p-ISSN: 2252-3006
e-ISSN: 2685-2411

Website-Based Application for Classification of Diabetes Using Logistic Regression Method

Muhamad Soleh[a1], Naufal Ammar[a2], Indrati Sukmadi[a3]
[a]Program Studi Informatika, Institut Teknologi Indonesia, Indonesia
e-mail: [1]muhamad.soleh@iti.ac.id, [2]ammarhtr@gmail.com, [3]indrati.sukmadi@iti.ac.id

Abstrak

Pembelajaran mesin merupakan salah satu ilmu yang mempelajari tentang bagaimana komputer mampu belajar dari data untuk meningkatkan kecerdasannya. Machine learning terdiri dari banyak metode klasifikasi, antara lain Neural Network, Support Vector Machine, Logistics Regression, dan lain – lain. Pada penelitian ini dilakukan proses klasifikasi dengan menggunakan metode Logistics Regression untuk kasus penyakit Diabetes. Diabetes adalah kenaikan glukosa dalam aliran darah karena kekurangan insulin yang bertanggung jawab untuk transfer glukosa dari darah ke jaringan atau sel. Penelitian ini dibuat dengan tujuan untuk memperbaiki penelitian sebelumnya. Data yang digunakan pada penelitian ini yaitu data yang sama dengan penelitian sebelumnya yang diterbitkan oleh Pima Indian Diabetes Dataset. Pada penelitian ini digunakan beberapa tahapan yaitu pre processing, proses, evaluasi, serta pengembangan aplikasi berbasis website. Data pada penelitian ini dibagi menjadi dua yaitu 75% untuk data training, dan 25% untuk data testing. Penelitian ini menghasilkan evaluasi dengan nilai akurasi sebesar 80%, yang itu berarti lebih baik dengan penelitian sebelumnya yaitu sebesar 75, 97%.

Kata kunci: *Pembelajaran Mesin, Logistics Regression, Diabetes, Aplikasi, Website.*

Abstract

Machine learning is a one of computer science field, machine-learning studies how computers are able to learn from data to improve their intelligence. Machine learning consists of many classification methods, including Neural Networks, Support Vector Machines, Logistics Regression, and others. In this study, a classification process carried out using the Logistics Regression method for cases of Diabetes. Diabetes is an increase in glucose in the bloodstream due to a lack of insulin, which is responsible for the transfer of glucose from the blood to tissues or cells. This study created with the aim of improving previous paper. The data used in this study are the same data as previous studies published by the Pima Indian Diabetes Dataset. In this study, several stages used, those are pre-processing, processing, evaluation, and website-based application development. The data in this study divided into two, 75% for training data, and 25% for testing data. This study produces an evaluation with an accuracy 80%, which means it is better than the previous paper, which is 75, 97%.

Keywords: *Machine Learning, Logistics Regression, Diabetes, Application, Website.*

1. Introduction

Machine learning defined as computer applications and mathematical algorithms that adopted by learning derives from data and produces predictions in the future. The learning process to acquire intelligence that goes through two stages, training and testing. Build computer programs with the intelligence increases automatically based on experience [1].

Previous paperers have carried out paper on the use of machine learning in the health sector. As paper conducted by [2], this paper used Convolution Neural Network (CNN) method in classifying patterns and evaluated 120 cases from two different hospitals. This method produces a pattern more superior to other methods and faster in classifying diseases present in the lung with larger and clearer image result. In other paper conducted by [3] This study uses the forward chaining method used to develop a machine learning prototype for the prognosis of dementia, using literature data on examining patients who have been diagnosed with dementia, including examination of blood pressure, blood lipid levels, blood sugar levels, vesicular and

abdominal inspection. In this paper, successfully carried out and obtained by paperer, a prognosis solution for dementia according to optimal examination results. In another paper conducted by [4] in this paper using the KNN algorithm in solving problems, this study got an accuracy for distinguishing HC and MS subjects reaching 80%. In this paper included 111 subjects, 71 healthy controls from the Alzheimer's disease neuroimaging initiative (ADNI) database and used 40 patients with multiple sclerosis [5].

Health is something expensive and desired by humans. Therefore, every human being wants to maintain his health in order to enjoy life comfortably. According to the World Health Organization, in its constitution, it defines health as "a state of complete physical, mental and social health". Today, it continues as an all-encompassing definition of health. This becomes the standard ideal condition for every human being [6]. Unfortunately, many humans cannot maintain their health, many diseases have emerged and harm the condition of the body, even in today's modern era, and many people suffer from diseases that are at risk of causing death.

