Prediction of Diabetes using Machine Learning

V. Karishma - AP21110010177
K.Sravani - AP21110010160

CONTENTS

ABSTRACT

Diabetes is a prevalent chronic disease affecting a significant portion of the global population. Early detection and accurate prediction of diabetes can play a crucial role in managing the condition and preventing complications. Machine Learning (ML) techniques have shown promising results in diabetes prediction based on patient data. In this study, we propose a user-understandable approach utilising the Random Forest classifier algorithm for accurate and interpretable diabetes prediction. The Random Forest classifier is an ensemble learning method that combines multiple decision trees to generate robust predictions. Its ability to handle high-dimensional datasets and capture complex relationships makes it well-suited for diabetes prediction To build our prediction model, we utilised a comprehensive dataset comprising various patient attributes, including age, body mass index (BMI), blood pressure, glucose levels, and medical history. Pre-processing techniques were applied to handle missing values and normalise the data, followed by feature selection to identify the most relevant attributes for diabetes prediction. The user-understandable representation of the model facilitated effective interpretation and communication of the prediction results. This allows healthcare professionals to explain the prediction rationale to patients, promoting shared decision-making and patient engagement.

INTRODUCTION

In an era where healthcare is increasingly intertwined with advanced technology, our mini project focuses on a vital aspect of public health - **THE PREDICTION OF DIABETES**. Diabetes, a chronic metabolic disorder affecting millions worldwide, demands early detection for effective management. Leveraging the power of data analytics and machine learning, our project aims to develop a predictive model that can identify individuals at risk of diabetes before clinical symptoms manifest. By harnessing a diverse array of health-related data, this endeavour strives to contribute to proactive healthcare strategies and empower individuals to take informed steps towards a healthier future. Join us on this innovative journey as we explore the potential of predictive modelling in revolutionising diabetes prevention and management.

LITERATURE SURVEY

SNO	Authors	Research Paper	Publication	Methodology	Conclusion
1	N.A. Farooqui , Ritika , A. Tyagi	Prediction Model for Diabetes Mellitus Using Machine Learning Techniques	International Journal of Science and Healthcare Research - 2020	Decision Tree, K-Nearest Neighbours	Our model uses different machine learning techniques viz. Decision tree, K-nearest Neighbor's, Random Forest and Support Vector Machine and predict the performance of different classification techniques
2	Dr. Mohammed Abdul Raheem , Shaik Ehetesham , Mohammad Faiz Ahmed Subhani, Sayed Abdul Zakir	AI Algorithm System for Prediction of Diabetes Using Progressive Web Appand IBM Cloud	International Journal of Science and Healthcare Research - 2020	Machine Learning, IBM Cloud, Artificial Intelligence algorithms, Artificial Neural Networks,	Its very important to detect Diabetes in the early stage. Although the accuracy achieved by these Machine Learning models are high, there are a few limitations in this project.

SNO	Authors	Research Paper	Publication	Methodology	Conclusion
3	Rinkal Keniya · Aman Khakharia · Vruddhi Shah · Vrushabh Gada · Ruchi Manjalkar · Tirth Thaker · Mahesh Warang · Ninad Mehendale	Disease prediction from various symptoms using machine learning	SSRN-2020	KNN(k-nearest neighbours)	The Weighted KNN model gave the highest accuracy of 93.5 % for the prediction of diseases using the symptoms.
4	Min Chen, Yixue Hao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang*	Disease Prediction by Machine Learning over Big Data from Healthcare Communities	IEEE -2017	CNN,KNN, Decision tree, NB	Compared to several typical prediction algorithms, the prediction accuracy of our propose algorithm.

PROBLEM STATEMENT IN EXISTING SYSTEM

- In the daily life of a common man prediction of a chronic disease like Diabetes is very crucial and also The prediction of these diseases can be achieved correctly using Machine Learning(ML) models with high accuracy.
- The current app consumes more of the human time for calculated and showing the relevant data from the model, A user Understandable way with less time consuming can be done to get user data predicted in efficient and understandable way .

