EMITTER International Journal of Engineering Technology
Vol. 6, No. 1, June 2018
ISSN: 2443-1168

Influence of Logistic Regression Models For Prediction and Analysis of Diabetes Risk Factors

Yufri Isnaini Rochmat Maulana, Tessy Badriyah, Iwan Syarif

Electronic Engineering Polytechnic Institute of Surabaya
Jl. Raya ITS Politeknik Elektronika, Kampus ITS Sukolilo, Surabaya
(031) 5947280
E-mail: yufriisnani.rm@gmail.com, tessy@pens.ac.id, iwanarif@pens.ac.id

Abstract

Diabetes is a very serious chronic. Diabetes can occurs when the pancreas doesn't produce enough insulin (a hormone used to regulate blood sugar), cause glucose in the blood to be high. The purpose of this research is to provide a different approach in handling with cases of diabetes, that's with data mining techniques using logistic regression algorithm to predict and analyze the risk of diabetes that is implemented in the mobile framework. The dataset used for data modeling using logistic regression algorithm was taken from Soewandhie Hospital on August 1 until September 30, 2017. Attributes obtained from the Hospital Laboratory have 11 attribute, with remove 1 attribute that is the medical record number so it becomes 10 attributes. In the data preparation dataset done preprocessing process using replace missing value, normalization, and feature extraction to produce a good accuracy. The result of this research is performance measure with ROC Curve, and also the attribute analysis that influence to diabetes using p-value. From these results it is known that by using modeling logistic regression algorithm and validation test using leave one out obtained accuracy of 94.77%. And for attributes that affect diabetes is 9 attributes, age, hemoglobin, sex, blood sugar pressure, creatin serum, white cell count, urea, total cholesterol, and bmi. And for attributes triglycerides have no effect on diabetes.

Keywords: diabetes, regression, recommendation, mobile, framework.

1. INTRODUCTION

Diabetes is one of the most serious chronic diseases. Diabetes canoccur when the pancreas doesn't produce enough insulin (a hormone used to regulate blood sugar), or when the body can not effectively use the insulin produced by the body itself [1, 2]. High blood glucose causes diabetes is not easily controlled so it will cause some of the consequences that can occur that include causing serious damage to the heart, blood vessels, eyes, kidneys, and

neurological disorders. Diabetes is a major cause of some diseases that attack the body, and can cause death [3, 4, 5].

In 2013, The World Health Organization (WHO) noted that diabetes accounts for 1.3 million deaths (2.4% of all deaths). And the latest report from WHO in 2016 there are more than 400 million people living with diabetes. The World Health Organization (WHO) estimates, more than 387 million people worldwide suffer from diabetes which is likely to double by 2030 [6, 7, 8, 9].

Diabetes is even referred to as a global epidemic that strikes low and middle income countries. A global comparison conducted by WHO found that the rate of increase of diabetics observed in the Southeast Asian Region and the Eastern Mediterranean Region was the highest [10, 11]. The load of diabetes does not only occur in the health sector, but also in social and economic sectors. In Indonesia Country, the percentage of adults with diabetes reached 8.5 percent or 1 in 11 adults who suffered from this disease. But the facts found in the field, 1 among 2 people with diabetes is still undiagnosed and not yet realize that he has diabetes.

Most cases of diabetes are type 2 diabetics whose 90% cause is due to lifestyle patterns that tend to be less physical activity, unhealthy and unbalanced diet and tobacco consumption (smoking) [12]. Therefore, the control of diabetes risk with preventive and promotive aspects in an integrated and comprehensive manner.

2. RELATED WORKS

In the preparation of this research, researchers are somewhat inspired and have been referenced by some of the previous studies relating to the background of the problem in this research.

Some pre-built frameworks are able to deal with several diseases such as control the health of children with problems limited ability, especially with special conditions, such as infants, sick children, and children with limited ability to move and thinking [13], then a framework for managing daily diet recommendations and exercise recommendations that can be done to keep of the health of the sufferer [9, 14].

There is also a semantic-based web-based framework to provide information needed by patients, such as finding information about food and exercise what is good for him [9, 15, 16]. By utilizing iOT technology some researchers also implement a framework that can control the diet and exercise with a device used by the user [17, 18, 19, 20].

3. ORIGINALITY

Diabetes is one of the most serious chronic diseases. Diabetes can occur either when the pancreas does not produce enough insulin or when the body can not effectively use the insulin produced by the body itself [1, 2]. From this background this research is expected to provide a new approach in diabetes problems. The approach is to create a system by taking the data mining modeling results using Logistic Regression algorithm. If previous research many use classification algorithm such as SVM until Decission Tree in solving

problems related to diabetes, then in this research using Logistic Regression algorithm with consideration to know the chances of someone having diabetes and attribute factors that influence. And in previous studies related to diabetes, most of the sampling data used is not real data taken directly from the Hospital, or Clinic. So in this research data used for data modeling is real data taken from Soewandhie Hospital Surabaya City, with attributes obtained from the hospital such as, medical record number, age, gender (L / P), body mass index (BMI), hemoglobin (gr / dl), white cell count (10 ^ 3 / ul), sugar pressure blood (mg / dl), serum creatin (mg / dl), urea (mg / dl), total cholesterol (gr / dl), triglycerides (gr / dl).