Diabetes is one of the diseases most feared by humans, because diabetes can harm parts of the body and can even cause death. Diabetes defined as a disease caused by the inability of the pancreas to produce insulin, which controls sugar in the blood. Ignoring and unchecked this desease, it will be dangerous because complications can occur such as heart disease, stroke, glaucoma or cancer. In Indonesia, based on data released by the World Health Organization (WHO) in 2016, there were 99,400 deaths caused by diabetes, of which 48,300 deaths occurred in productive age [7]. Diabetes ranks as one of the most chronic diseases in the world. Statistics show that in 2013 about 382.8 million people between the ages of 20 and 79 diagnosed with diabetes worldwide. During 2013, 4.6 million deaths occurred and cost $ 548 billion in medical expenses [8].

Paper conducted by [7] Logistic Regression Method, also known as Logistic Regression analysis, this method often used in various fields, such as data mining, automatic disease diagnosis, economic prediction and other fields. [8]. Data used comes from the Pima Indians Diabetes Dataset. Pima Indian data itself comes from the National Institute of Diabetes, Digestive, and Kidney Disease by taking samples of 768 women of Pima Indian descent aged 21 years and over. Of the 768 data samples, 500 samples not diagnosed with diabetes while 268 samples diagnosed positive for diabetes. The data in the study divided into 75% for training data and 25% for testing data. In this paper, it resulted in an accuracy of 75.97%. Based on those problems, this paper conduct Website-Based Application for Classification of Diabetes Using Logistic Regression Method was conduct by the author based on those problems.

In this paper, the process of oversampling data carried out using the One Point Cross over technique. The addition of data carried out with the aim of balancing the number of patients detected with diabetes and non-diabetes. Data addition was done by taking 232 samples of patients with detected diabetes, then from these 232 samples a new sample was generated through the One Point Cross Over process, so that the number of samples became 1000, with 500 samples detected diabetes and 500 samples not diabetes. This oversampling technique expected to increase the accuracy of the logistic regression model, as done in this paper [10].

2. Reseach methodology

The reseach methodology used in completing the paper, which consists of:

1. Data collection / Acquaring
   Retrieve and collect diabetes classification data. The data is secondary data taken from the internet.
2. Data Pre-porcessing
   In the pre-processing stage, carried out several steps. Missing value process, correlation coefficient analysis, data cleaning process, data addition process and data partitioning process.
3. Requirement Analysis and Software Design
   Analyze system requirements and identify information needs based on observations. The system implemented using the Python programming language.
4. Implementation
   Implement logistic regression algorithms
5. Testing and Evaluation

p-ISSN: 2252-3006
e-ISSN: 2685-2411

The test includes testing the diabetes classification using Logistic Regression method. The testing results analyzed to obtain conclusions from the entire paper series.

2.1. Dataset

The dataset process by retrieving diabetes medical examination data from the internet. Data obtained from [11]. The data called "Pima Indian Diabetes Databases" published by the "National Institute of Diabetes and Digestive and Kidney Diseases". The data nine attributes and 768 samples. Of the 768 sample data, there are 500 samples of undetectable diabetes and 268 samples of detected diabetes.

Pima Indian data took samples of 768 people of Pima Indian descent between the ages of 21 years and over. The Pima Indian data has nine attributes. These are Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Body Mass Index (BMI), Diabetes Pedigree Function, Age, and Outcome.

2.2. Data Pre Processing

In the pre-processing stage, carried out several steps. Missing value process, correlation coefficient analysis, data cleaning process, data addition process and data partitioning process. Mising value can be handle in several ways, in previous paper the detected data for the mising value replaced with a value of zero, in this paper, mising value was handle by replacing the data with the average value of each attribute. The data cleaning process carried out with the aim of deleting columns / attributes with the least correlation, the deleted attributes based on the correlation value of each data attribute of the independent variable to the dependent variable. Analysis of the correlation coefficient divided into several types, in this paper the Pearson correlation coefficient analysis used as in equation 1:

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}} \quad (1)$$

r = Pearson Correlation
n = Number of Data
X = Independent Variable
Y = Dependent Variable

Table 1. Pearsons Correlation Result

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
Pregnancies	1	0,149021347	0,24633527	0,033918637	-0,01811	0,08054	-0,01615	0,5382	0,24547
Glucose	0,14902135	1	0,219765132	0,158060093	0,396137	0,23146	0,13716	0,2667	0,49288
BloodPressure	0,24633527	0,219765132	1	0,130403192	0,010492	0,28122	0,00047	0,3268	0,16288
SkinThickness	0,03391864	0,158060093	0,130403192	1	0,24541	0,53255	0,1572	0,0206	0,17186
Insulin	-0,0181124	0,396136563	0,010491597	0,245410498	1	0,18992	0,15824	0,0377	0,1787
BMI	0,08053802	0,231463748	0,281221732	0,532551906	0,189919	1	0,15351	0,0257	0,31225
DiabetesPedigree Function	-0,01615088	0,137158309	0,000471125	0,157196074	0,158243	0,15351	1	0,0336	0,17384
Age	0,53816885	0,266673434	0,326791303	0,020582297	0,037676	0,02575	0,03356	1	0,23836
Outcome	0,24546646	0,492884103	0,162879099	0,171856814	0,178696	0,31225	0,17384	0,2384	1