PROPOSED SYSTEM

- The patient's medical data is taken as the requirements for the model.
- A Huge records of data related to diabetes patients is used to train the model .
- The model RandomForestClassifier is trained with the dataset related to diabetic patients.
- The model is then made to predict the output according to patients data provided.
- Finally a user-friendly and user understandable way the data is presented using the web technologies and framework .

OBJECTIVES:

- The objective is to predict the disease of Diabetes from the data provided by the user using a proper Machine Learning Model(ML).
- Providing Understandable and accurate prediction to user.

Project Domain

❖Using Random Forest Classifier in Machine Learning Model

Requirement analysis

Functional Requirements

- The user is required to provide the medical details regarding which the model can predict the output as confirmation for Diabetes disease

Non-Functional Requirements

- Understandable output
- Available
- User friendly
- Platform Independent

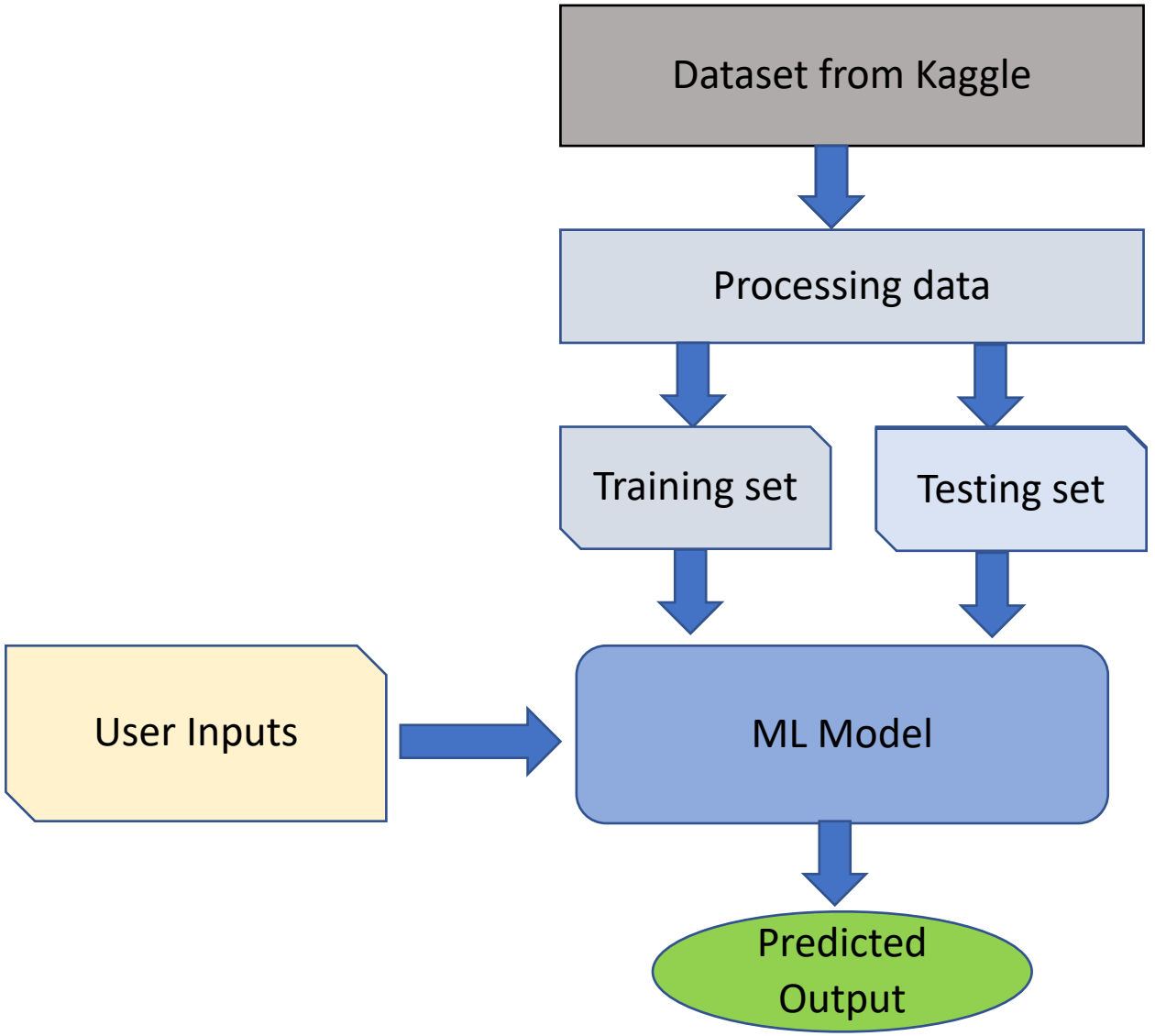
Software Requirements

- Language : Python version - 3 above
- Technologies: Web ,Scripting and Machine Learning
- Operating System – Windows 7 and above

