BVRIT HYDERABAD College of Engineering for Women

Department of Information Technology

Team 1

A.S.Anusha - 18WH1A1201

M.V.Harini- 18WH1A1233

D.Harshitha - 18WH1A1253

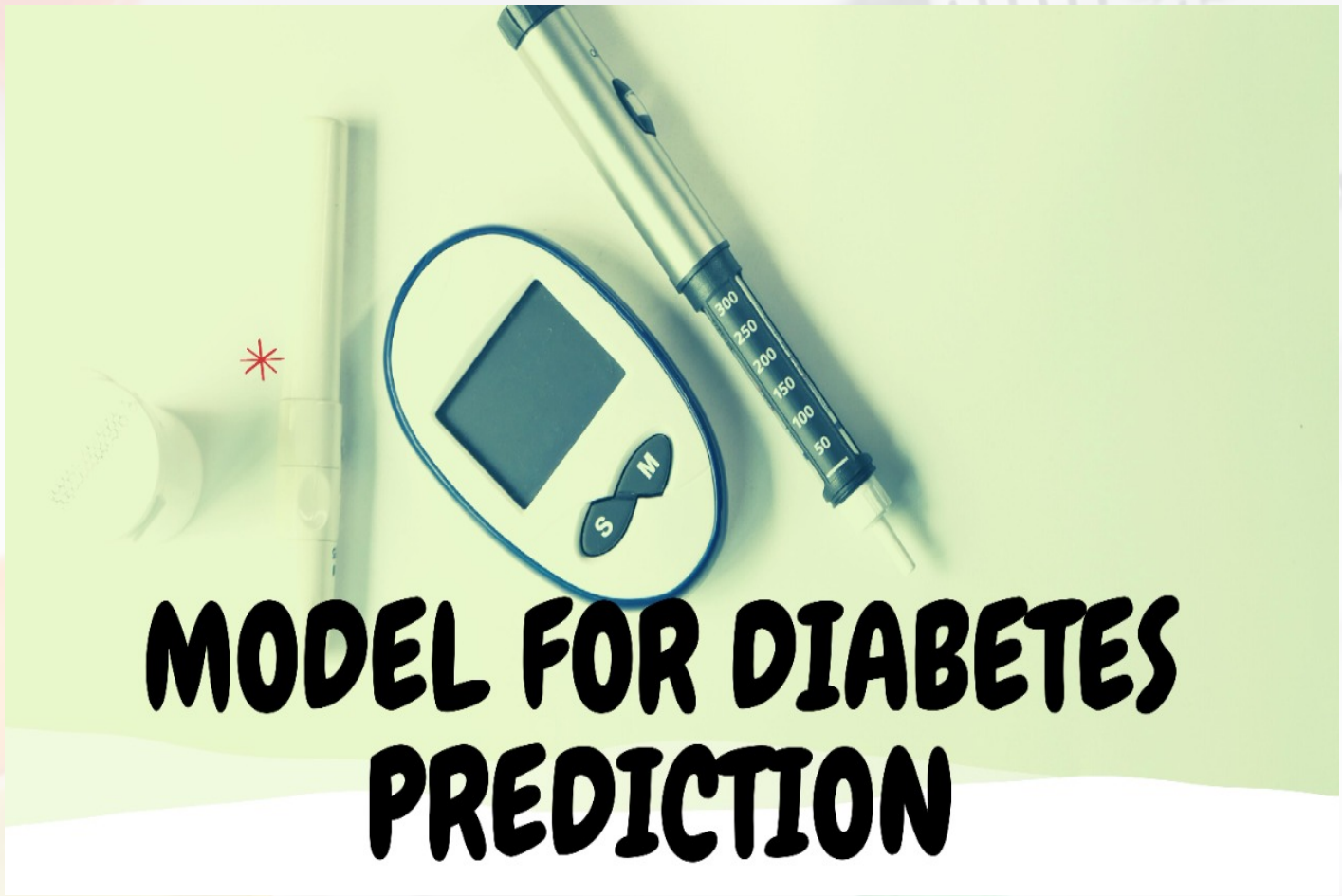
K.Divya- 18WH1A1258

Under the guidance of :

Guide Name: Dr T.Subetha

Designation: Assistant Professor

Title of the project



Agenda

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

Abstract

Diabetes mellitus is a chronic disease characterized by hyperglycemia. It may cause many complications.In this project, we focus on building machine learning 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 predict diabetes mellitus using various ML algorithms like decision trees,random forests and LightGBM.

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 ail in early detection of Diabetes Mellitus using Predictive analysis.

Literature Survey

Title	Outcome
Analysis of diabetes mellitus for early prediction using optimal features selection. Author: N. Sneha & Tarun Gangil	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
Predicting Diabetes Mellitus With Machine Learning Techniques. Author: Quan Zou	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.
Prediction of Diabetes using Classification Algorithms Author: Deepti Sisodia and Dilip Singh Sisodia	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.