From table 1, it known the lowest order of correlation for each variable on the dependent variable. In this paper, two independent variables with the lowest correlation value removed, therefore the Blood Pressure and Skin Thickness variables removed because they had low correlation values. Feature Selection is a method used to optimize the performance of the classifier. The way it works based on a large reduction in feature space, namely by eliminating the less relevant attributes and by using the feature selection algorithm to increase accuracy [12].

In this paper, "One Point Cross Over" technique was used to increase the amount of data as needed. There were 500 target data samples that had a value of 0 and 268 samples that had a value of 1. In order for the number of target data values to be the same, 500 data is 0

p-ISSN: 2252-3006
e-ISSN: 2685-2411

and 500 data is 1, it is necessary to do the One Point Cross Over technique. One Point Crossover is a data exchange technique that carried out by exchanging genes from one chromosome with another chromosome to produce new chromosomes through one intersection point [13]. The cut point is obtained by generating a random number with a limit of 1 to n (chromosome length. The resulting random number used as the chromosome intersection point. For example, two chromosomes have a length of 6 and the resulting random number is 4, then genes 1 to 4 will be cut with genes 5 to 6. Then, genes 1 to 4 on the first chromosome will be crossovered with genes 5 and 6 on the second chromosome, and vice versa.

Figure 1. *One Point Crossover*

The process of partitioning data in this paper carried out in two stages. The first was dividing the data into two parts, the independent variable or feature data, which usually symbolized by X, and the dependent variable or target data, which usually symbolized by Y. After dividing the independent and dependent variables, then the dependent and independent variable data are partitioned into two parts, training data and testing data, in this study 75% training data and 25% testing data were used, so that the data distribution became X training, Y training, and X testing, Y testing.

2.3. Logistic Regression Process

The process of Logistic Regression is divided into two stages, the first is to find the value of weights and constants and the second stage is to evaluate the Logistic Regression model with the weights and constants that have been found.

2.3.1. Finding the Weight Value and Constant Value

The essence of linearly separable based machine learning algorithms is to find the weight and constant values in the equation of the line. To get the weight and constant values from training data used with equation 2

$$f(x) = w\theta + w1 * x1 + w2 * x2 + \cdots wn * xn \quad (2)$$

$f(x)$ = logistic regression function
$w\theta$ = constant value
xi = independent variable
wi = wight value

Then initialize the weights and constants, perform mathematical operations according to the logistix regression equation. After getting the results, perform the sigmoid function operation with the formula 3

$$y' = \frac{1}{1+e^{-f(x)}} \quad (3)$$

y' = sigmoid function as predicted value
$f(x)$ = logistic regression function

Then calculate the error value with formula 4 as follows:

$$Error = y - y' \quad (4)$$

y : *dependent variable as true value*

p-ISSN: 2252-3006
e-ISSN: 2685-2411

y' : sigmoid function as predicted value

update weights and constants value using equation 5 - 7:

$$coeff\_update = (y - y') * y' * (1 - y') \quad (5)$$

$$w\theta baru = w\theta + \propto * coef\_update \quad (6)$$

$w\theta baru$: updated constants value
$\propto$: *learning rate* (set as 0,3)
$w\theta$: initial constants value
$coef\_update$: coefficient *update*

$$wbaru = w + \propto * coef\_update * x \quad (7)$$

$wbaru$: updated weight value
w : initial weight value
x : *independent variable*

Repeat the process of finding the value of weights and constants and the iteration process will stop at the stoping criteria if the error value has reached 0.0001 and the maximum number of iterations in this research is 1000 iterations.

2.2.2 Evaluasi Logistics Regression

The evaluation of the logistic regression performance in this research uses a Confusion matrix. Confusion matrix is a tabular representation of the actual and predicted values of data as shown in Table 2 [14].

Tabel 2. *Confusion Matrix*

True Value	**Predicted Value**	
	Class 1	**Class 0**
Class 1	TP	**FN**
Class 0	**FP**	**TN**

Each column in the matrix represents the prediction class, while each row represents the events in the actual class [15]. From the four sections, an evaluation will be generated in the form of accuracy, F1-score, precision, support and recall. Precision is used to measure the probability of classifier exactness, while recall is used to measure the probability of classifier completeness. Unlike precision and recall, the F1-score tries to compare the balance between precision and recall, while support is used to calculate the amount of data.