Hardware Requirements

- Processor Support : i3 onwards
- RAM : Minimum 4GB
- Hard Disk : Minimum 2GB

System Architecture

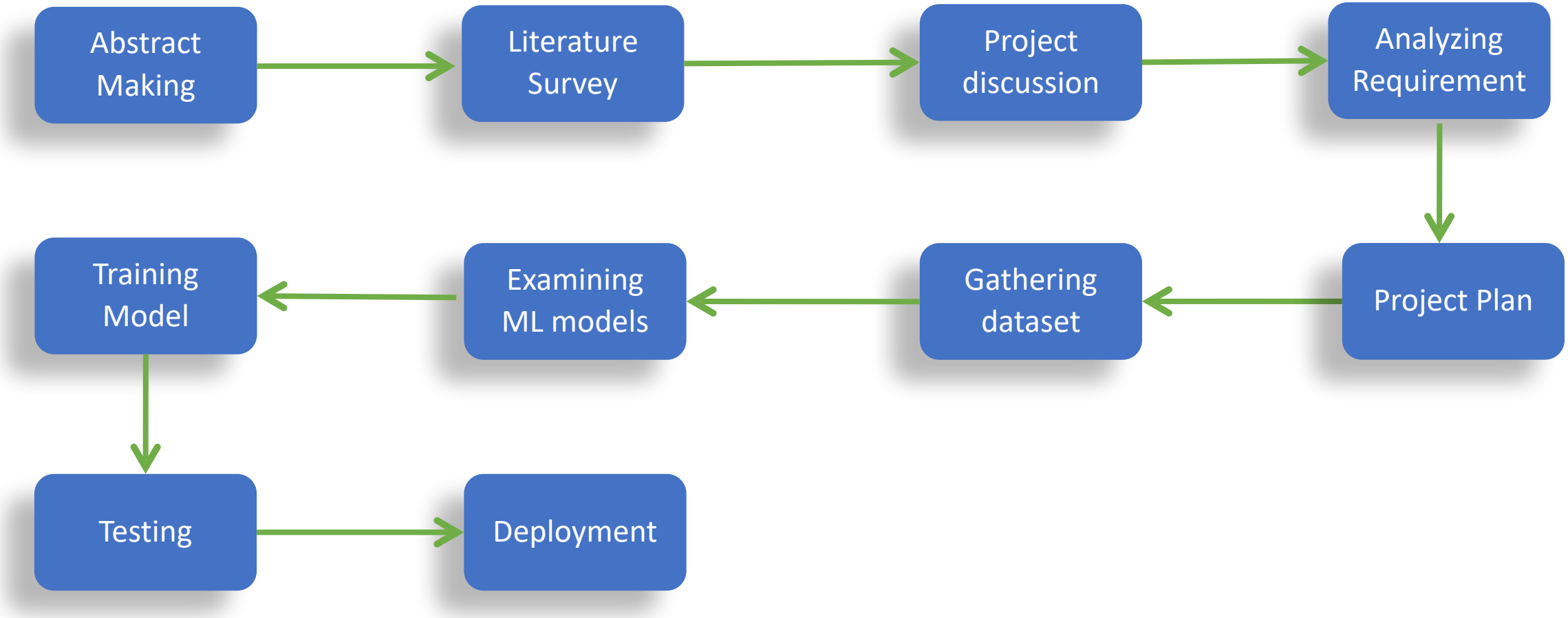


Methodology

Random Forest Classifier :

- In Random forest n number of random records are taken from the data set having k number of records.
- Individual decision trees are constructed for each sample.
- Each decision tree will generate an output.
- Final output is considered based on **Averaging** for Classification and regression respectively.

