y-P. Chen, The Developing of Fuzzy System for Multiple Time Series Forecasting with Generated Rule Bases and Optimized Consequence Part, *International Journal of Engineering Trends and Technology*, Vol. 68, No. 12, pp. 118-122, 2020.

[4] A. Efendi, S. Handoyo, A.P.S. Prasojo, and Marji, The Implementation Of The Optimal Rule Bases Generated By Hybrid Fuzzy C-Mean And Particle Swarm Optimization, *Journal of Theoretical & Applied Information Technology*, Vol. 97, No. 16, pp. 4453-4453, 2019.

[5] H. Kusdarwati, and S. Handoyo, System for Prediction of Non Stationary Time Series based on the Wavelet Radial Bases Function Neural Network Model, *Int J Elec & Comp Eng (IJECE)*, Vol. 8, No. 4, pp. 2327-2337, 2018.

[6] S. Handoyo, H. Kusdarwati, Implementation of Fuzzy Inference System for Classification of Dengue Fever on the villages in Malang, *In IOP Conference Series: Materials Science and Engineering*. Vol. 546, No. 5, pp. 052026, 2019.

[7] A.P. Worth, M.T. Cronin, The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects, *Journal of Molecular Structure: THEOCHEM*, Vol. 622, No. 1-2, pp. 97-111, 2003.

[8] J. Zhu, T. Hastie, Classification of gene microarrays by penalized logistic regression, *Biostatistics*, Vol. 5, No. 3, pp. 427-443, 2004.

[9] L. Khairunnahar, M.A. Hasib, R.H. Rezanur, M.R. Islam, M.K. Hosain, Classification of malignant and benign tissue with logistic regression, *Informatics in Medicine Unlocked*, Vol. 16, pp. 100189, 2019.

[10] Z.Y. Algamal, H.M. Lee, A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification, *Advances in data analysis and classification*, Vol.3, No. 3, pp. 753-771, 2019.

[11] I. Halushchak, Z. Novosad, Y. Tsizhma, & A. Zagorodnyuk, Logistic Map on the Ring of Multisets and Its Application in Economic Models, *Mathematics and Statistics*, Vol 8, No.

4, pp.424 – 429, 2020.
https://doi.org/10.13189/ms.2020.080408

[12] A. Widodo and S. Handoyo, The Classification Performance Using Logistic Regression And Support Vector Machine(Svm)., *Journal Of Theoretical & Applied Information Technology, Vol.* 95, No. 19, pp. 5184-5193, 2017.

[13] A. Robles-Velasco, P. Cortés, J. Muñuzuri, L. Onieva, Prediction of pipe failures in water supply networks using logistic regression and support vector classification, *Reliability Engineering & System Safety*, No. 196, pp. 06754, 2020.

[14] W.H. Nugroho, S. Handoyo, and Y.J. Akri, An Influence of Measurement Scale of Predictor Variable on Logistic Regression Modeling and Learning Vector Quantization Modeling for Object Classification, *International Journal of Electrical and Computer Engineering(IJECE)*, Vol. 8, No. 1, pp: 333-343, 2018.

[15] W. Jia, Y. Deng, C. Xin, X. Liu, W. Pedrycz, A classification algorithm with Linear Discriminant Analysis and Axiomatic Fuzzy Sets, *Mathematical Foundations of Computing*, Vol. 2, No. 1, pp. 73-85, 2019.

[16] K. Al-Dulaimi, V. Chandran, K. Nguyen, J. Banks, I. Tomeo-Reyes, Benchmarking HEp-2 specimen cells classification using linear discriminant analysis on higher order spectra features of cell shape, *Pattern Recognition Letters*, No. 125, pp. 534-41, 2019.

[17] R. Fu , Y. Tian , T. Bao, Z. Meng, P. Shi, Improvement motor imagery EEG classification based on regularized linear discriminant analysis, *Journal of medical systems*, Vol. 43, No. 6, pp. 147-169, 2019.

[18] T. L. Hayes, C. Kanan, Lifelong machine learning with deep streaming linear discriminant analysis, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 210-221, 2020.

[19] Y. Liu, J. Zeng, L. Xie, X. Lang, S. Luo, H. Su, An improved mixture robust probabilistic linear discriminant analyzer for fault classification, *ISA transactions*, Vol. 1; No. 98, pp.227-236, 2020.

[20] S.A. Khan, Z.A. Rana, Evaluating performance of software defect prediction models using area under precision-Recall curve (AUC-PR), *In 2019 2nd International Conference on Advancements in Computational Sciences (ICACS)*, pp. 1-6, 2019.

[21] H.R. Sofaer, J.A. Hoeting, C.S. Jarnevich, The area under the precision−recall curve as a performance metric for rare binary events, *Methods in Ecology and Evolution*, Vol. 10, No. 4, pp. 565-577, 2019.

[22] F. Emmert−Streib, S. Moutari, M. Dehmer, A comprehensive survey of error measures for evaluating binary decision making in data science, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 9, No. 5, pp. 1289-1303, 2019.

[23] A. Tharwat, Classification assessment methods, *Applied Computing and Informatics*, pp. 1-25, 2018.

[24] I. Silva, N. J. Eugenio, A Systematic Methodology to Evaluate Prediction Models for Driving Style Classification, *Sensors*, Vol. 20, No. 6, pp. 1679-1692, 2020.

[25] H. Kusdarwati, S. Handoyo, Modeling Treshold Liner in Transfer Function to Overcome Non Normality of the Errors, *In IOP Conference Series: Materials Science and Engineering*, Vol. 546, No. 5, pp. 052039, 2019.

[26] N. Hooda, S. Bawa, P.S. Rana, Fraudulent firm classification: a case study of an external audit, *Applied Artificial Intelligence*, Vol. 32, No. 1, pp. 48-64, 2018.