3. Results and Discussion

The implementation of this research is the application of the Logistic Regression method for diabetes classification techniques using streamlit as a library in python-based web development [16]. This section will explain the user interface of the application, the application made with the aim of making a diagnostic prediction whether a person has diabetes or not.

3.1. Performance Comparison of Logistic Regression Model

The performance of the logistic regression model in this research measured using four parameters, accuracy, precision, recall, and F1-Score. The higher the value of the four

p-ISSN: 2252-3006
e-ISSN: 2685-2411

evaluation parameters, the better the resulting model will be. With the same treatment as previous studies. Table 3 is a table of classification report results obtained from this research.

Table 3. *Classification Report*

class	Precission	Recall	F1-Score	Support
0	0,78	0,86	0,81	125
1	0,84	0,75	0,79	125
Accuracy		**0,80**		**125**

From the table 3, the precission value in class 0 is 78%, the precission value in class 1 is 84%, the recall value in class 0 is 86%, the f1-score value in class 0 is 81% and the f1-score value in class 1 by 79%. From this research, an accuracy of 80% is obtained, which means it is better than previous research.

Table 4. Comparison of logistic regression performance results

Evaluation	Result [7]	Result
Accuracy	75,97 %	80 %
Precision	76,92 %	84 %
Recall	51,72 %	75 %
F1-score	61,86 %	79 %

Table 4 is the result of a comparison between this research and previous research [7]. Based on the table, it`s known that this research has better classification reports than previous research. There are several process-engineering features carried out in this research, including handling misising values, feature selection, and data oversampling which in fact have an effect on the research results.

3.2. Application view

The main display contains several menus, including the data retrieval menu, the pre-processing data menu, the train test split menu, the Logistic Regression process menu, the Logistic Regression method evaluation menu, and the data prediction feature menu as shown in Figure 2. In Figure 3, the data collection menu displays the initial data used in this research, while the data in this menu is the same data as the previous research, "Pima Indian Diabetes Dataset".

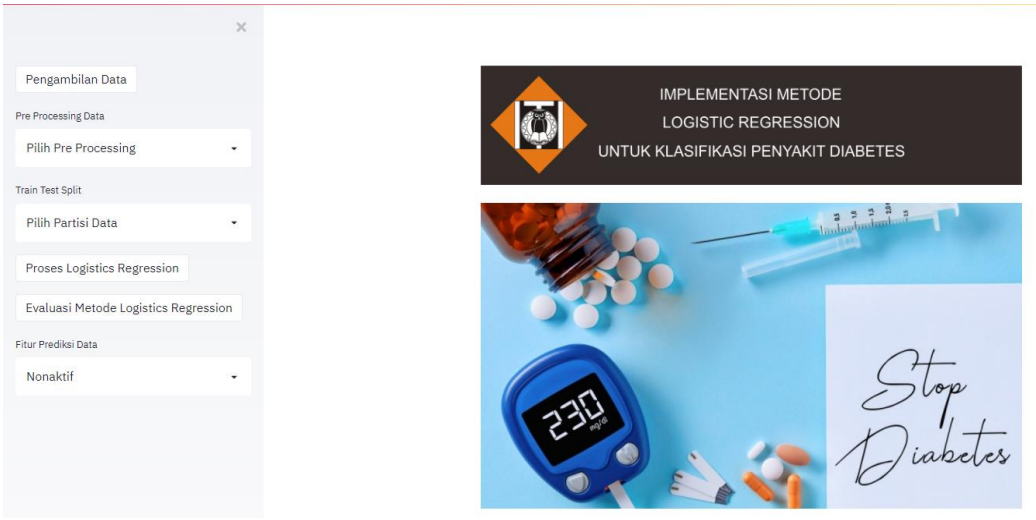

Figure 2. Application View

p-ISSN: 2252-3006
e-ISSN: 2685-2411

Data Pima Indian Diabetes Datasets

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Diab
0	6	148	72	35	0	33.6000	
1	1	85	66	29	0	26.6000	
2	8	183	64	0	0	23.3000	
3	1	89	66	23	94	28.1000	
4	0	137	40	35	168	43.1000	
5	5	116	74	0	0	25.6000	
6	3	78	50	32	88	31	
7	10	115	0	0	0	35.3000	
8	2	197	70	45	543	30.5000	
9	8	125	96	0	0	0	
10	4	110	92	0	0	37.6000	

Informasi Data

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
count	768	768	768	768	768	768
mean	3.8451	120.8945	69.1055	20.5365	79.7995	31.9926
std	3.3696	31.9726	19.3558	15.9522	115.2440	7.8842
min	0	0	0	0	0	0
25%	1	99	62	0	0	27.3000
50%	3	117	72	23	30.5000	32
75%	6	140.2500	80	32	127.2500	36.6000
max	17	199	122	99	846	67.1000

