iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** V **Month of publication:** May 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Medalert-A Diabetic Predictor Using Logistic Regression Statistical Model

Shwetank Shukla[1], Raunak Jain[2], Shikhar Dubey[3]
Final Year Students, IMS Engineering College

Abstract: Diabetes is a type of disease in which your blood glucose, or blood sugar, levels are too high. In this research we are focusing on Diabetes patients who have been in a hospital for treatment and we want to classify the ones who are at high risk that they're going to be readmitted to the hospital that is, that somehow their treatment or the follow up to their treatment or something isn't likely to be sufficient and they're going to wind up being re-hospitalized, which is, as we can imagine a serious problem. It's expensive, it's dangerous for the patients, etc. so there's a lot of reasons why this is an important area. Objective of this research is to design a predictor that predict which hospitalized diabetes patients will be readmitted for their condition at a later date with help of statistical models like logistic regression model In this research, the criteria are the performance to minimize the error function.

I. INTRODUCTION

We are going to work on classifying diabetes patients who have been in a hospital for treatment and we want to classify the ones who are at high risk that they're going to be readmitted to the hospital; that is, that somehow their treatment or the follow up to their treatment or something isn't likely to be sufficient and they're going to wind up being re-hospitalized, which is, as we can imagine a serious problem. It's expensive, it's dangerous for the patients, etc.

We very well know prevention is better than cure so providing an alert regarding their vulunerability towards a specific medication is always be very helpful.

The incidence of diabetes has doubled in the last ten years in the worldwide. About 200 million people are infected and about six percent increase in the annual prevalence of diabetes in the world Human for a long time suffered from different diseases that in some cases, been able to diagnose illnesses and offer solutions in order to improve it, but unfortunately, at times, due to lack of diagnosis remains asymptomatic in patients for a long time and may endanger the patient's life. One of the decision support models application is in the medical field and diagnosis of diseases such as diabetes. Delay in diagnosis and prediction of diabetes due to inadequate control of blood glucose increases Capillaries and macrovascular complications risk, ocular diseases and Kidney failure

So proposed a model to predict diabetes that doctors can be useful as a model to help predict diabetes. In this research, examined the relationship between complications in diabetic patients and their properties such as blood glucose, blood pressure, height, weight, and hemoglobin and body mass index of the patients. The aim of this research is to predict complications based on their symptoms.

Based on the Diabetes Research Center reports, the incidence of diabetes has doubled in the last ten years in the worldwide and About 200 million people are infected and about six percent increase in the annual prevalence of diabetes in the world.

Since diabetes is a chronic disease and import irreparable damage to the limbs and vital organs in the body, using intelligent tools can improve detection methods and disease control and is a great help to the doctors. According to the Diabetes Center Research that has shown that with early diagnostics of patients at risk can be prevented 80 percent of chronic complications of type II diabetes or delayed them.

There are two types of diabetes, type Ⅰ and type Ⅱ diabetes, type I diabetes also called insulin dependent and type II diabetes called relative insulin deficiency.

Logistic regression model of new computational methods for machine learning, knowledge

Representation and finally apply the obtained Knowledge in order to predict the response of output are complex systems. Logistic Regression is a very basic method, in particular, it's a very old algorithm. It dates back at least 50 years, but it really works It's simple, it's fast and it often competes with the best machine learning methods it uses the loss function which minimizes the error the ROC Curve (Receiver operating characteristic) was found in 1950, for the first time was used to detect radio signals with noise. Recently it was discovered that the curves have significant applications in making medical decisions.

©IJRASET: All Rights are Reserved

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

II. INTRODUCTION TO DIABETES

Diabetes or diabetes mellitus, is a metabolic disorder (metabolic) in the body. In this disease it destroys the ability to produce insulin in the patient's body or the body becomes resistant to insulin the and therefore the produced insulin cannot perform its own normal function. The key role of insulin is to lower down blood sugar by different mechanisms. There are three main types of diabetes mellitus which are as :

A. Type1 results from the pancreas's failure to produce enough insulin. In Type I diabetes, the destruction of beta cells in pancreatic leads to impaired insulin production this form of diabetic was previously referred to as "insulin-dependent diabetes mellitus" (IDDM) or "juvenile diabetes". The cause is still unknown.