Stages of data mining techniques used in this research are Data Collection, Data Preprocessing which includes Replace Missing Value, Normalization, and Feature Extraction, Algorithms used are Logistic Regression, and Data Validation using Leave One Out Validation. The result of the data mining is prediction and analysis of diabetes disease in the form of performance measure covering accuracy, f - measure, precision, recall, accuracy error (RMSE), and any attribute analysis that is very influential to diabetes.

From result of modeling, will be able to provide preventive and self-control against diabetes. For provide prevention, in this research with give recommendation. There is two recommendation, that is food recommendation and exercise recommendation. While a person can control health life and control blood glucose pressure for diabetics.

4. SYSTEM DESIGN

This research uses system design concept from data mining, where the process stages are Data Collection, Data Preprocessing, Data Mining, Validation Testing, and Classification Result as shown in Figure 1.

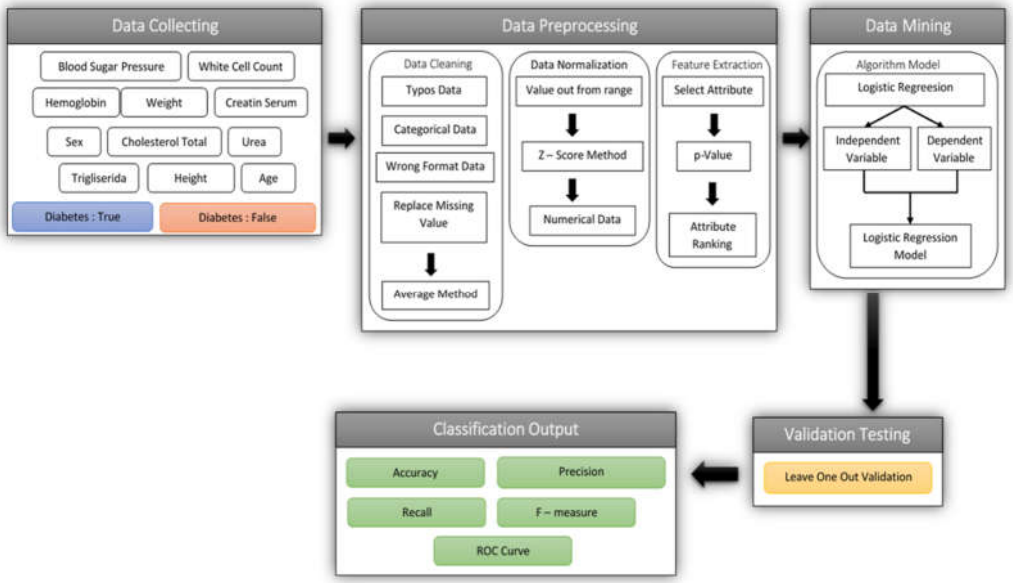

Figure 1. Design System

4.1. Data Description

Data retrieval in this research was conducted at Soewandhie Hospital in Surabaya City. The data taken are in the form of patient administrative data and laboratory data of patient in period of August 1 - September 30, 2017. Specifically, the data needed in the research are diabetic data and non diabetics data. Descriptions of both data are as follows :

- Amount of data there is 1550 data
- Amount of class there is 2 class :
  1. True (Diabetes) = 596 data
  2. False (Non-Diabetes) = 954 data
- Amount of attribute of data there is 11 :
  1. Medical Record Number
  2. Age
  3. Sex (M / F)
  4. BMI (Body Mass Index)
  5. Hemoglobin (gr/dl)
  6. White Cell Count (10^3/ul)
  7. Blood Glucose Pressure (mg/dl)
  8. Creatin Serum (mg/dl)
  9. Urea (mg/dl)
  10. Cholesterol Total (gr/dl)
  11. Trigliserida (gr/dl)

Due to attribute of Medical Record Number is not required, it can be directly removed and become 10 attributes.

- Amount of missing value in each attribuet :
  1. Age = 0,12 %
  2. Sex (M / F) = lengkap
  3. BMI (Body Mass index) = 50,45%
  4. Hemoglobin = 6,45%
  5. White Cell Count = 6,45%
  6. Blood Glucose Pressure = 6,45%
  7. Creatin Serum = 6,45%
  8. Urea = 6,45%
  9. Cholesterol Total = 6,45%
  10. Trigliserida = 6,45%

From the data obtained and already described as above, then the percentage of people who suffered from diabetes is 37.59% and people who do not have diabetes is the rest of the percentage. That is equal to 62,41%. So it can be assumed because the percentage of people suffering from diabetes is above 10%, then the data used is balance (balance data).