Figure 3. Data Retrieval Menu

3.3. Pre-Processing Menu

The Pre Processing menu contains the steps that are prepared for the Logistic Regression method process, as for the sub menu on the pre-processing menu, namely the mising value menu, the correlation coefficient menu, the data-cleaning menu, and oversampling data menu. The Mising Value menu contains improvements to the independent variable data, which is zero and replaced with the average value of each attribute. The Correlation Coefficient menu contains correlation data for each attribute; in the Correlation Coefficient Menu, you can also see the correlation order between each independent variable and the dependent variable. The Data Cleaning menu contains data through several stages, the data cleaning process with removing attributes no longer needed in this research. Data cleaning is one of most popular step in data mining, such [17] use data cleaning for gathering Tourism object in Bali.The next menu is oversampling, this menu displays additional data with the one point cross over technique, while additional data in this research made with the aim of balancing the number of class 0 and class 1. Figure 4 shows the display of one of the preprocessing menus.

p-ISSN: 2252-3006
e-ISSN: 2685-2411

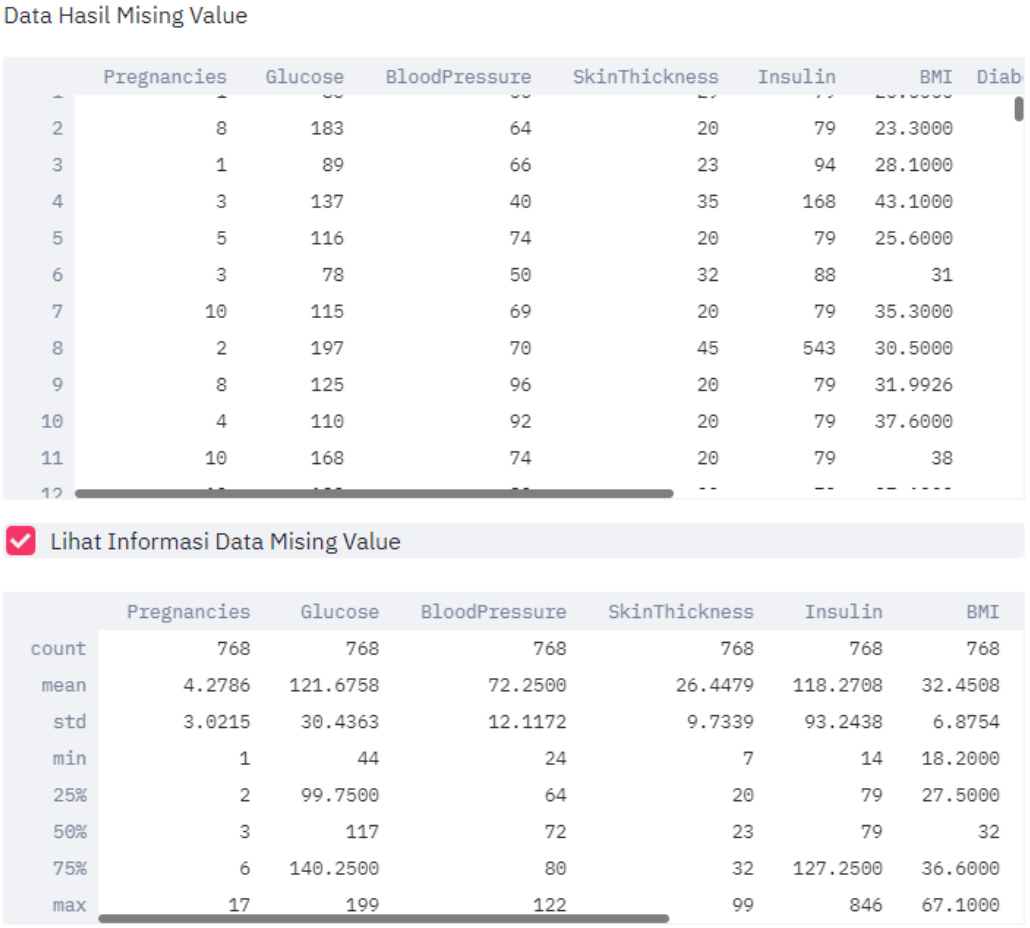

Figure 4. One of the pre-processing menus

3.4. Train Test Split menu

The train test split menu contains the distribution of train and test data, the training data functions to conduct training on machine learning, while the testing data functions to test the Logistic Regression model, in this menu, 75% of training data is used and 25% testing data. The display of the train test spit menu as shown in Figure 5.