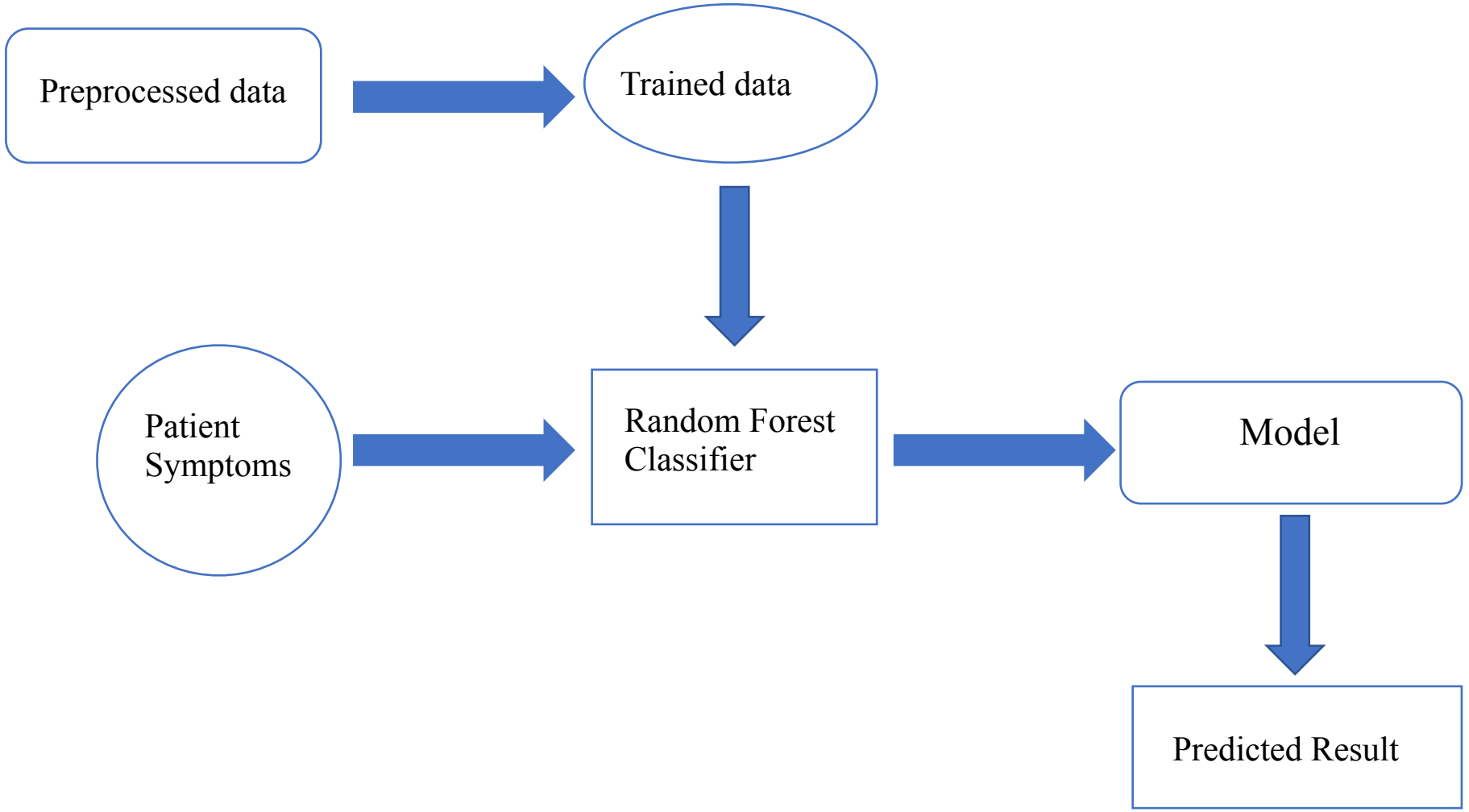
Project plan



CONTENTS

- Abstract
- Design
- Implementation
- Result Analysis
- Attainment of objective
- Conclusion
- Future scope of Project