B. Type2 begins with a condition in which cells fail to respond to insulin properly there is a progressive insulin resistance in the body and eventually may lead to the destruction of pancreatic beta cells and defects in insulin production. As the disease progresses a lack of insulin may also develop. This form was previously referred to as "non-insulin-dependent diabetes mellitus" (NIDDM) or "adult-onset diabetes". The most common cause is excessive body weight and not enough exercise.

C. Gestational diabetes is the third main category of diabetic and occurs when a pregnant women without having a previous history of diabetes develop high blood sugar levels.

III. LOGISTIC REGRESSION

Logistic Regression is one of the oldest statistical method which dates back more than 50 years ago. The basic idea of this method is to minimise the error function although this method is old but it really works It is simple, it's fast and it often competes with the best machine learning methods. The error function is

$$\text{Min}\ \frac{1}{n}\sum_{i=1}^{n} log\left(1 + e^{-y_i f(x_i)}\right)$$

using statistical techniques that known as Maximum Likelihood to estimate coefficients and then obtained logistic regression equation

$$\text{Where } f(x_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \ \ldots\ldots \beta_p x_{ip} = \sum_{j=1}^{p} \beta_j x_{ij}$$

For example let us take income predictor in x values will be the values from the data set values of different attributes like age education , weekly hours , race ,country etc. and finding β values that best fit. β are also called as regression coefficients.

$$income(x) = 2 * age + 3 * education + 4 * hours + \cdots$$

$$min_{\beta 1\, \beta 2..\beta n} \sum_{i=1}^{n} log\left(1 + e^{-y_i} \sum_{j=1}^{p} \beta_j x_{ij}\right)$$

IV. METHOD

In this paper first presented a brief introduction to logistic regression how it will minimise the classification error how it will fit the data the key principle in statistical learning theory is the principle of Ockham's razor. Now Ockham's razor is the idea that the best models are simple models that fit the data well. At first the data is taken and prepared so that we can apply our algorithm efficiently and predict well. The required parameter for calculating the function is taken and viewed then these values are used to calculate loss function again emphasis is on minimising the loss function.

Here in our data set if the value >=.5 which means yes there are chances that he/she could have diabetic in future and if the value is less than .5 than the answer is no . We can also that we have done binary classification here 0/1 yes/no. The ROC curve for this logistic regression classifier is shown and analysed in conclusion section.

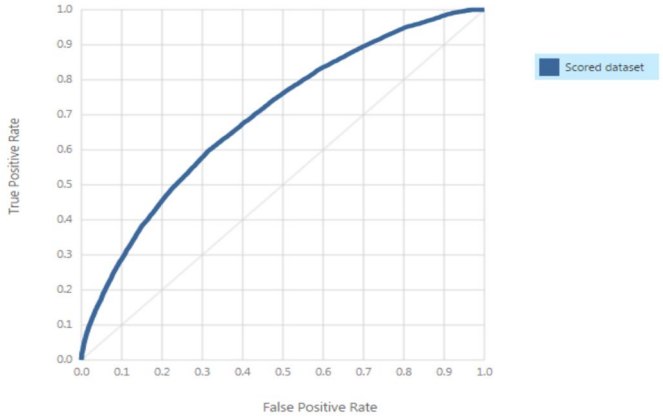
V. CONCLUSION

Logistic regression is a statistical tool used in order to modeling and analyzes the data. After preparing the data we train the model

©IJRASET: All Rights are Reserved

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

for training purpose we split the data a part of data is used in training the model and other is used for evaluation of the model. The ROC curve the model must be above the straight line the staright line in ROC curve shows random guessing which means poor model



Examine this ROC curve. Notice that the bold blue line is well above the diagonal grey line, indicating the model is performing better than random guessing.

Analysing the curve we examine the statistics for the model evaluation, noting the area under the curve (AUC), the number of true positives, false positives, false negatives, and true negatives in the confusion matrix, and the accuracy, recall, precision, and F1 score Correctly classified values (True Positive + True Negative) values should outnumber the errors (False Negative + False Positive).

An Accuracy of greater than 0.5 indicates that the scores are correct more often than not.

There are nearly equal numbers of True Positive and False Negative scores.

The following points should be noticed

A. Correctly classified values (True Positive + True Negative) values should outnumber the errors (False Negative + False Positive).

B. An Accuracy of greater than 0.5 indicates that the scores are correct more often than not.

C. The curve is well above the straight line(m=1) which means this is not random guessing

We have created a classification model using the Two-Class Logistic Regression algorithm. An initial evaluation of the model seems to indicate that it provides better results than random guessing, but more evaluation is needed, and it may be that the model could be improved.