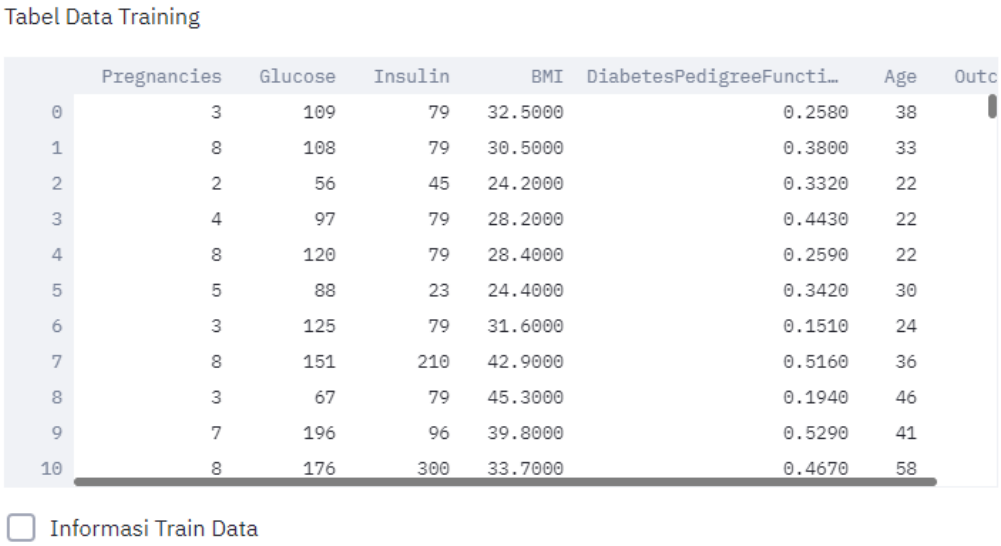

Figure 5. Train Test Split menu

p-ISSN: 2252-3006
e-ISSN: 2685-2411

3.5. Logistic Regression Process Menu

The Logistic Regression process menu contains the Logistic Regression prediction results; the prediction obtained from the sigmoid function in the Logistics Regression process. The Logistic Regression menu display design shown in Figure 6

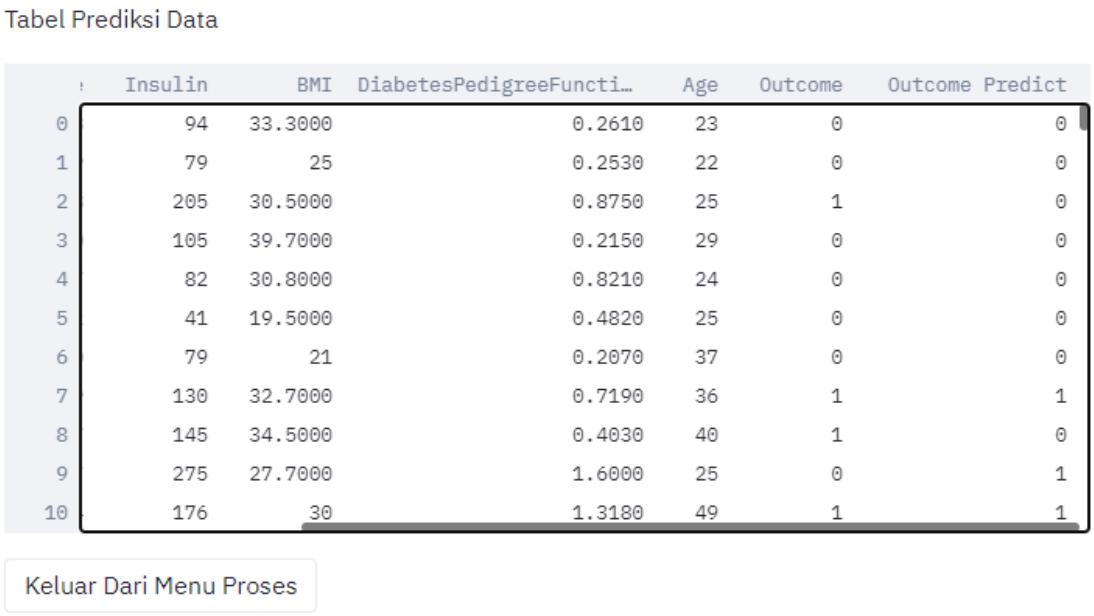

Figure 6. Logistic Regression Process Menu

3.6. Logistic Regression Method Evaluation Menu

The Logistic Regression Evaluation Menu contains information about the evaluation of the Logistic Regression method in the form of Confusion Matrix and Classification Report tables. Figure 7 is a display image of the Logistic Regression Method Evaluation Menu.