4.2. Data Preprocessing

At this preprosessing stage it is aimed at making raw data that already obtained into qualified data. Because this process will be can increase the accuracy and reduce the error rate ratio in the process of data mining.

Methods performed at the preprocessing stage is Replace Missing Value, Normalization, and Feature Extraction. Preprocessing technique is done to handle some cases in raw, including :

- Incomplete, that is data that lacks attribute values or contains only aggregate data (example : BMI = “”)
- Noisy, that is data have error and outliers (example : Age = -20)
- Inconsistent, that is data containing discrepancies in letters and numbers (example : Sex there’s M / F, changed to 0 / 1)

4.2.1. Replace Missing Value

Replace missing value is a process performed if there’s a data value that is empty or contains only the aggregate that has been described above. In this research the technique used to replace missing value is Average Method. Where is looking for the average value of the whole data then the average value will be set on the empty value attribute.

The mean or average is a measure of statistically centralized tendencies as well as the median and mode. The average calculation is done by sum of all the data values of a sample group, then divided by the number of samples. So if a group of random samples with the number of samples n, then it can be calculated average of the sample with the following formula :

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + xn) \quad (1)$$

If it is denoted by sigma notation, then the above formula will be as follows :

$$\bar{x} = \frac{1}{n}\sum_{n=1}^{n} xi \quad (2)$$

Description :

$\bar{x}$ = Average Count
xi = Sample Value
n = Number of Samples

4.2.2. Normalization

Normalization is the process for scaling the attribute values of the data so that it can fall in a certain range. In this research the technique used for normalization is Z - Score. Please note for normalization process in this research is not used in all attributes,

because basically this normalization process is done to normalize the scale of the value of attributes that have a range between values far apart. The attributes of normalization techniques are triglycerides and the amount of white blood sugar.

Z - Score is a normalization method that is found based on the mean (mean value) and standard deviation of the data. The Z - Score method is very useful if it does not know the actual and minimum value of the data. To calculate Z - Score is by using the following formula :

$$\text{Z} - \text{Score} = \frac{\text{X} - \bar{x}}{\text{S}} \quad (3)$$

Description :

X = Value of Subject
$\bar{x}$ = Average Value
s = Standard Deviation

In the normalization stage it is also necessary to calculate the standard deviation of the formula to determine the following variance.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_1\right)^2}{n\,(n-1)} \quad (4)$$

The standard deviation formula (standard deviation) of the above variance formula is as follows :

$$s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_1\right)^2}{n\,(n-1)}} \quad (5)$$

Description :

s^2 = Varians
s = Standard Deviation
x_i = Value of x
n = Sample Size

4.2.3. Feature Extraction

This stage is part of the preprocessing stage, in this stage the process of selecting features to determine the effect of features on the class. The purpose of this feature selection process is to improve the classification performance of diabetes by examining the effect of each attribute on the class and also to get the rank of attribute to the classification performed. The method used to perform feature selection is to use p-Value.

P - Value is the smallest probability value of a hypothesis testing so that the observed static value of the test is still meaningful. To determine the selection of features that have an important effect on the class, hypothesis is done by using hypothesis test with α (α = 0,05), where:

- If p-Value < α, then Ho accepted (it mean attribute significant effect on class)
- If p-Value > α, then Ho rejected (it mean attribute not significant effect on class)

Description :

α : The relationship constant between two variables
Ho : Hypothesis Zero

Reffer from hypothesis of determining attributes that directly affect risk of diabetes with α (α = 0,05), then obtained p – value is sequence as in table 1 as below.

Table 1. P – Value of Each Attribute

Attribute	P - Value
Sex	0,033
Age	0,027
BMI	0,016
Hemoglobin	0,001
White Cell Count	0
Blood Presure	0
Creatin Serum	0
Cholesterol Total	0
Trigliserida	0,356

So by looking at the table then the attribute whose p – value < 0,05 in sequence is sex, age, bmi, hemoglobin, white cell count, blood presure, creatin serum, and cholesterol total. While for attibute which value of p – value > 0,05 is triglyceride. So in the process of classification of the class need to use all attributes except triglyceride attribute because the influence is not significant to the class.

