

Diabetes Prediction Using Different Machine Learning Approaches

¹Shejal Kale, ²Priti Rahane, ³Mansi Ghumare, ⁴Snehal Patil

Department of Computer Engineering
Matoshri College of Engineering and Research Centre,
Nashik-422105

Abstract: Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affects other organs of human body. Diabetes can be controlled if it is predicted earlier. To achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by constructing models from datasets collected from patients. In this work we will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. Which are K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Gradient Boosting (GB) and Random Forest (RF). The accuracy is different for every model when compared to other models. The Project work gives the accurate or higher accuracy model shows that the model is capable of predicting diabetes effectively. Our Result shows that Random Forest achieved higher accuracy compared to other machine learning techniques Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affects other organs of human body. Diabetes can be controlled if it is predicted earlier. To achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by constructing models from datasets collected from patients. In this work we will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. Which are K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Gradient Boosting (GB) and Random Forest (RF). The accuracy is different for every model when compared to other models. The Project work gives the accurate or higher accuracy model shows that the model is capable of predicting diabetes effectively. Our Result shows that Random Forest achieved higher accuracy compared to other machine learning techniques

Keywords: Diabetes, Machine, Learning, Prediction, Dataset, Ensemble

INTRODUCTION

Diabetes is noxious diseases in the world. Diabetes caused because of obesity or high blood glucose level, and so forth. It affects the hormone insulin, resulting in abnormal metabolism of carbs and improves level of sugar in the blood. Diabetes occurs when body does not make enough insulin. According to (WHO) World Health Organization about 422 million people suffering from diabetes particularly from low or idle income countries. And this could be increased to 490 billion up to the year of 2030.

MOTIVATION

Diabetes is major cause of death in the world. Early prediction of disease like diabetes can be controlled and save the human life. To accomplish this, this work explores prediction of diabetes by taking various attributes related to diabetes disease. For this purpose we use the Pima Indian Diabetes Dataset, we apply various Machine Learning classification and ensemble Techniques to predict diabetes.

LITRATURE SURVEY

This chapter discuss brief literature regarding the project. Literature survey is mainly used to identify information relevant to the project work and know impact of it within the project area. It defines as till yet how many surveys have been done knowledge of latest technology and implementation designs.

Sr.no	Title	Details
1	implementation of Diabetic Retinopathy Prediction System using Data Mining	Diabetic retinopathy (DR) is the most common cause of newly diagnosed blindness every year. Annual eye checking for diabetic patients are suggested in order to find and treat DR in a timely manner, since blindness from this condition is preventable with early identification.
2	Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining	The disease predictions have been explored using various methods of data mining. The use of medical data set on the prediction of diabetic mellitus has been analyzed.

LIMITATION OF EXISTING SYSTEM

- Costing: The Existing system is high cost and this is main reason most of the system is failed.
- Technology Complexity: Most of system is the complex to understand, Not user friendly as compare to our proposed system
- Time Consuming Feature: In existing system, the performance is low and most of the time system gets hanged due to load.
- Not Easy to Understand: Systems re complex to understand and they were not user friendly

EXPERIMENTAL SETUP

Secure Login System

- The system is containing the connections with the user system so login and registration system must be full proof.
- To make this happen, user will register first and then process for login model 2.Data set Uploading
- we are providing the user with a feature to upload the data set and wait for the result 3.Prediction model
- prediction model of churn is process it will process the current dataset and uploaded dataset and give output.

Hardware and Software Requirements

1. Hardware Requirements

- Processor core i3
- RAM 4gb
- HDD 500 gb

2. Software Requirements

- MySQL Database MySQL is an open source relational database management system(RDBMS).It uses a standard form of the well known SQL data language and works quickly and works will even with huge dataset
- Language: Python Python is a popular programming language. It was created by Guido van Rossum, and released in 1991. It is used for: (a) web development (server-side). (b) software development.
- Operating System: Windows 8 and above Windows is the most widely used operating system for desktop and laptop computers. Develop by Microsoft, Windows preliminary runs on x86 based computers . Windows provides Graphical User Interface and desktop Environment in which application displayed in resizable,movable windows on screen.

SCOPE:

R detection is solely based on existing patient records. Now a day's medical data growing tremendously and we need to process that data for detection. But it is time consuming hence data mining techniques helps to get rid from this issue. We use neural network (NN) and naïve bayes for classification. According to comparison results NN gives better accuracy than naïve bayes and time and memory required for NN is less as compared to native bayes.

PROBLEM STATEMENT:

Diabetes is a most common disease caused by a group of metabolic disorders. It is also known as Diabetic mellitus. It affects the organs of the human body. It can be controlled by predicting this disease earlier. If diabetics patient is untreated for a long time, it may lead to increase blood sugar. Now a days, Healthcare industries generating large volume of data. Machine Learning algorithms and statistics are used to predict the disease with the help of current and past data. Machine learning techniques helps the doctors to predict early stage for diabetics. Diabetics patient medical record and different types of algorithms are added in dataset for experimental analysis. we use logistic regression, random forest, decision tree classifier and gradient boosting to predict whether a patient has diabetes based on diagnostic measurements. Performance and accuracy of the applied algorithms is discussed and compared.

