

DIABETES PREDICTION USING MACHINE LEARNING

Project Members:- 2127062 Shreyash Somvanshi

2127063 Sujay Shinde

2127076 Prajwal Rudrapwar

2127073 Vedanti Mane

2127032 Trupti Kharat

Guided by: Prof. Rajkumar Panchal sir

Department of Artificial Intelligence & Data Science
Vidya Pratishthan's Kamalnayan Bajaj Institute of College and Technology
Vidyanagari, Baramati-413133

INTRODUCTION

Motivation of the Project

Motive of this project is to use the Artificial Intelligence technologies like Machine Learning in healthcare sector to increase efficiency and accuracy of results.

Literature Survey

This Diabetes Prediction Model makes use of various classification Algorithms to categorize patients into diabetic and non-diabetic. Currently we have achieved 80% accuracy but it can be made more precise with help of appropriate data processing.



PROBLEM STATEMENT

Make use of new emerging technologies in the healthcare to reduce time and efforts.

Goals and Objectives

To use modern technologies to increase accuracy and automation in healthcare.

Statement of Scope

This Diabetes Prediction using Machine Learning works for Healthcare domain. It has 80% accuracy in predicting diabetes based on the factors like patient's insulin, glucose level, age, etc. factors. This Diabetes Prediction Model makes use of various classification Algorithms to categorize patients into diabetic and non-diabetic.



ABSTRACT

Diabetes is a chronic disease with the potential to cause a worldwide healthcare 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. Diabetes is a disease caused due to the increase level of blood glucose. This high blood glucose produces the symptoms of frequent urination, increased thirst, and increased hunger. Diabetes is a one of the leading cause of blindness, kidney failure, amputations, heart failure and stroke.



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. 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.



AREA PROJECT

- This Diabetes Prediction using Machine Learning works for Healthcare domain. It has 79% accuracy in predicting diabetes based on the factors like patient's insulin, glucose level, age, etc.

Technical Keywords

Machine Learning, Diabetes, Decision tree, Train-Test-Split, K nearest neighbour, Logistic Regression, Support vector Machine, Accuracy



LITERATURE SURVEY

- The analysis of related work gives results on various healthcare datasets, where analysis and predictions were carried out using various methods and techniques. Various prediction models have been developed and implemented by various researchers using variants of machine learning algorithms.
- Aiswarya Iyer (2015) used classification technique to study hidden patterns in diabetes dataset. Naïve Bayes and Decision Trees were used in this model. Comparison was made for performance of both algorithms and effectiveness of both algorithms was shown as a result. B.M. Patil, R.C. Joshi and Durga Toshniwal (2010) proposed Hybrid Prediction Model which includes Simple K-means clustering algorithm, followed by application of classification algorithm to the result obtained from clustering algorithm. In order to build classifiers C4.5 decision tree algorithm is used.



METHODOLOGY

- 1 Import the dataset with various patient records on parameters like age, insulin, glucose, blood pressure.
- 2 Clean the dataset i.e. remove the unwanted constraints (Preprocessing)
- 3 Perform train-test-split on processed dataset.
- 4 Use the preferred algorithm.
- 5 Check Accuracy
- 6 Predict and Compare



DATASET DESCRIPTION

- Dataset used for this model is located on
<https://raw.githubusercontent.com/ShreyashSomvanshi/Datasets/main/Diabetes.csv>
- It consists the detailed records of patients. The attributes in datasets are Age, Insulin, Pregnancies, Glucose, DPF, BMI, Blood pressure, etc.
- Dataset consists of 768 rows and 9 columns



LIBRARIES and FUNCTIONS used:

Libraries:

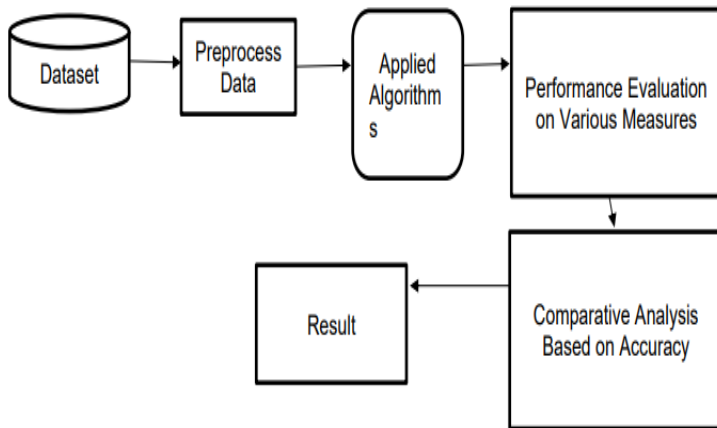
- 1 pandas
- 2 numpy
- 3 sklearn

Functions/Modules used from sklearn:

- 1 `sklearn.model_selection import train_test_split`
- 2 `sklearn.metrics import mean_absolute_error`
- 3 `sklearn.linear_model import LogisticRegression`
- 4 `sklearn.linear_model import LinearRegression`
- 5 `sklearn.metrics import accuracy_score`
- 6 `sklearn.neighbors import KNeighborsClassifier`
- 7 `sklearn.naive_bayes import GaussianNB`
- 8 `sklearn.svm import SVC`
- 9 `sklearn.tree import DecisionTreeClassifier`



ARCHITECTURE



Proposed Model Diagram

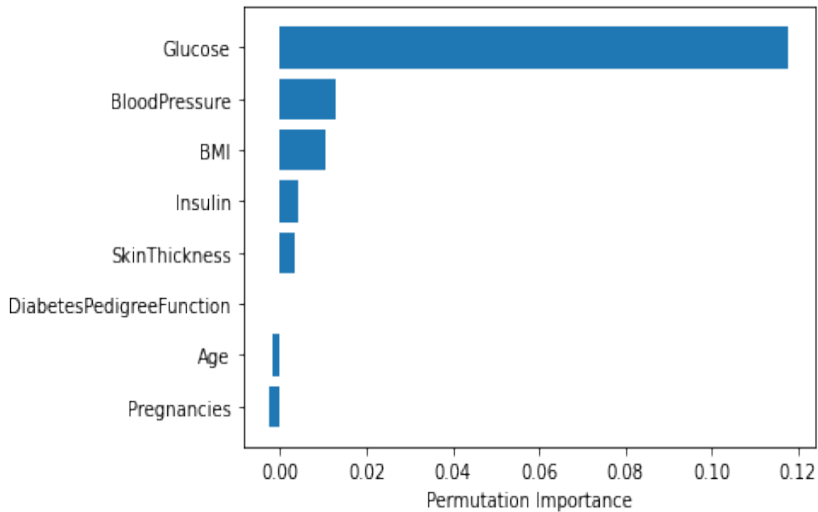


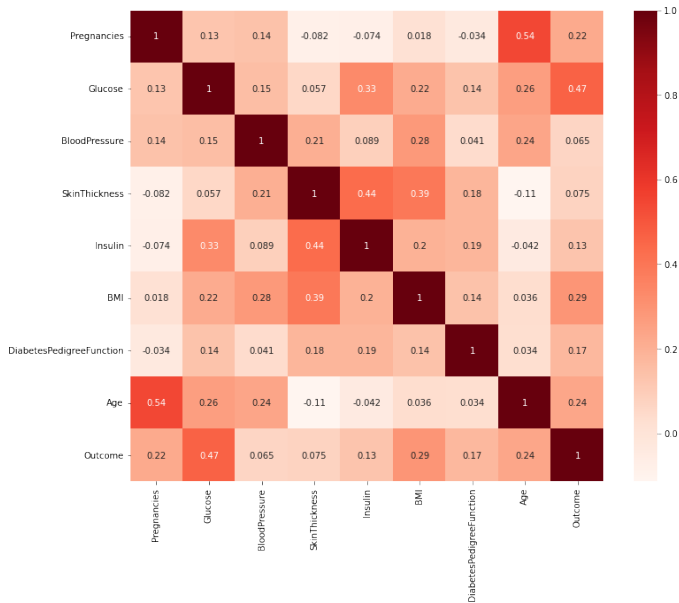
Algorithm's Accuracy:

Algorithm	Testing Accuracy	Training Accuracy
Logistic Regression	0.7792	0.7636
Naive Bayes	0.7619	0.7673
Decision Tree	0.7533	1.0
SVM	0.7489	0.7896
KNN Classifier	0.7489	0.7896



Result Analysis





APPLICATION

As the technology in each domain is progressing rapidly, this Machine Learning based Diabetes Prediction Model can also come in normal usage. This can also be updated with the future technological trends. It can adapt the changes without much complications. Now-a-days the covid self testing kits are available in the market similarly this can also be used for testing the Diabetic diseases without going to the hospitals.



CONSTRAINTS

- Medical tests are required to get the accurate values of insulin , glucose , blood pressure of patients.
- It does not give 100% accurate results.
- Results may change.



SOFTWARE and HARDWARE RESOURCES:

Software Resources:

- Python 3.9
- Jupyter notebook
- Google Colab
- Github

Hardware Resources:

- A Desktop or Laptop with minimum 4GB RAM 500 MB Storage space and intel core i3 processor.
- Smartphone with internet connectivity.



CONCLUSION

One of the important real-world medical problems is the detection of diabetes at its early stage. In this study, various machine learning algorithms are applied on the dataset and the classification has been done using various algorithms. In this study, systematic efforts are made in designing a system which results in the prediction of diabetes. During this work, four machine learning classification algorithms are studied and evaluated on various measures. Experiments are performed on YBI Foundation Diabetes Database. Experimental results determine the adequacy of the designed system with an achieved accuracy of 79% Using Decision Tree Algorithm. In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.



THANK YOU