4.3. Data Modelling

Variable / attribute that used for modelling data in this research using data from Dr. Soewandhie Hospital and then in this research using logistic regression algorithm with description as follows :

- Response variable (Y) that's percentage of people at risk of diabetes
- Independent variable (X), that's percentage of sex (X_1), percentage of age (X_2), percentage of BMI (X_3), percentage of hemoglobin (X_4), percentage of white cell count (X_5), percentage blood glucose (X_6), percentage of creatin serum (X_7), percentage of urea (X_8), percentage of cholesterol total (X_9), and percentage of trigliserida (X_{10})

Logistic regression is one model for predicting the relationship between the category response variable with one or more continuous predictor variables or categories [21, 22, 23, 24]. The response variable consists of two categories: y = 1 "success" and y = 0 "fail" [21, 22]. In such circumstances, the variable y follows the Bernoulli distribution for each single observation. The probability function for each observation is given as follows [21, 25].

$$f(y_i, \pi_i) = \pi_i^y (1 - \pi)^{1-y_i} ; y = 0,1 \quad (6)$$

Description :

- Where if, y = 0, then f (y) = 1 - π
- And If y = 1, then f (y) = π

And equation for logistic regression function can be written as follows.

$$\pi(x) = \frac{e^{g(x)}}{1+ e^{g(x)}} \quad (7)$$

Description :

$\pi(x)$: chance of success with probability value $0 \leq \pi(x) \leq 1$
$g(x)$: logit equation of logistic regression

From equation as above, so equation of Logistic Regression model as follows.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)}{1+\exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)} \quad (8)$$

Description :
$\pi(x)$: chance of success with probability value $0 \leq \pi(x) \leq 1$
β_j : parameter value with j = 1, 2, 3, ..., p

exp : exponen function (exponen is the opposite of natural logarithm. While natural logarithm is a logarithmic form but with constant value 2.71828182845904 or commonly rounded to 2.72)

4.4. Performance Measure

Validity test from performance of the algorithm to produce accuracy that is in the form of confusion matrix, precision, recall, and F-measure, where the method formula - the method is as follows :

- Confusion Matrix

$$\text{True Positive Rate} = \frac{TP}{(TP+FN)} \quad (9)$$

$$\text{False Positive Rate} = \frac{FP}{(FP+TN)} \quad (10)$$

$$\text{Succes Rate} = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (11)$$

$$\text{Error Rate} = 1 - \text{Success Rate} \quad (12)$$

- $\text{Precision} = \frac{correct}{correct+falsePositive}$ (13)
- $\text{Recall} = \frac{correct}{correct+falseNegative}$ (14)
- $\text{F – measure} = \frac{2 \text{ x precision x recall}}{precision+recall}$ (15)

Description :

correct = The number of slots filled correctly
falsePositive = Number of filled but incorrect slots
falseNegative = Number of unallocated slots

And the error ratio of the predicted result pattern is an error that occurs between the predicted data and the actual data. The error is represented using the mean squared error (MSE), that is the difference of the squares between the predicted value data and the data being observed, while root mean square error (RMSE) is the root of MSE.

5. EXPERIMENT AND ANALYSIS

Experiment in this research will do with two scenario, that's the first for experiment with logistic regression algorithm and the second experiment with compare logistic regression algorithm and support vector machine algorithm.

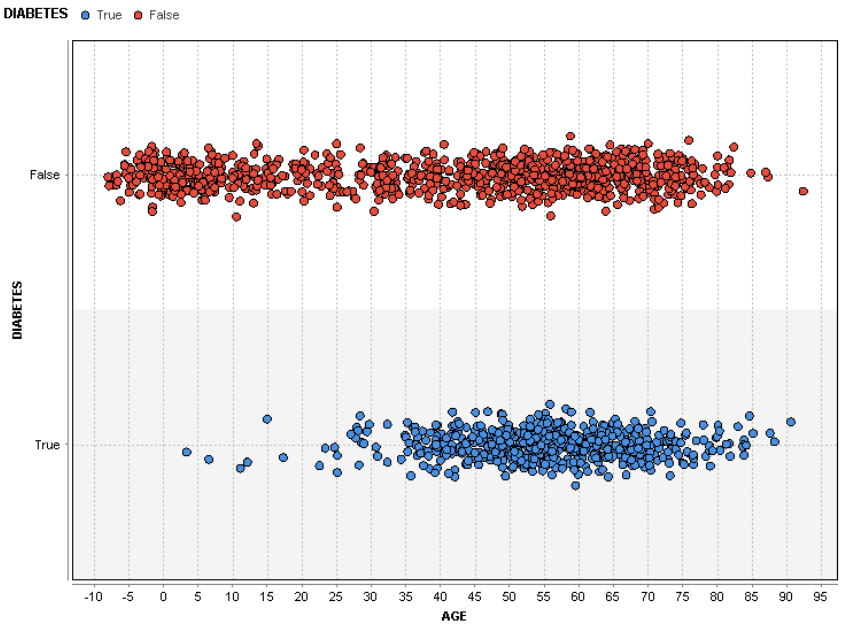

Figure 3. Plot Diabetes Patients

Figure 3 shows the plot of diabetics from the dataset already taken. For x - axis selected age attribute, and y - axis selected attributes of blood sugar. Seen that in the age range 35 years and over is age susceptible to diabetes.

