WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Diabetes Prediction Using Machine Learning

Robin Roy, Alan Prasad, Sherin Mary Andrews

UG student,Bca Final Year, Saintgits College Of Applied Science, Kottayam, 686532, India

A B S T R A C T

Diabetes is an illness that is caused due to the high glucose level in the human body. Diabetes should not be ignored, if we ignore Diabetes then it may cause some serious problems in human beings like heart-related problems, kidney problems, blood pressure, eye damage and it can also affect other organs of the human body. If we can predict before it occurs then we can easily control them. By applying various Machine Learning Techniques we can easily predict Diabetes in the human body. Machine learning techniques Provide better results for prediction by constructing models from datasets collected from patients. The Main algorithms focused on are K-Nearest Neighbour (KNN) Algorithm, Support Vector Machine (SVM), Decision Tree (DT), Gradient Boosting (GB) and, Random Forest (RF) algorithms.

Keywords: K-Nearest Neighbour (KNN) Algorithm, Support Vector Machine (SVM), Decision Tree (DT), Gradient Boosting (GB) and, Random Forest (RF) algorithms.

1. Introduction

World Health Organization (WHO) states that people suffering from Diabetes are particularly from lower-income Countries. It is found among countries like India, China and Canada. Diabetes can be controlled through various health controlling methods. Diabetes can be caused due to high blood levels and it affects the hormone insulin that results in up normal metabolism of crabs and it improves the level of sugar in the blood. Some prediction says that Diabetes is one of the major cause of death in the world. About 400 million people suffering from Diabetes from lower-income Countries. By the year 2030, it can be an increase in numbers from the latest prediction. In India, about 40 Million peoples are diabetic of the population of 100 million. To set the Indian Diabetic Dataset there are various Machine Learning classifications and Technique to predict and analyze diabetes. It is the method that is used with the help of Computers or Machines to get accurate data. It also helps in getting an efficient result to build the knowledge in various classifications from the collected Dataset that can be used for the Diabetes prediction. It is a tough way to choose the best technique from Machine Learning but mostly we apply popular classification and ensemble methods on this type of Dataset for prediction.

2. Diabetes Type

Diabetes is also called a “Touch of sugar”. It is a disease that occurs when blood sugar is too high. The main source of energy comes from the food that we eat. Too much glucose in the blood can cause health problems so takes steps to avoid Diabetes and manage it by a stay in healthy.
Mostly there are two types of Diabetes
Type1: when the body does not make any insulin then the immune system is attacked and destroyed by the cells in the pancreas which make the insulin. It is mostly diagnosed in children ,adults and young people.
Type2: when the body does not use the insulin that produces at any age it can occur. Usually in the age category of middle-aged and older people, which is the most common type of diabetes.

* *Corresponding author.*
E-mail address: robinjrroy123@gmail.com

2.1 Symptoms of diabetes

- Frequent Urination
- Increased Hunger
- Weight loss
- Blurry vision
- Increased Thirst
- Sores that do not heal

3. Methodology

The main aim is to find a model to predict Diabetes with better accuracy. For that different classifications and Algorithms are used to predict Diabetes

3.1 Phases of Prediction

Dataset Description:
It is a collection of data that lists values for each of the variables such as demographic factors like Height and Weight of an object. Each value taken is known as a datum. It also consists of documents or files.

Data Pre-processing:
It is the most important process. It is used for converting the data into the dataset. The major task is to remove the unwanted data and filling the missed value. Machine Learning Techniques on the dataset are essential for accurate results.

Apply Machine Learning:

Different classifications and techniques are used to predict diabetes. The main objective to apply Machine Learning Techniques is to identify the performance of these methods and find accurate data. It plays a major role in prediction.

4. Various Machine Learning Techniques

4.1 Support Vector Machine:

Support Vector Machine is also known as SVM. It creates a hyperplane or set of a hyperplane in high dimensional space that separates two classes. It can be used for classification and regression. It can also classify the entities which are not supported by data. Hyperplane performs the separation to the training point of any class.

Algorithm

i. First select the hyperplane that divides the better class.
ii. We have to calculate the distance between the planes and the data to find the better hyperplane which is called Margin.
iii. If the chance of miss conception is high then the distance between the classes is low
iv. Where the Margin = distance to positive point + Distance to a negative point. To select the class which has a high margin.

4.2 Decision Tree:

The decision tree is a basic classification method in Diabetes prediction. The decision tree can be used when the response variable is categorical. It has a main tree-like structure-based model which describes the classification process. They can be of any type like the graph, discrete, text, continuous, etc.

Algorithm

i. First construct a tree with nodes, in which the nodes an input feature.
ii. The select feature predicts the output from the input feature.
iii. From each node of the tree highest information gain is calculated.
iv. Repeat step 2 to get a subtree and use the feature which is not being used in previous nodes.

5. Conclusion

In today's scenario, diabetes has great importance in concerning its severe complications. Nowadays Diabetes is one of the biggest reasons for death

worldwide. The main objective focuses on the identification of diabetes using Machine Learning algorithms. It is mainly useful for people like physicians to predict diabetes in its initial stage. From the analyses of machine learning, conventional treatments and solutions are given to the patients. This system makes more relevant results. A diabetes prediction system is useful for hospitals and doctors. So that doctors can treat patients in a better way. By the use of machine learning algorithms for Diabetes prediction, we will get more accurate and efficient results.

REFERENCES

Iqbal H. Sarker, *"Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus"*. International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February 2019.

Rahul Joshi, *"Analysis and prediction of diabetes diseases using machine learning algorithm"*: International Research Journal of Engineering and Technology Volume: 04 Issue: 10 | Oct -2017.

Dheeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, *"Diabetes Disease Prediction Using Data Mining "*.International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017.

K.VijiyaKumar, B.Lavanya, I.Nirmala, S.Sofia Caroline, *"Random Forest Algorithm for the Prediction of Diabetes "*.Proceeding of International Conference on Systems Computation Automation and Networking, 2019.

Nonso Nnamoko, Abir Hussain, David England, *"Predicting Diabetes Onset: an Ensemble Supervised Learning Approach "*. IEEE Congress on Evolutionary Computation (CEC), 2018.

Tejas N. Joshi, Prof. Pramila M. Chawan, *"Diabetes Prediction Using Machine Learning Techniques"*.Int. Journal of Engineering Research and Application, Vol. 8, Issue 1, (Part -II) January 2018, pp.-09-13.

International Journal of Advanced Research in Computer and Communication Engineering. Vol. 9, Issue 7, July 2020.