Prediction of Diabetes Using Machine Learning

Ch.Niharika, A.Hima Bindu, A.Mounika, G.Kiranmai

Abstract : Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the whole world. By 2035, this will be doubled as 592 million *COVID*-19 pandemic has rapidly affected our day-to-day life disrupting the world trade and movements. In this situation diabetic patient standing in the queue is a critical issue. Diabetes mellitus or simply diabetes is a disease caused due to the increase level of blood glucose. Various traditional methods, based on physical and chemical tests, are available for diagnosing diabetes. However, early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors as diabetes affects human organs such as kidney, eye, heart, nerves, foot etc. Data science methods have the potential to benefit other scientific fields by shedding new light on common questions. One such task is to help make predictions on medical data. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience. The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques. This project aims to predict diabetes via four different supervised machine learning methods including: KNN, Logistic regression, Decision Tree, Random Forest. This project also aims to propose an effective technique for earlier detection of the diabetes disease.
Keywords: Machine Learning, Supervised, Knn, Decision tree, Random forest, Logistic Regression

* Correspondence Author
 Ch.Niharika ,Assistant Professor, Department of CSE,
Usha Rama College of Engineering and Technology,
India.
Email:chava.niharika@gmail.com
A.Hima Bindu, Department of CSE,
Usha Rama College of Engineering and Technology,
India,
Email:himabindualuri001@gmail.com
A.Mounika, Department of CSE,
Usha Rama College of Engineering and Technology,
India,
Email:anumalamounika1892@gmail.com
G,Kiranmai, Department of CSE,
Usha Rama College of Engineering and Technology,
India,
Email:kiranmaigade@gmail.com

Prediction of Diabetes Using Machine Learning

1. Introduction

Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood sugar. Hyperglycaemia, or raised blood sugar, is a common effect of uncontrolled diabetes and, over time, leads to severe damage to many of the body's systems, especially the nerves and blood vessels. In 2014, 8.5% of adults aged 18 years and older had diabetes worldwide. In 2016, diabetes was the direct cause of 1.6 million deaths, and in 2012 high blood glucose was the cause of another 2.2 million deaths. There are two types of diabetes, namely, Type 1 and Type 2. Type 1 diabetes (previously known as insulin-dependent, juvenile, or childhood-onset) is characterized by deficient insulin production and requires daily administration of insulin. The cause of Type 1 diabetes is not known, and it is not preventable with current knowledge. Type 2 diabetes (formerly called noninsulin-dependent or adult-onset) results from the body's ineffective use of insulin. Type 2 diabetes comprises what the majority of people with diabetes around the world have and is mostly the result of excess body weight and physical inactivity.

According to the World Health Organization (WHO), to help prevent type 2 diabetes and its complications, people should achieve and maintain healthy body weight; be physically active—at least 30 minutes of regular, moderate-intensity activity on most days. More activity is required for weight control, eat a healthy diet, avoiding sugar and saturated fats intake, and avoid tobacco use—smoking increases the risk of diabetes and cardiovascular diseases. Treatment of diabetes involves diet and physical activity, along with lowering blood glucose and the levels of other known risk factors that damage blood vessels .It is evident from the WHO reviewed literature that the incidence of diabetes mellitus is ever increasing throughout the world in both developed and developing countries. A significant number of people living in both developed and developing countries are ever becoming sedentary. Moreover, although there is evidence that the complications of diabetes can be prevented, there are still diabetics who lack the required knowledge and skills to manage and control their condition using available technology for healthy living and changed lifestyles. Changing lifestyles require deliberate effort. Therefore, diabetics must take the ultimate responsibility for their care and treatment using available technology-related systems. Technologies such as meal recommendation systems, physical activity monitoring and tracking, notification systems for taking drugs, and that they may have about their condition. The machine learning and computer vision have made producing applications to automate tasks requiring intelligent behaviour, learning, and adaptation possible, hence, providing solutions to real-life problems such as diabetes management. The diabetes prediction system aims at determining the required nutrition of patients and recommends meals to meet these needs, notify patients to take their medication on time, identify the factors that influence blood sugar, and educate patients on them. The developed system also encourages patients to exercise and keep track of their activities and keep doctors updated on the patient's condition by sending those pdfs of patient's data securely. These proposed objectives of the software modules help in delivering the required assistance in diabetes management through software engineering technology and machine learning. In machine learning we are using knn, logistic regression, decision tree, random forest algorithms for prediction.