5.1. Logistic Regreesion Algorithm

Table 1 shows the overall estimation of the attributes used in this research that can significantly influence the data modeling of diabetes classification.

Table 1. Estimation of Attribute Effect with Logistic Regression Modeling

Attribute	Coefficient	Std. Error	z-Value	p-Value
Sex (X1)	-0,023	0,279	-0,084	0,033
Age (X2)	0,011	0,009	1,263	0,027
BMI (X3)	-0,083	0,034	-2,415	0,016
Hemoglobin (X4)	0,248	0,077	3,223	0,001
White Cell Count (X5)	-0,161	0,043	-3,709	0
Blood Pressure (X6)	-0,057	0,006	-10,015	0
Creatin Serum (X7)	-2,309	0,614	-3,76	0
Urea (X8)	-0,096	0,022	-4,338	0
Cholesterol Total (X9)	-0,019	0,004	-4,555	0
Trigliserida (X10)	**0,004**	**0,004**	**0,924**	**0,356**
Intercept	20,211	8,58	2,355	0,018

By testing alpha = 0.05 as it has been done in the feature selection process it is obtained almost all attributes have a significant effect on the class. The attributes are X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, and X_9. Then for logit equation is :

$$g(x) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$

$$g(x) = 20{,}211 - 0{,}023X_1 + 0{,}011X_2 - 0{,}083X_3 + 0{,}248X_4 - 0{,}161X_5 - 0{,}057X_6 - 2{,}309X_7 - 0{,}096X_8 - 0{,}019X_9$$

And referring to the formula number 3, then the logistic regression equation model is as follows :

$$\pi(X_i) = \frac{\exp(20{,}11 - 0{,}023X_1 + 0{,}011X_2 - 0{,}083X_3 + \ldots + 0{,}019X_9)}{1 + \exp(20{,}11 - 0{,}023X_1 + 0{,}011X_2 - 0{,}083X_3 + \ldots + 0{,}019X_9)}$$

After get the model from logistic regression, it will obtained chance model for determine risks of diabetes based on 9 variable inside of logistic regression model.

$$\text{Log}(p/1\text{-}p) = 20{,}211 - 0{,}023X_1 + 0{,}011X_2 - 0{,}083X_3 + 0{,}248X_4 - 0{,}161X_5 - 0{,}057X_6 - 2{,}309X_7 - 0{,}096X_8 - 0{,}019X_9$$

From the chance model as above, it will obtained 4 categories level risks of diabetes, namely:

1. >= 0,0 healthy category,
2. >= 0,5 low category,
3. >= 0,7 average category.
4. And = 1 high category

To illustrate how ROC calculations use logistic regression methods, this experiment has 1550 data, where scores on class attributes are predicted attributes and the results are attributes that are monitored. On a discredited score with predictive numbers from 0.0 to 1.0, binary results with 0 and 1 as scores.

From the model that has been established using logistic regression, accuracy testing for Logistic Regression model using Leave One Out Validation. With data proportion between data training 70% and data testing 30%. And here are the results of performance measure shown in table 2. For ROC Curve logistic regression model shown in figure 4.

Table 2. Accuracy Use Leave One Out Validation

Leave One Out Validation					
RMSE	Precision	Recall	F - Measure	ROC Area	Accuracy
0,1837	0,949	0,948	0,947	0,991	94,77%

Result of the validation test scenario using leave one out validation in this research obtained value of RMSE, precision, recall, f-measure, roc area, and accuracy. From table 3, value of RMSE is 0,1837 that show if value of RMSE

is closer to 0, so variation of value generated by the model in this research is closer to the variation of observation value. Value of precision that obtained is 0,949 which show accuracy of value generated in the modelling equals the observation value, while value of recall that obtained is 0,948 which show success of model in this research resulting value is the same as the observation value. And value of F – Measure in this research is 0,947 which the value is combination from value of precision and value of recall. For accuracy that obtained is 94,77% which is high accuracy value. In this reseach can resulting good model to used solve the problem, so with the accuracy of this research can already be used to determine the risk level of a person suffering from diabetes.

