Diabetes Prediction

Harsh A Patil.
Manjunath Balekundri.
Rakshita Shivapoji.
Ankita Bhosale.

1 Problem statement

Diabetes mellitus, a chronic metabolic disorder characterized by elevated blood glucose levels, has become a global health concern with a significant impact on individuals, healthcare systems, and economies. Early detection and intervention play a crucial role in managing diabetes and preventing complications.Therefore, there is a pressing need for an accurate and reliable predictive model that can identify individuals at high risk of developing diabetes.

2 Abstract

Diabetes mellitus is one of the most common human diseases worldwide and may cause several health-related complications.It is an illness caused because of high glucose level in a human body.Diabetes should not 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.This paper builds up a model for the prediction of diabetes using machine learning. The supervised machine learning algorithms used for prediction model is logistic regression.

Key Words: Diabetes, Machine Learning, Prediction, Dataset.

3 Introduction

Diabetes is one of the most common human diseases and has become a significant public health concern worldwide. Earlier,diabetes cases were reported primarily among the middle and old population. But in recent years,the young children have also been reported as diabetic. 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. Machine Learning Is a method that is used to train computers or machines explicitly. Various Machine Learning Techniques provide efficient result to collect Knowledge by building various classification and ensemble models from collected dataset. Various techniques of Machine Learning can capable to do prediction, however it's tough to choose best technique. Thus for this purpose we apply popular classification and ensemble methods on dataset for prediction.

4 Literature Survey

1. The main objective of this paper is to develop a system that can perform early prediction of diabetes for a patient with higher accuracy using different machine learning techniques, including Support Vector Machine, Logistic Regression, and Artificial Neural Network. The paper aims to propose an effective technique for earlier detection of the diabetes disease.[1]

2. The main objective of the paper is to design and implement a system for early prediction of diabetes in a patient using various machine learning techniques. The paper explores the use of different classification and ensemble algorithms to predict diabetes and compares their performance to achieve the best accuracy. The proposed system aims to assist healthcare professionals in making early decisions to cure diabetes and save human lives.[2]

3. The aim of this paper titled "Prediction of Diabetes Disease using Machine Learning Model" is to propose a machine learning model that can predict the occurrence of diabetes with high accuracy. The authors have used supervised learning algorithms such as Logistic Regression,Artificial Neural Network, and Decision Tree to create the analytics models for finding whether the patient is diabetic or not. The paper also highlights the importance of early diagnosis and prevention of diabetes-related health hazards.[3]

4. The objective of the paper "Prediction of diabetes disease using machine learning algorithms" is to develop an effective model with high precision to predict diabetes. The study aims to utilize machine learning algorithms,

specifically the K-nearest neighbor algorithm and support vector machine, to create a model that can accurately predict the occurrence of diabetes. The ultimate goal is to contribute to the field of healthcare by providing a reliable method for predicting diabetes using advanced computational techniques.[4]

5 Supervised learning Algorithms

5.1 Logistic Regression

Logistic regression is a supervised machine learning algorithm.LR helps us in solving classification problems, it uses S-curve instead of a straight line for fitting the points. Logistic is taken from the function logit that is used in this method of classification.Main aim of logistic regression is to best fit which is responsible for describing the relationship between target and predictor variable. Logistic regression is a based on Linear regression model. Logistic regression model uses sigmoid function to predict probability of positive and negative class.Sigmoid function

$$P = 1/1 + e - (a + bx) \tag{1}$$

Here P = probability, a and b = parameter of Model.

