

# Prediction of Diabetes Using Machine Learning

## ABSTRACT

Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. 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.

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 based on the factors that are known to be associated with diabetes in a human body.

## INTRODUCTION

### *Overview*

Various health ailments such as diabetes can be predicted using numerous methods when it comes to the field of Machine Learning. In our project, keeping our scope, knowledge and experience, we have proceeded with the task. Testing the model using various Machine Learning algorithms such as decision trees, random forests and neural networks has been done. The model with the best accuracy has been chosen as the best-fit model.

### *Purpose*

Our project aims to predict if an individual is suffering from diabetes or not by taking the required inputs from him/her. If the individual is predicted to have diabetes, a piece of advice is given to keep it in control. Otherwise, some mandatory precautions that are to be taken to keep oneself away from diabetes are suggested.

## LITERATURE SURVEY

### *Existing problem:*

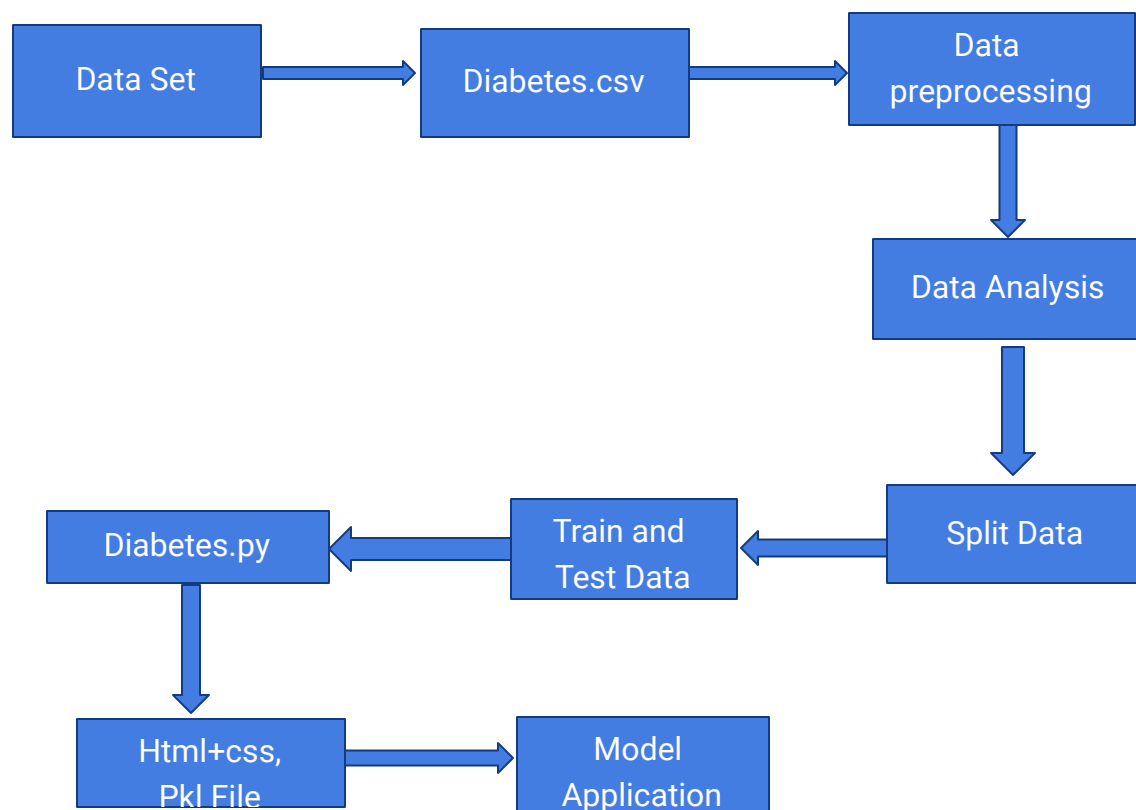
Due to its continuously increasing occurrence, more and more families are influenced by diabetes mellitus. Most diabetic patients know little about their health quality or the risk factors they face prior to diagnosis.

### *Proposed Solution:*

This model, by taking the required inputs, predicts whether a person has diabetes or not. If a person has diabetes then this model gives some advice to the person or else gives some precautions to the person.

## THEORITICAL ANALYSIS:

### *Block Diagram:*



**Hardware/Software designing:**

- Strategy: Matching the problem with the solution.
- Dataset preparation and preprocessing- Data collection, data visualization, labeling, data selection, data preprocessing, data transformation.
- Dataset splitting into train data and test data.
- Modeling- Model training, model evaluation and testing. Improving predictions with ensemble methods.
- Model deployment.

**EXPERIMENTAL INVESTIGATION:**

During our investigation, we got to know all the parameters are required to predict whether a person has diabetes or not. All the parameters are equally important.

**RESULT**

Based on all the input parameters- Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction and Age, the model predicts whether the individual has diabetes or not. If the individual has diabetes some advice is displayed and if the individual does not have diabetes, some precautions will be displayed.

**ADVANTAGES AND DISADVANTAGES****Advantages:**

- After prediction, the individual can immediately go to treatment if he has Diabetes.
- Mostly all predictions are accurate.
- Easy and user friendly User Interface.

**Disadvantages:**

- If user enter wrong inputs by mistake, the model might give a wrong prediction. In case diabetes treatment is not started at the earliest, the individual's health will

be affected.

## APPLICATIONS

- This model can be used to predict the prescence of Diabetes.
- This model can be used in pre-clinical and public health screening programs.

## CONCLUSION

Diabetes is a heterogeneous group of diseases. It's characterized by chronic elevation of glucose in the blood. Diabetes mellitus is a disease, which can cause many complications. How to exactly predict and diagnose this disease by using machine learning is worth studying. The main motto of our project is to Diabetes Milletus Prediction. The result, which only used fasting glucose, has a better performance especially in Luzhou dataset. It means that fasting glucose is the most important index for prediction. But using fasting glucose alone we cannot achieve the best result. So if want to predict accurately, we need more indexes. In addition, by comparing the results of three classifications, we can find there is not much difference among random forest, decision tree and neural network, but random forests are obviously better than the rest of the algorithms in some cases.

## FUTURE SCOPE

Various other key features in the medical records like type of diabetes, family history(hereditary information), emotianal status, lifestyle habits need to be collected and analysed. It will be interesting to perform a more exhaustive exploration of additional features in the dataset.

## BIBILOGRAPHY APPENDIX

- [Dataset](#)
- Model

## APPLICATION BUILDING

- HTML5 and CSS3
- Flask