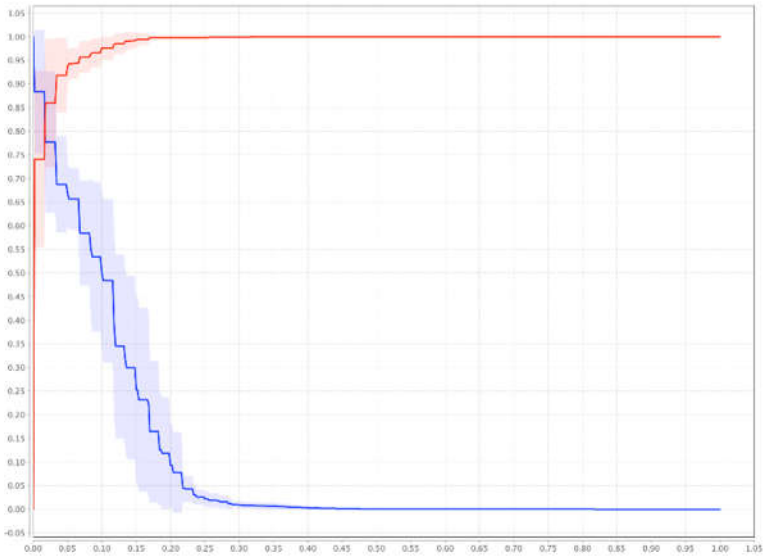

Figure 4. ROC Curve

With value of ROC Area is 0,991 which show in table 3 and show ROC Curve in figure 4. Red line show as ROC and blue line is ROC thresold, horizontal line take from false positive data, while vertical line take from true positive data. With value of ROC Area is 0,991 which show in table 3 and ROC Curve in figure 4. Red line show as ROC and blue line is ROC thresold, horizontal line taken from false positive data, while vertical line taken from true positive data.

Table 3. Example data test of Risk Level of Diabetes with Logistic Regression Algorithm

Hemoglobin	Blood Glucose Pressure	Creatin Serum	Urea	Cholesterol Total	Trygliceride	BMI	Predicted	Observed	Risk
12,5	226	1,2	48	148	82	33,6	0,9	1	moderate
13,4	213	0,2	42	275	221	23,3	0,8	1	moderate
14,6	315	1,2	42	215	122	39,8	1	0	high
14,3	303	1,2	29	208	180	29	1	0	high
15,4	300	1,2	32	285	197	37,7	1	1	high

In table 3 shows example data test level risks of diabetes that are obtained from the calculation of logistic regression method and chance model

method. From example data test as above can predicted for risks of a person suffering from diabetes is >= 0,7, so can categorized average or moderate. Test in this research is data number 1 and number 2 If a person have chance >= 0,5, so so can categorized low. Test in this research is can't find the data. If a person have chance >= 0,0, so so can categorized healthy. Test in this research is can't find the data also. While if a person have chance = 1, so can categorized high. Test in this research is data number 3, data number 4, data number 5, and data number 6.

5.2. Improve System Framework

If in previous research or system, researcher only work for modelling, analyze, and test data, so in this research try to make system can integrated between artificial intelegence that's data mining and recommendation system. So this research can get result from testing data that's risks of diabetes, and then will given recommendation about preventive action towards diabetics. In this research there's 2 recommendation, that's food recommendation and exercise recommendation.

5.2.1. Food Recommendation

At the consensus of PERKENI (Perkumpulan Endikronologi Indonesia) it has been established that the recommended standard is a meal with a balanced composition of 60-70% carbohydrates, 10-15% protein, and 20-25% fat. Daily calorie needs, using Harris-Benedict Equation [26] :

1. Ideal Weight
   Ideal Weight = [(Height in cm – 100) x 1 kg] x 90% (16)
2. Basal Calorie Needs
   If women, Basal Calorie Needs = Ideal Weight x 25 kkal (17)
   If men, Basal Calorie Needs = Ideal Weight x 30 kkal
3. Correction (Age Factor, Activity Factor, etc)
   Correction = Calorie Basal Needs x Activity (Light, Medium, Heavy) (18)

5.2.2. Exercise Recommendation

Men every day need approximately 2500 kcal and woman every day need approximately 2000 kcal. For can get ideal weight :

- Men can reduce calories by 500 kcal, making it to 2000 kcal
- Women can reduce calories by 500 kcal, making it to 1500 kcal

5.2.3. Integrated Data Mining and Recommendation

From the results of experimental data that have been done using function logistic regression, which is divided into several categories. So the recommendations will be given will refer to the test results data.

Table 4. Example Result data testing using logistic regression method

Gender	Age	Hemoglobin	Blood Glucose Pressure	Creatin Serum	Urea	Cholesterol Total	Tryglicer ide	BMI	Predicted	Observed	Risk
M	49	12,5	226	1,2	48	148	82	33,6	0,9	1	moderate
F	51	15,4	300	1,2	32	285	197	37,7	1	1	high

Table 4 show result of prediction risk a person suffer from diabetes, if a someone is on category level average or moderate so food and exercise recommendation will be adjusted, also about category level is high. But if a person on category level healthy, so still given recommendation with purpose for healthy life and can prevention from diabetes.