6 Flow Chart

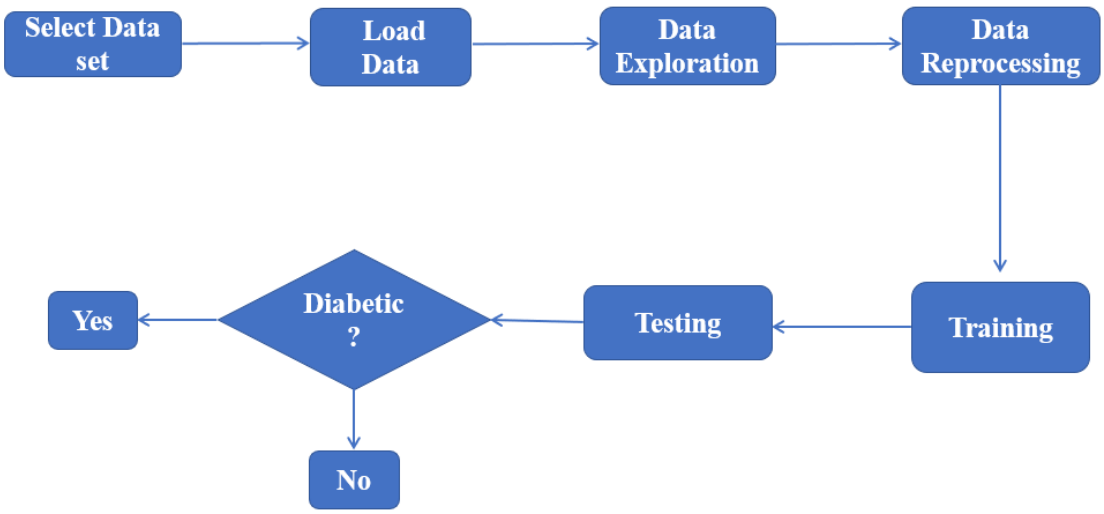

Figure 1: Flow Chart.

