

BVRIT HYDERABAD College of Engineering for Women

Department of Information Technology

Team 1

A.S.Anusha - 18WH1A1201

M.V.Harini- 18WH1A1233

D.Harshitha - 18WH1A1253

K.Divya- 18WH1A1258

Under the guidance of :

Guide Name: Dr T.Subetha

Designation: Assistant Professor

Title of the project



Agenda

- Abstract
- Introduction
- Problem Definition
- About Dataset
- Flow of the Model
- Literature Survey
- References

Abstract

Diabetes mellitus is a chronic disease characterized by hyperglycemia. It may cause many complications. In this project, we focus on models to determine whether a patient admitted to an ICU has been diagnosed with a particular type of diabetes, Diabetes Mellitus. Since the dataset has 371 attributes, we use feature engineering to optimise the data and then implement using various ML algorithms like decision trees, random forests and LightGBM to predict diabetes mellitus.

Introduction

- Diabetes mellitus (DM), is a group of metabolic diseases in which there are high blood sugar levels over a prolonged period.
- According to the growing morbidity in recent years, in 2040, the world's diabetic patients will reach 642 million.
- We focused on model to determine whether a patient admitted to an ICU has been diagnosed with a particular type of diabetes, Diabetes Mellitus.

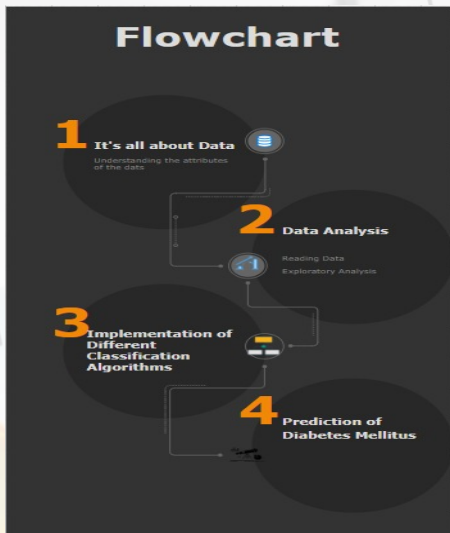
Problem Definition

- The constant hyperglycemia of diabetes is related to long-haul harm, brokenness, and failure of various organs, particularly the eyes, kidneys, nerves, heart, and veins.
- The objective of this project is to make use of significant features, design a prediction algorithm using Machine learning and find the optimal classifier to give the closest result comparing to clinical outcomes.
- The proposed method aims to focus on selecting the attributes that aid in early detection of Diabetes Mellitus using Predictive analysis.

About Dataset

- The dataset consists of 1,30,157 rows and 317 columns in our dataset with different data types like int, float, string etc. . .
- **Splitting of Data**
 - 70% of data is used for training
 - 30% of data is used for testing.


Flow of the Model



Literature Survey

Title	Outcome
<p>Analysis of diabetes mellitus for early prediction using optimal features selection.</p> <p>Author: N. Sneha & Tarun Gangil</p>	<p>The point of this examination is to the finding of diabetes illness, which is a standout amongst the most vital infections in the restorative field utilizing Generalized Discriminant Analysis (GDA) and Least Square Support Vector Machine</p>
<p>Predicting Diabetes Mellitus With Machine Learning Techniques.</p> <p>Author: Quan Zou</p>	<p>Principal component analysis (PCA) and minimum redundancy maximum relevance (mRMR) to reduce the dimensionality. The results showed that prediction with random forest could reach the highest accuracy when all the attributes were used.</p>
<p>Prediction of Diabetes using Classification Algorithms</p> <p>Author: Deepti Sisodia and Dilip Singh Sisodia</p>	<p>Decision Tree, SVM and Naive Bayes are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD) which is sourced from UCI machine learning repository.</p>

References

- 
- <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0175-6>
 - <https://www.frontiersin.org/articles/10.3389/fgene.2018.00515/full>
 - <https://www.sciencedirect.com/science/article/pii/S1877050918308548>