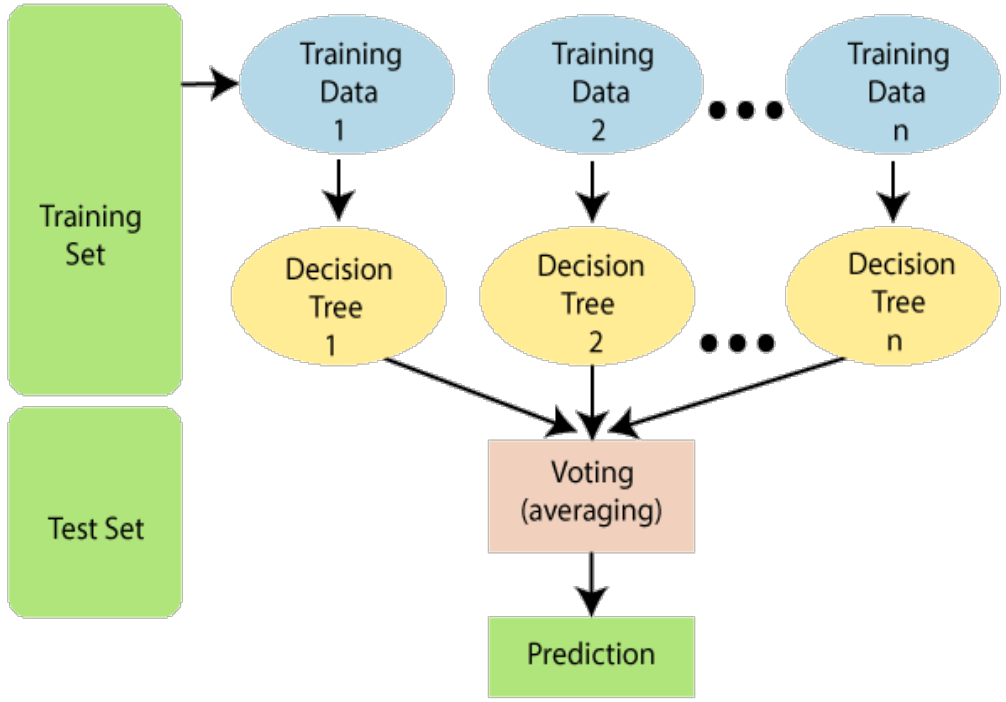
Design


Preprocessed data
Trained data
Patient Symptoms
Random Forest Classifier
Model
Predicted Result

Implementation

Random Forest Classifier :