7 Data Visualization

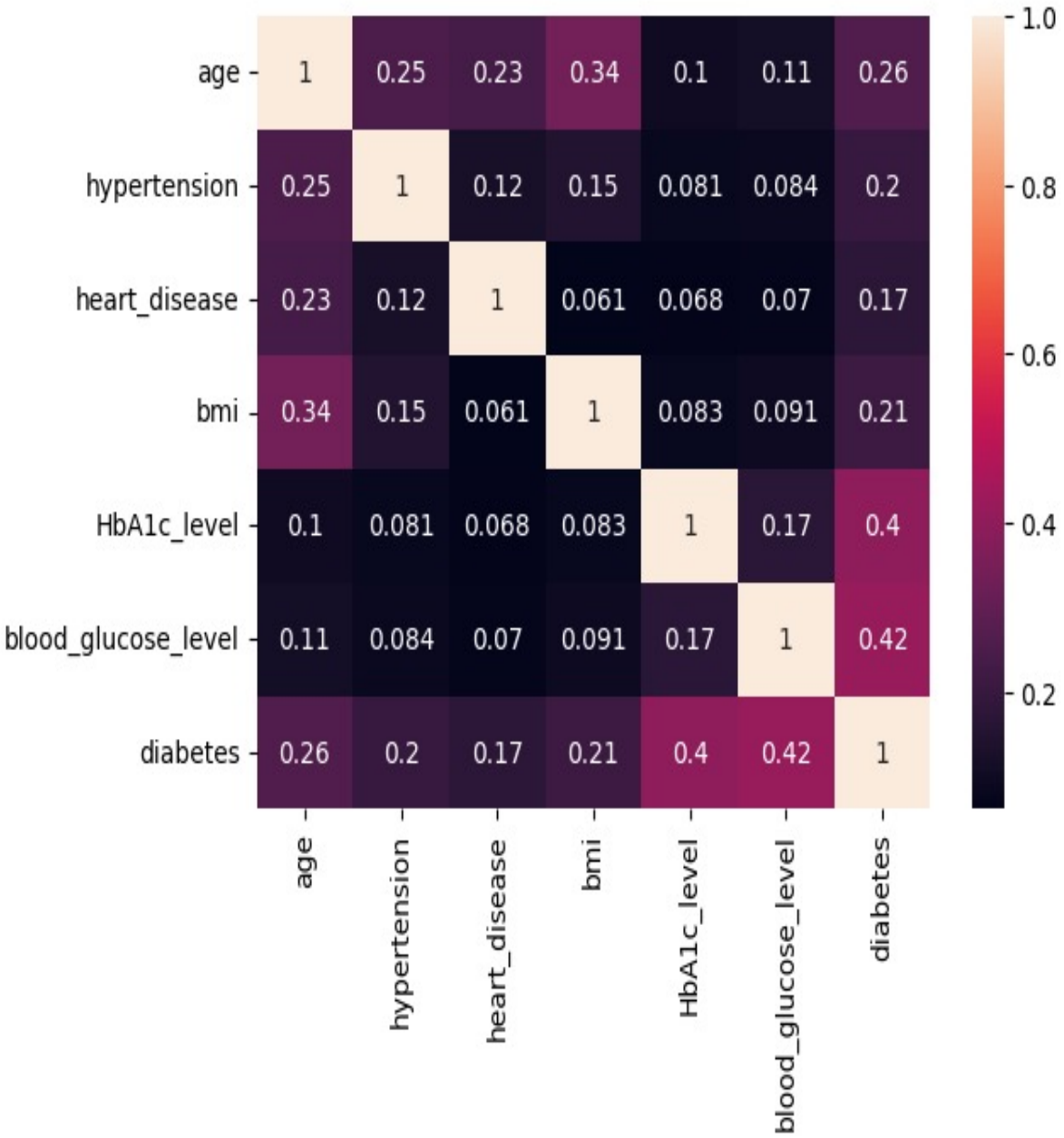

Figure 2: Heat Map of Data.

A Correlation Matrix Heatmap serves as a visual instrument for depicting the correlation coefficients among distinct variables within a dataset. This graphical representation offers an intuitive means to discern relationships and interdependencies between pairs of variables, providing a comprehensive overview of the data's underlying patterns. The correlation matrix heatmap is a crucial analytical tool that contributes to a deeper understanding of the intrinsic connections embedded in the dataset.

7.1 Correlation of Data

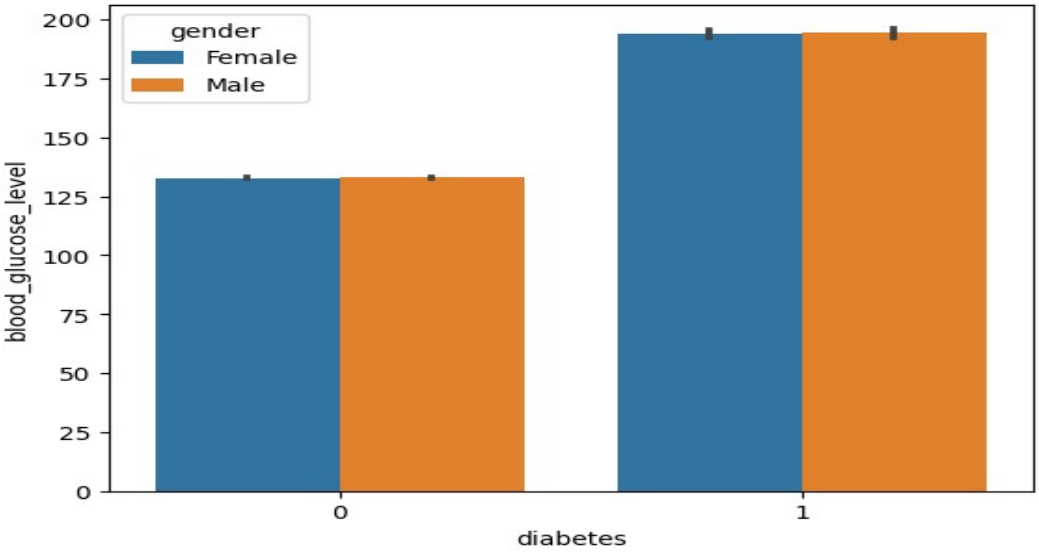

Figure 3: Figure showing how blood glucose level is related to diabetes.