Literature Survey

Title	Outcome
Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications Author: Umair Muneer Butt,Sukumar Letchumanan, Mubashir Ali	In this paper for diabetes classification and early-stage identification, three different classifiers have been employed, i.e., random forest (RF), multilayer perceptron (MLP), and logistic regression (LR).
Research on Diabetes Prediction Method Based on Machine Learning Author: Jingyu Xue1st,a, Fanchao Min	In this paper,Support Vector Machine (SVM), Naive Bayes classifier and LightGBM to train on the actual data of 520 diabetic patients and potential diabetic patients aged 16 to 90. Through comparative analysis of classification and recognition accuracy, the performance of support vector machine is the best.
Diabetes prediction model based on an enhanced deep neural network Author: Huaping Zhou, Raushan Myrzashova	In this paper the model is mainly built using the hidden layers of a deep neural network to prevent overfitting.It showed much accuracy with DLPD (Deep Learning for Predicting Diabetes) model. The best training accuracy of the diabetes type data set is 94.02174%.

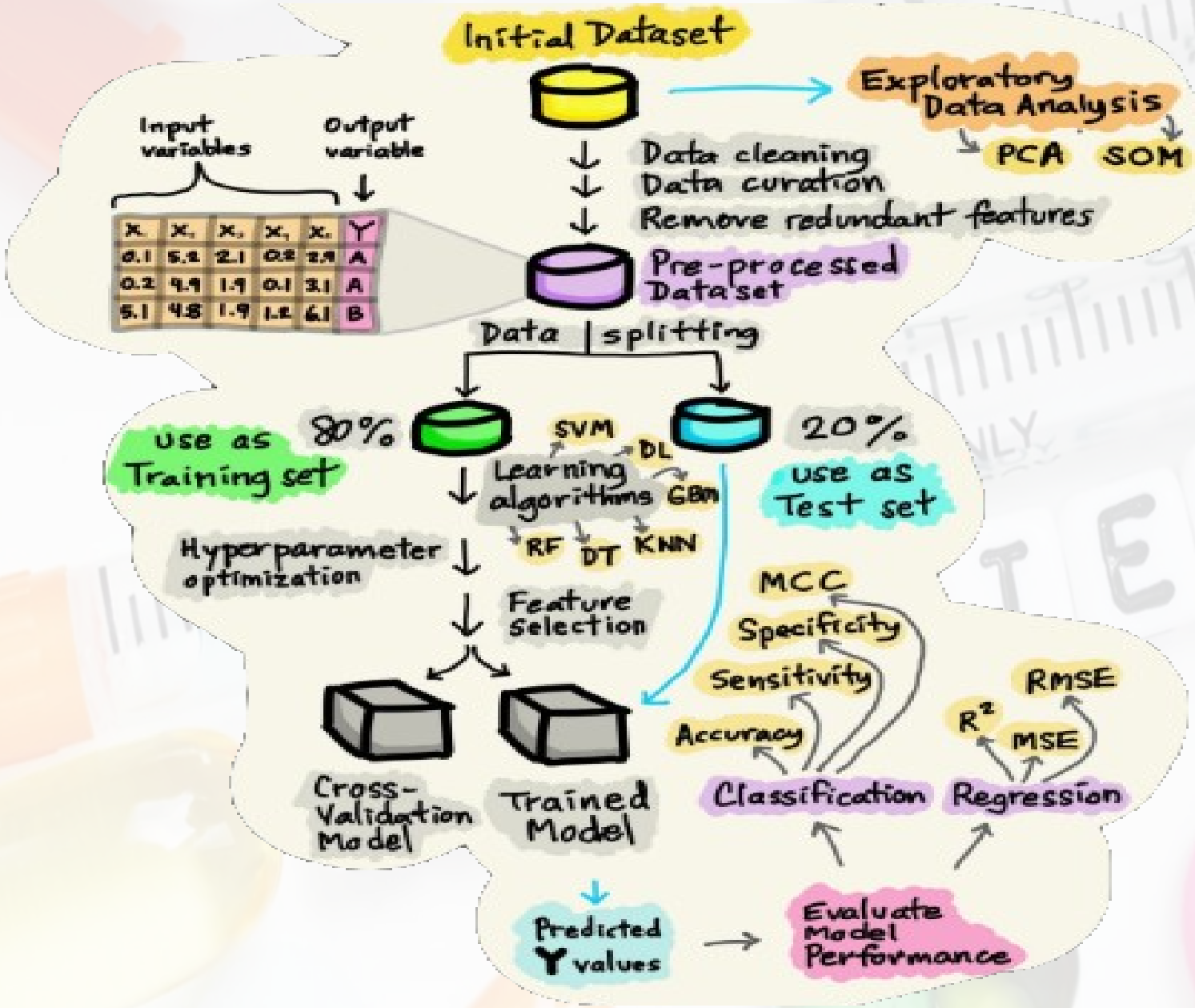
Literature Survey

Title	Outcome
LGBM Classifier based Technique for Predicting Type-2 Diabetes Author: B. Shamreen Ahamed, Dr. Meenakshi Sumeet Arya	The author has used the existing PIMA Indian Dataset for diabetes prediction and detection using LGBM Algorithm. Accuracy is 95.20% Therefore by using the LGBM classifiers, we can develop a data model for diabetes detection and prediction
Prediction of Gestational Diabetes Based on LightGBM Author: Fan Hou,ZhiXiang Cheng	This paper says that it is important to identify the risk of diabetes. Artificial intelligence assisted diabetes genetic risk prediction rematch is selected to construct the LightGBM prediction model and compare with Random Forest and XGBoost on the ROC curve.The results show that the AUC of LightGBM is 85.2%.