REFERENCES

A. Journal Article Referencing

[1] Prediction of Diabetes by Using Artificial Neural Network, Logistic Regression Statistical Model and Combination of Them

[2] A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management by Kevin Plis, Razvan Bunescu and Cindy Marling Jay Shubrook and Frank Schwartz

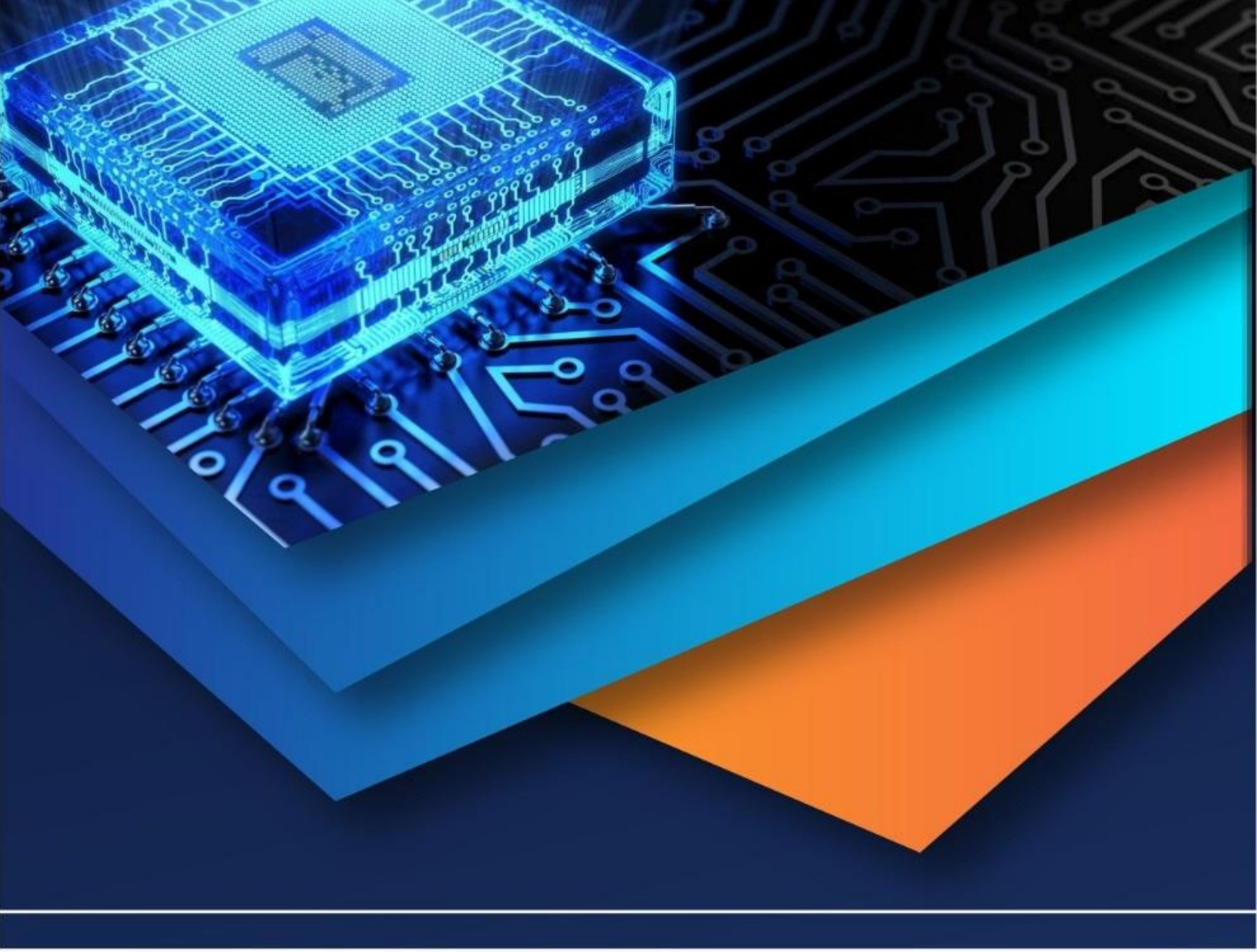
[3] Computational thinking. JM Wing - Communications of the ACM, 2006

B. Book Referencing

[1] Applied logistic regression DW Hosmer Jr, S Lemeshow, RX Sturdivant

[2] Applied logistic regression analysis S Menard

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)