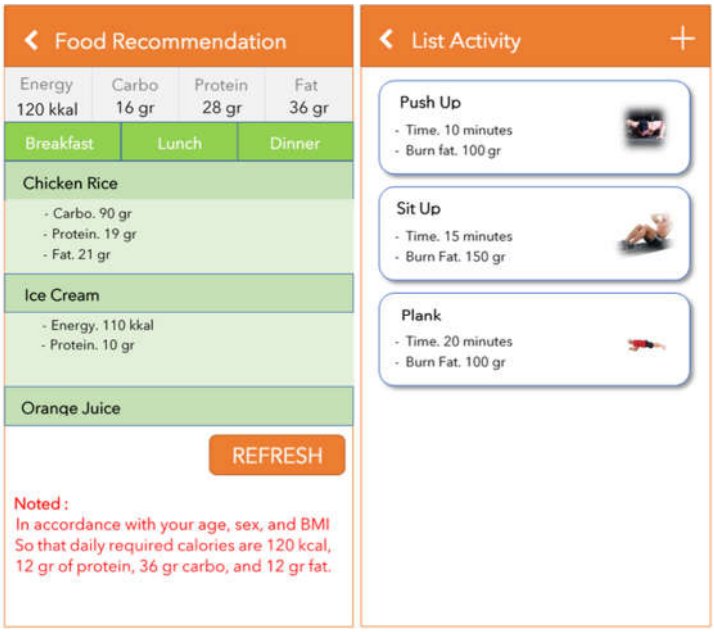

Figure 5. Food and Exercise Recommendation

Figure 5 show result of recommendation that given after a person prediction done, seems like in food recommendation system that given is food types and content inside food which must be fulfilled by a person in order to remain able to control his health, because it will affect a person with diabetics. And for exercise recommendation that given is adjusted to activity of a person with diabetics (low activity, medium activity, and high activity).

6. CONCLUSION

Based on the analysis that has been done, obtained the conclusion. Of the attributes used in this research using Logistic Regression algorithm, almost all attributes affect the prediction accuracy results for diabetes, there is only one attribute that not effect in diabetes disease is trigliserida attribute because it has a value of more p - value of threshold.

Logistic Regression Algorithm can be used as a method to predict and analyze diabetes with dataset taken from Soewandhie Hospital. The accuracy level is 94,77% by using validation Method Leave One Out Validation. Therefore, it can be concluded that logistic regression methods have good discrimination score to predict and analyse diabetes. The prominent advantage of the logistic regression methods is the near perfect 100% ROC score (the Logistic Regression method).

From result and analyze using logistic regression that already done, so system can given food and exercise recommendation with purpose a person can control healthy life. Because diabetics is affected by diet and activity.

REFERENCES

[1] Prayitno, Agus, Wibawa, Andi Dharma, Purnomo, Mauridhi, Hery. **"Early Detection Study of Kidney Organ Complication Caused By Diabetes Mellitus using Iris Image Color Constancy"**. International Conference of Information, Communication Technology and System (ICTS), 146 - 149, 2016

[2] Basar, Md Abul, Alvi, Hassan Nomani, Bokul, Gazi Nowrin, Khan M, Shahriar, Anowar, Farzana, Huda, Mohammad Nurul, Al Mamun, Khondaker Abdullah. **"A Review on Diabetes Patient Lifestyle Management Using Mobile Application"**. 18th International Conference on Computer and Information Technology, 379 - 385, 2015

[3] S. Goyal and J. a. Cafazzo. **"Mobile phone health apps for diabetes management: Current evidence and future developments"**. Qjm vol. 106, no. 12, pp. 1067–1069, 2013

[4] George, Eleni I, Protopappas, Vasilios C, Mougiakakou, Stavroula G. **"Short-term vs. Long-term Analysis of Diabetes Data: Application of Machine Learning and Data Mining Techniques"**. 2013

[5] Ojugo, A A, Eboka, A O, Yoro, R E, Yerokun, M O, Efozia, F N. **"Hybrid Model for Early Diabetes Diagnosis"**. Second International Conference on Mathematics and Computers in Science and in Industry, 55 - 64, 2015

[6] Srikanth, Panigrahi, Deverapalli, Dharmaiah. **"A Critical Study of Classification Algorithms Using Diabetes Diagnostics"**. IEEE 6th International Conference on Advanced Computing, 245 - 249, 2016

[7] Tulu, B et al. **"Design Implications of User Experience Studies The Case of a Diabetes 9Wellness App"**. 49th Hawaii Internasional Conference on System Sciences, p. 3437 – 3482, 2016

[8] Pagoto, S, Schneider K, Jojic M, DeBiasse M, and Mann D. **"Evidence-based strategies in weight-loss mobile apps"**. American journal of preventive medicine, 45, (5), p. 576-582 , 2013

[9] Siahaan, Elisa Julie Irianti, Cholissodin, Imam, Fauzi, M. Ali. **"Sistem Rekomendasi Bahan Makanan Bagi Penderita Penyakit Jantung Menggunakan Algoritma Genetika"**. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 1 No. 11, 1406 - 1415, 2017 ("Food Recommendation System For Heart Disease Patients Using Genetic Algorithm")

[10] Wicaksono, Andri Permana, Badriyah, Tessy, Basuki, Ahmad**. "Comparison of The Data-Mining Methods in Predicting The Risk Level of Diabetes"**. EMITTER International Journal of Engineering Technology, Vol. 4 No. 1, 2016

[11] M. Cable. **"Mobile Diabetes Management Tools"**. no. November, pp. 24–26, 2011.