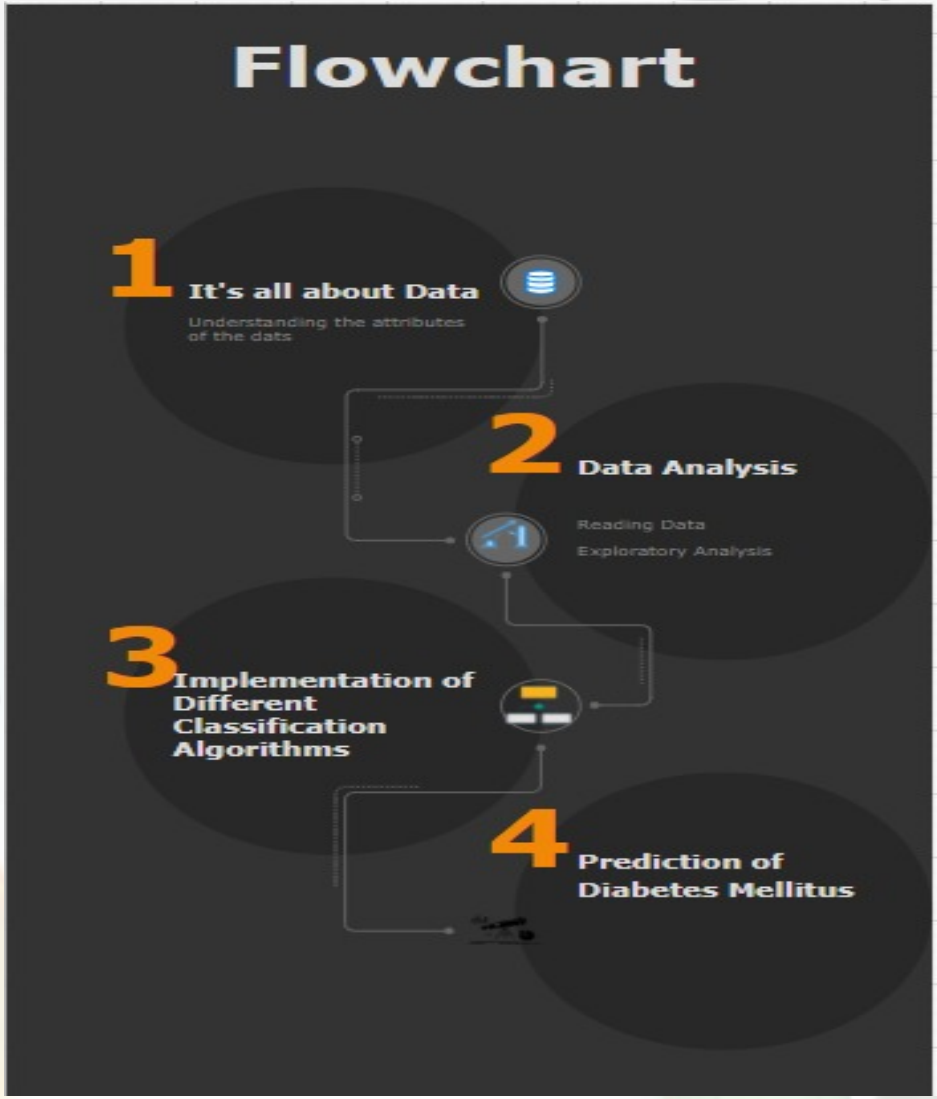
About Dataset

- The dataset consists of 1,30,157 rows and 317 columns with different data types like int, float, string etc. . .
- **Splitting of Data**
  - 80% of data is used for training
  - 20% of data is used for testing.

Design of the project



Flow of the Model



Conclusion

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 SVM, Knn, Random Forest, Decision Tree, Logistic Regression and Gradient Boosting classifiers are used. And 87% classification accuracy has been achieved. The Experimental results can assist health care to take early prediction and make early decision to cure diabetes and save humans life.

References

- https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0175
- https://www.frontiersin.org/arti cles/10.3389/fgene.2018.00515 /full
- https://www.sciencedirect.com/science/article/pii/S1877050918308548
- https://jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-020
- https://www.hindawi.com/journals/jhe/2021/9930985/
- https://iopscience.iop.org/article/10.1088/1742-6596/1684/1/012062/pdf
- https://ejmcm.com/article$_9403_15c24bd9c676c28d90c3fc5fad8b42ea.pdf$
- https://dl.acm.org/doi/10.1145/3433996.3434025


THANK YOU.