2. DESIGN

Systems design is the method by which the architecture, elements, modules, interfaces and information of a system are defined to meet specific requirements. It could be viewed as applying the system theory to product development. Object-oriented evaluation and design methods become the most widely used computer system design techniques. The project contains following modules:
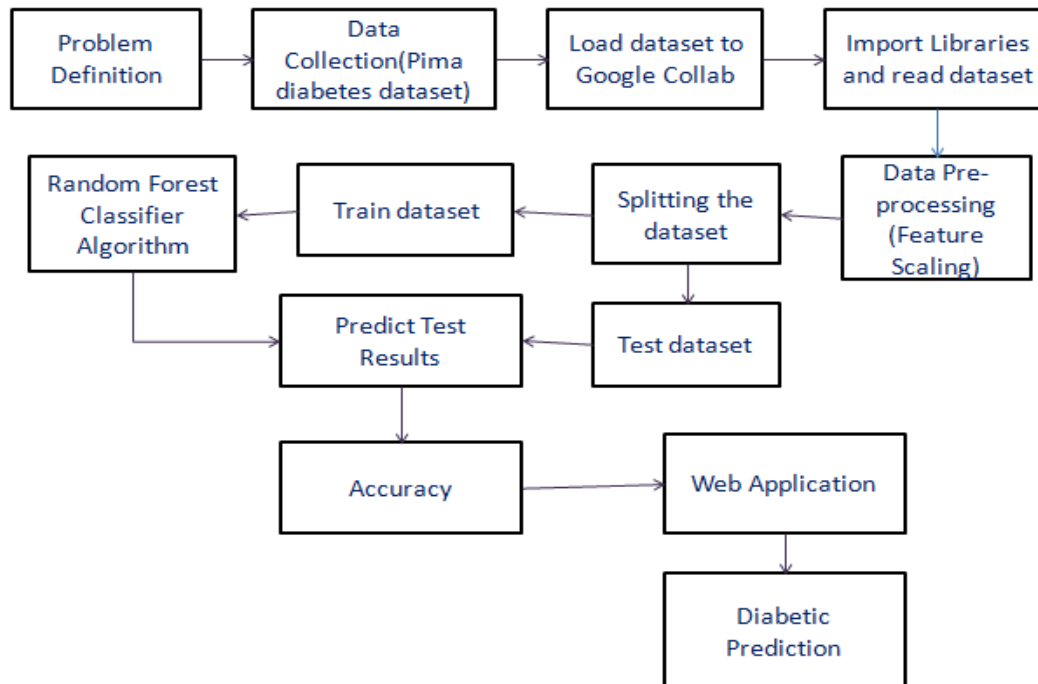
GATHERING DATASET: Dataset is the heart of this project. We gathered dataset from the website called Kaggle.com in which there will be already existing diabetes records of the patients. By using those values we predict the diabetes.

PREPROCESSING DATASET: Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

TRAIN USING RANDOM FOREST CLASSIFIER: With the help of existing dataset we train the data of 75% and 25% for test data. When the user enter the values it directly goes to the dataset and the process of testing and training takes place using random forest classifier. In random forest classifier there will be a number of decision trees which produces the most accuracy.

PREDICTING THE DIABETES: When user enters the values of their health information, the values are compared with the existing dataset. When compared we can finally get the prediction whether it is positive (0) or negative (1).

3. Figure

4. ANALYSIS

 System Analysis is first stage according to System Development Life Cycle model. This System Analysis is a process that starts with the analyst. Analysis is a detailed study of the various operations performed by a system and their relationships within and outside the system. One aspect of analysis is defining the boundaries of the system and determining whether or not a candidate should consider other related systems.

During analysis, data is collected from the available files, decision points, and transactions handled by the present system. Logical system models and tools are used in analysis. Training, experience, and common sense are required for collection of the information needed to do the analysis. Logical system models and tools are used in analysis. Training, experience, and common sense are required for collection of the information needed to do the analysis. This System Analysis is a process that starts with the analyst. Analysis is a detailed study of the various operations performed by a system and their relationships within and outside the system.

Random forest is a most popular and powerful supervised machine learning algorithm capable of performing both classification, regression tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction. Random decision forests correct for decision trees habit of over fitting to their training set. This algorithm runs efficiently on large databases and it has higher classification accuracy .

5. RESULTS

Machine Learning algorithms are applied on the dataset and the classification has been done using various algorithms of which Random Forest gives highest accuracy of 97%. We have seen comparison of machine learning algorithm accuracies with pima datasets. It is clear that the model improves accuracy and precision of diabetes prediction with this dataset compared to existing dataset. For all of these when a web application is created in which it can store all the related diabetes dataset by using Machine Learning which will analyze the data provided by an individual through application and it will advice which type of food and exercises are required to avoid/reduce the diabetes issues. Further this work can be extended to find how likely non-diabetic people can have diabetes in next few years.

Prediction of Diabetes Using Machine Learning

6. CONCLUSIONS

The main aim of this project was to design and implement Diabetes Prediction Using Machine Learning Methods and Performance Analysis of that methods and it has been achieved successfully. The proposed approach uses various classification and ensemble learning method in which Knn, Random Forest, Decision Tree and Logistic Regression classifiers are used. And 77% classification accuracy has been achieved. The Experimental results can be asst health care to take early prediction and make early decision to cure diabetes and save humans life.

REFERENCES

[1] Ravi Sanakal, Smt. T Jayakumari,-Prognosis of Diabetes Using Data mining Approach- Fuzzy C Means Clustering and Support Vector Machine, International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 2 May 2014.

[2] M. Durairaj, G. Kalaiselvi., "Prediction Of Diabetes Using Soft Computing Techniques- A Survey" in International Journal Of Scientific & Technology Research Volume 4, Issue 03, March 2015.

[3] Veena Vijayan V., Aswathy Ravikumar, "Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus" in International Journal of Computer Applications (0975 – 8887) Volume 95– No.17, June 2014.

[4] Velu C.M, K.R.Kashwan, 2013."Visual Data Mining Techniques for Classification of Diabetic Patients", IEEE International Advance Computing Conference (IACC),pp-1070-1075.