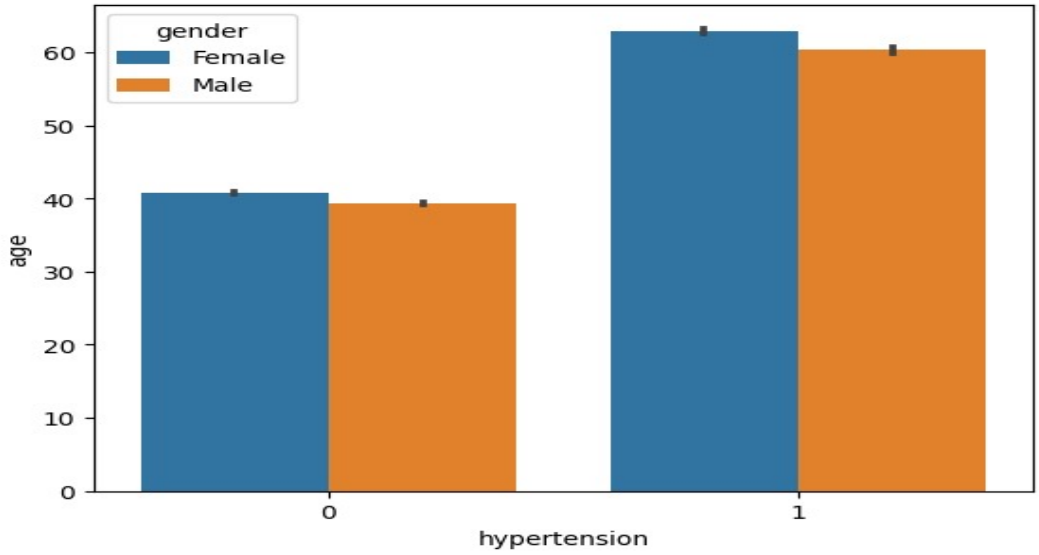

Figure 4: Figure showing how age is related to hypertension.

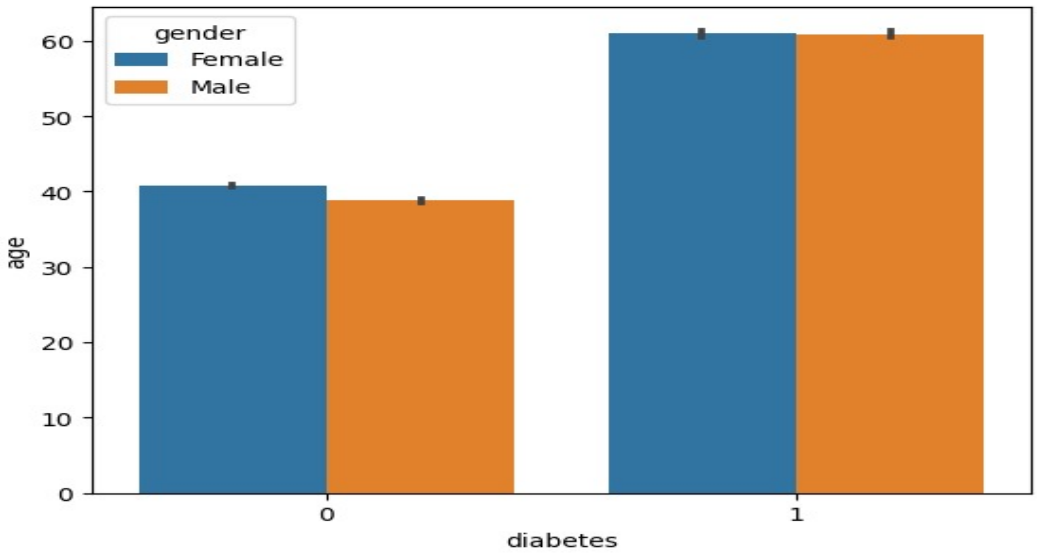

Figure 5: Figure showing how age is related to diabetes

8 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, Logistic Regression and Gradient Boosting classifiers are used. And 77 classification accuracy has been achieved. The Experimental results can be asst health care to take early prediction and make early decision to cure diabetes and save humans life.

References

[1] T. N. Joshi and P. M. Chawan, "Logistic regression and svm based diabetes prediction system," *International Journal For Technological Research In Engineering*, vol. 5, pp. 4347–4350, 2018.

[2] M. Soni and S. Varma, "Diabetes prediction using machine learning techniques," *International Journal of Engineering Research & Technology (Ijert) Volume*, vol. 9, 2020.

[3] A. Sharma, K. Guleria, and N. Goyal, "Prediction of diabetes disease using machine learning model," in *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020*, pp. 683–692, Springer, 2021.

[4] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, "Prediction of diabetes disease using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, p. 284, 2022.