[12] Douali, Nassim, Dollon, Julien, Jaulent, Marie-Christine. **"Personalised Prediction of Gestational Diabetes Using a Clinical Decision Support System"**. IEEE, 2015

[13] Sevani, Nina. **"Personal Health Care Framework for Children"**. International Conference on Data and Software Engineering, 166 - 170, 2015

[14] Salman, Galih Afan, Prasetio, Yen Lina, Kanigoro, Bayu, Anggi. **"Aplikasi Rekomendasi Pola Makan Berbasis iOS"**. ComTech Vol. 3 No. 2, 796 – 807, 2012 ("Application of iOS-Based food Recommendations")

[15] Al-Nazer, Ahmed, Helmy, Tarek, and Al-Mulhem, Mohammed. **"User's Profile Ontology-Based Semantic Framework for Personalized Food and Nutrition Recommendation"**. Procedia Computer Science 32, 101 – 108, 2014

[16] Tang, Y Y, Zhang, Bob, Shu, Ting. **"Using k-NN With Weights To Detect Diabetes Mellitus Based On Genetic Algorithm Feature Selection"**. Proceding of the 2016 Internasional Conference on Wavelet Analysis and Pattern Recognition, 2016

[17] Al-Taee, Majid A, Al-Nuaimy, Waleed., Al-Ataby, Ali, Muhsin, and Zahra J. **"Mobile Health Platform for Diabetes Management Based on the Internet-of-Things"**. IEEE Jordan Conference on Applied Electrical Engineesing and Computing Technologies (AEECT), 2015

[18] Perez, Javier Andreu., Leff, Daniel R, Ip H M D, and Yang, Guang-Zhong. **"From Wearable Sensors to Smart Implants – Towards Pervasive and Personalised Healthcare"**. IEEE, 2015

[19] Blount, M. et al. **"Remote Healthcare Monitoring Using Personal Care Connect"**. IBM Systems Journal Vol. 46, No. 1, 2017

[20] Rahman, Ruhani Ab, Aziz, Nur Shima Abdul, Yusof, Mat Ikan et al. **"IoT-based Personal Health Care Monitoring Device for Diabetics Patiens"**. IEEE, 2017

[21] Kotimah, Muinah Kusnul, dan Wulandari, Sri Pingit. **"Model Regresi Logistik Biner Stratifikasi Pada Partisipasi Ekonomi Perempuan Di Provinsi Jawa Timur"**. Jurnal Sains dan Seni Pomits, Vol. 3, No. 1, 2014 ("Binary Stratification Logistic Regression Model on Women's Economic Participation in East Java Province")

[22] Al-Nazer, Ahmed, Helmy, Tarek, and Al-Mulhem, Mohammed. **"Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner Dan Radial Basis Function Network di Bank "X" Cabang Kediri"**. Jurnal Sains dan Seni Pomits, Vol. 3, No. 2, 2014 ("Classification analysis of Credits using Binary Logistic Reggresion and Radial Basis Function Network at "X" Bank in Kediri branch office")

[23] Aditya, Ahmad Reza, Suparti, Sudarno**. "Ketepatan Klasifikasi Pemilihan Metode Kontrasepsi Di Kota Semarang Menggunakan Bootstrap Aggregating Regresi Logistik Multinomial"**. Jurnal Gaussian, Volume 3, Nomor 1, 2015 ("Accuracy Classification Selection of Contraception Method In Semarang City Using Bootstrap Aggregating Multinomial Logistic Regression")

[24] Maharani, Indah Irma, Hardinsyah, dan Sumantri, Bambang. **"Aplikasi Regresi Logistik Dalam Analisis Faktor Risiko Anemia Gizi Pada Mahasiswa Baru IPB"**. Jurnal Gizi dan Pangan, 36 - 43, 2007 ("Application of Logistic Regression in Nutrition Anemia Risk Factor Analysis in New Student IPB")

[25] D, Hosmer, & S, Lemeshow**. "Aplied Logistic Regression"**. USA: John Wliey & Sons, 2000

[26] Picolo, Michele Ferreira, et. al. **"Harris-Benedict Equation and Resting Energy Expenditure Estimates in Critically Ill Ventilator Patients"**. American Journal Of Critical Care, Volume 25, 2016