SYSTEM ARCHITECTURE

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. The algorithms like K nearest neighbour, Logistic Regression, Random forest, Support vector machine and Decision tree are used. The accuracy of the model using each of the algorithms is calculated. Then the one with a good accuracy is taken as the model for predicting the diabetes.

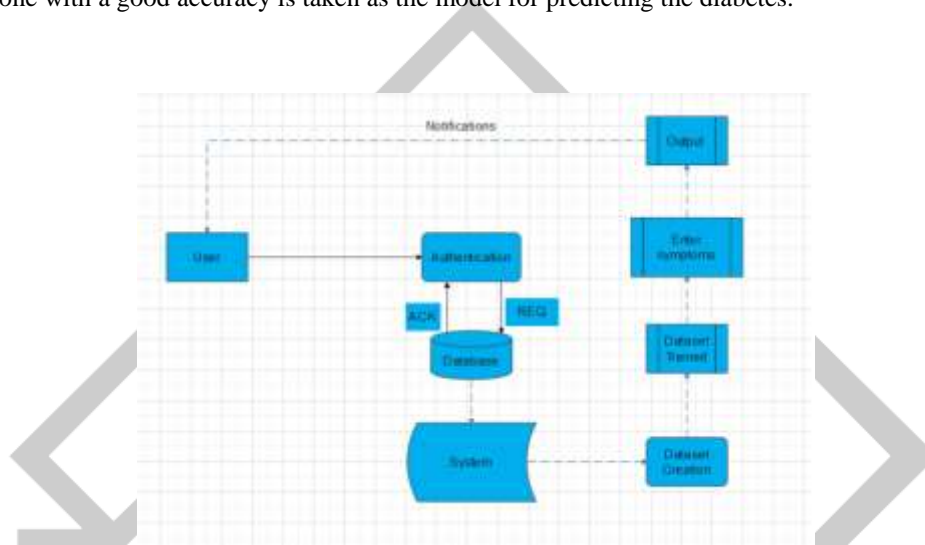


Fig -1: System Architecture Diagram

The presence of disease has been identified using the appearance of various symptoms. However, the methods use different features and produces varying accuracy. The result of prediction differs with the methods/measures/ features being used. Towards diabetic prediction, a Disease Influence Measure (DIM) based diabetic prediction has been presented. The method pre processes the input data set and removes the noisy records. In the second stage, the method estimates disease influence measure (DIM) based on the features of input data point. Based on the DIM value, the method performs diabetic prediction. Different approaches of disease prediction have been considered and their performance in disease prediction has been compared. The analysis result has been presented in detail towards the development.

ADVANTAGES

1. Easy to use
2. High Performace
3. Scalable

METHODOLOGY

The single problem can be solved by different solutions. This considers the performance parameters for each approach. Thus considers the efficiency issues.

- Problem Solving Methods are concerned with efficient realization of functionality. This is an important characteristics of Problem Solving Methods and should be deal with it explicitly.
- Problem Solving Methods achieve this efficiency by making assumptions about resources provided by their context (such as domain knowledge) and by assumptions about the precise definition of the task. It is important to make these assumptions explicit as it give the reason about Problem Solving Methods.

- The process of constructing Problem Solving Methods is assumption-based. During this process assumptions are added that facilitate efficient operation rationalization of the desired functionality

CONCLUSION

In our project the result is classified into Yes or No. If the result is classified into No then we use time prediction module. Time Prediction - here we predict the "time" of getting the diabetes disease. We analyze the result of the diabetes prediction and check the accuracy of the diabetes prediction, time taken to compute the accuracy of the diabetes prediction, correctly classification and incorrectly classification of result of the diabetes prediction. We have used KNN Algorithm to predict the diabetes where result is classified into Yes or No and also for time prediction module same KNN Algorithm is used. We compared the testing data and actual data to get the accuracy of our project

REFERENCES

- Azra Ramezankhani, Omid Pournik, Jamal Shahrabi, Fereidoun Azizi and Farzad Hadaegh, "An Application of Association Rule Mining to Extract Risk Pattern for Type 2 Diabetes Using Tehran Lipid and Glucose Study Database", Int J Endocrinol Metab, April 2015.
- Han Wu, Shengqi Yang, Zhangqin Huang, Jian He and Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", Elsevier Informatics in Medicine, vol. 10, pp. 100-107, 2018.
- A Abbasi, LM Peelen, E Corpeleijn, YT van der Schouw, RP Stolk, AM Spijkerman et al., "Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study", BMJ, 2012. • et al., "Mining constrained association rules to predict heart disease", IEEE 13th International Conference on Data Mining, pp. 433, 2001.
- M. F. Faruque, Asaduzzaman and I. H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus", 2019 International Conference on Electrical Computer and Communication Engineering (ECCE), pp. 1-4, 2019.
- M Sabibullah, V Shanmugasundaram and Priya K Raja, "Diabetes Patient's Risk through Soft Computing Model", International Journal of Emerging Trends Technology in Computer Science, vol. 2, no. 6, 2013.
- G. Tripathi and R. Kumar, "Early Prediction of Diabetes Mellitus Using Machine Learning", 2020 8th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1009- 1014, 2020.
- VZ. Tafa, N. Pervetica and B. Karahoda, "An intelligent system for diabetes prediction", 2015 4th Mediterranean Conference on Embedded Computing (MECO), pp. 378-382, 2015.