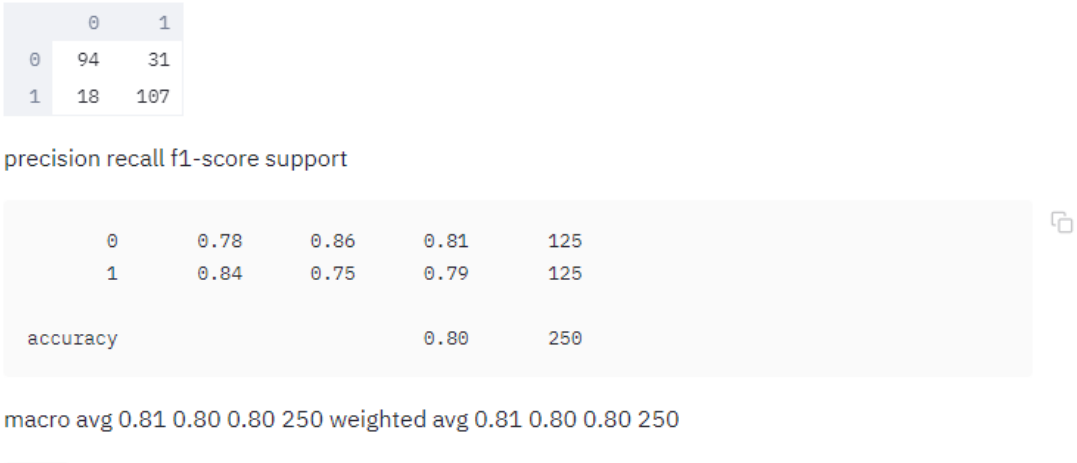

Figure 7. Logistic Regression Method Evaluation Menu

3.7. Data Prediction Menu

The data prediction made with the aim that the user can provide manual input for each variable, and the user can find out the prediction results on the system. The data prediction menu display shown in Figure 8. Data prediction is the main key of machine learning. There are some algorithm can be used in machine learning. One of them is logistic regression. The other one such as support vector regression. Machine learning can solve problem in real life, such as

p-ISSN: 2252-3006
e-ISSN: 2685-2411

forecasting the Number of Traffic Accidents Using the Support Vector Regression Method [18] and many others.

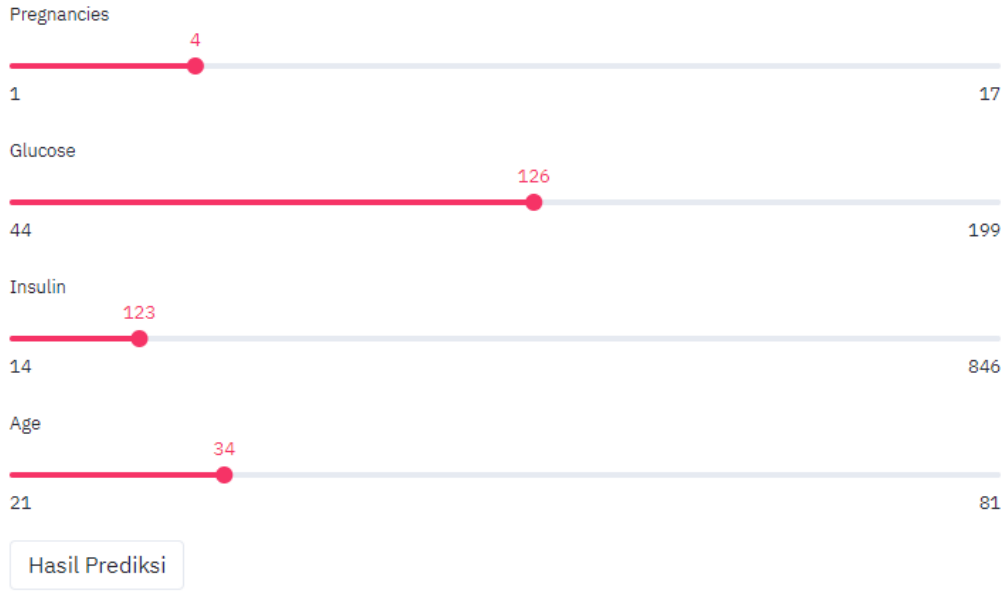

Figure 8. Data Prediction Menu

4. Conclusion

From this whole series of research, several conclusions can, including the following:

1. In this research, the Logistics Regression method used for the classification of diabetes, and software made based on website using streamlit.
2. To make predictions using the Logistic Regression method consists of several stages, the first is the pre-processing of data, the Logistic Regression Process and the evaluation of the Logistic Regression method.
3. There are differences in pre-processing in this research with previous research, in previous reseach including data cleansing and oversampling data. Correlation coefficient process using for data cleaning process. Oversampling data using the One Point Crossover technique.
4. The results of the evaluation of the Logistic Regression method in this study obtained a predictive accuracy value of 80%, which means that it is better than previous studies with a predictive accuracy of 75.97%.

References

[1]. A. Roihan, P. A. Sunarya and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang," *Indonesian Journal on Computer and Information Technology,* vol. Vol. 5 No.1, pp. 75-82, 2020.

[2]. Marios Anthimopoulos, stargios chistodoulidis, Lucas ebner, Adreas christe dan Stavroula mougiakakou, *"Lung Pattern Classification for Interstitial Luang Disiases Using a Deep Convulutional Neural Network"*, IEEE Trans Med Imaging. 2016 May;35(5):1207-1216. doi: 10.1109/TMI.2016.2535865. Epub 2016 Feb 29.

[3] Rifqi Hammad, Julia Kurniasih, Nur Fitrianingsih Hasan, Christin Nandari Dengen, Kusrini. "Prototipe Machine Learning untuk Prognosis Penyakit Demensia" IPTEK-KOM, Vol. 21 No. 1, Juni 2019: 17 - 29

[4] Sana Rebbah, Daniel Delahaye, Stepane Puechmorel, Pierre Marechal, Florenco Nicol. "Classification of Multiple Sclerosis Patients Using a Histogram based KNN Algorithm". OHBM 2019, pertemuan tahunan ke 25 organisasi pemetaan otak manusia, juni 2019, Roma, Italia.

[5]. F. D. Telaumbanua, P. Hulu, T. Z. Nadeak, R. R. Lumbantong and A. Dharma,

"Penggunaan Machine Learning Di Bidang Kesehatan," *Jurnal Teknologi dan Ilmu Komputer Prima,* vol. 3 No. 1, pp. 57-64, 2018.

[6]. D. Margaret F. Schulte, Healthcare Delivery in the U.S.A, New York: CRC Press, 2013.

[7]. F. I. Kurniadi dan P. K. Vinnia , "Perbandingan Regresi Linear dengan Heaviside Activation Function dengan Logistic Regression untuk Klasifikasi Diabetes," *ULTIMATICS,* Vol. 1, No. 1, pp. 7-10, 2018.

[8] M. Alotaibi and M. Albalawi, "A Mobile Gestational Diabetes Management and," *2018 9th IEEE Control and System Graduate Paper Colloquium (ICSGRC 2018), 3 - 4 August 2018, Shah Alam, Malaysia,* pp. 193-196, 2018.

[9]. Y.-h. Wang, Y. Ou, X.-d. Deng, L.-r. Zhao dan C.-y. Zhang, "The Ship Collision Accidents Based on Logistic Regression and Big Data," *The 31th Chinese Control and Decision Conference (2019 CCDC),* pp. 4438-4440, 2019.

[10] M Soleh1, E R Djuwitaningrum1, M Ramli1 and M Indriasari, "Feature engineering strategies based on a One-p oint Crossover for fraud detection on Big Data Analytics" Published under licence by IOP Publishing Ltd, Journal of Physics: Conference Series, Volume 1566, 4th International Conference on Computing and Applied Informatics 2019 (ICCAI 2019) 26-27 November 2019, Medan, Indonesia

[11] https://www.kaggle.com/uciml/pima-indians-diabetes-database

[12]. O. Somantri dan M. Khambali, "Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika," J NTETI, Vol. %1 dari %26, No.3, pp. 301-306, 2017.

[13] J. Suryaputra, C. Lubis and T. Sutrisno, "PEMILIHAN CROSSOVER PADA ALGORITMA GENETIKA UNTUK PROGRAM APLIKASI PENGENALAN KARAKTER TULISAN TANGAN," Jurnal Ilmu Komputer dan Sistem Informasi, pp. 69-72.

[14] A. Chowdhury, . S. Tejas and T. K, "Predicting whether songs will be hit using Logistic Regression," International Journal Of Engineering And Computer Science, vol. 6, p. 22434, 2017.

[15] M. Nawawi dan R. Marliansyah, "Klasifikasi Tingkat Popularitas Siswa Berdasarkan Aktifitas Komunikasi Siswa Menggunakan Smartphone dengan Teknik Logistic Regression," Prosiding Annual Paper Seminar 2018 Computer Science and ICT, Vol. %1 dari %24, No.1, pp. 251-254, 2018.

[16] www.streamlit.io

[17] WIDIARI, Ni Putu Ayu; SUARJAYA, I Made Agus Dwi; GITHA, Dwi Putra. Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali. Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi), [S.l.], p. 137-145, july 2020. ISSN 2685-2411

[18] APRIYANTI, Ni Putu Ratindia; PUTRA, I Ketut Gede Darma; PUTRA, I Made Suwija. Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression. Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi), [S.l.], p. 72-80, june 2020. ISSN 2685-2411