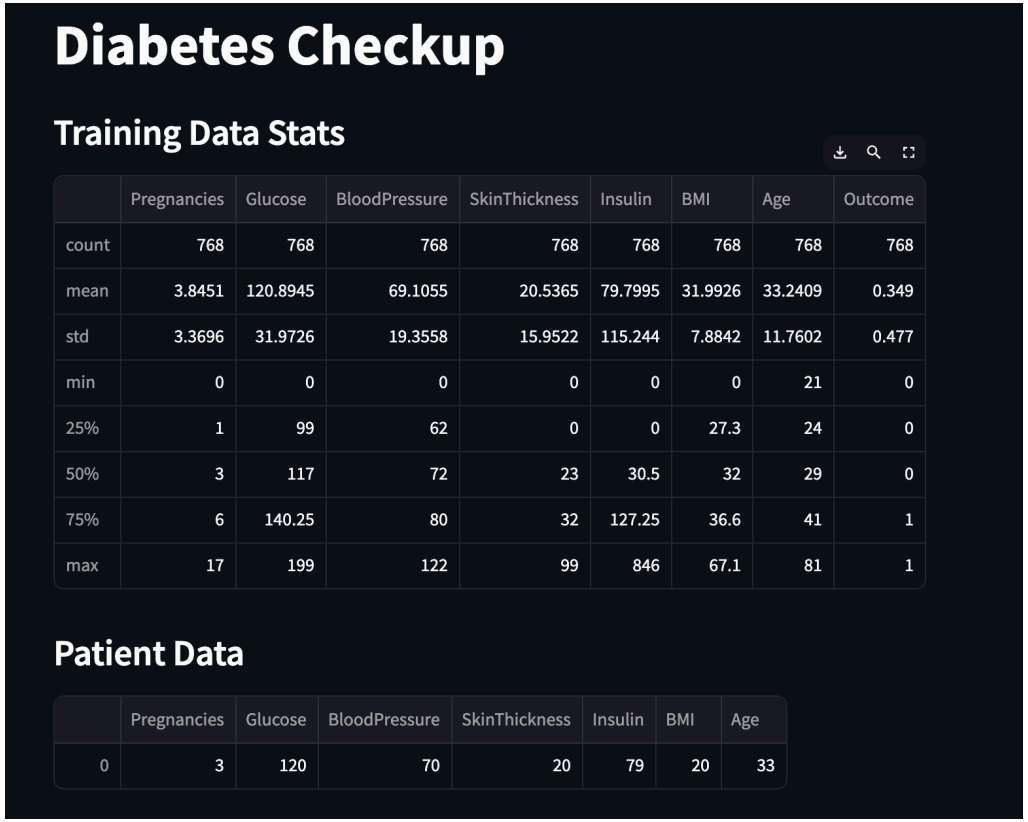
- It is a variously levelled tree-like design that settles on choices by recursively dividing the information into subsets in view of the values of input features.
- The construction of the tree comprises nodes, branches, and leaves. Nodes address choices, branches address the potential results of decisions, and leaves address the last predictions.
- Decision trees make parts in view of component values to make nodes. The objective is to make splits that best separate the data into classes for classification. The splitting process continues until stopping criteria are met. For classification, a leaf node typically represents a class label. The final prediction in a random forest is obtained by aggregating the predictions of all individual decision trees; this can be done by majority voting.



Modules Used

- pandas
- streamlit
- PIL
- numpy
- matplotlib.pyplot
- sklearn
- seaborn

Result Analysis



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Age	Outcome
count	768	768	768	768	768	768	768	768
mean	3.8451	120.8945	69.1055	20.5365	79.7995	31.9926	33.2409	0.349
std	3.3696	31.9726	19.3558	15.9522	115.244	7.8842	11.7602	0.477
min	0	0	0	0	0	0	21	0
25%	1	99	62	0	0	27.3	24	0
50%	3	117	72	23	30.5	32	29	0
75%	6	140.25	80	32	127.25	36.6	41	1
max	17	199	122	99	846	67.1	81	1

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Age
0	3	120	70	20	79	20	33

Fig-1: Interface of the proposed model

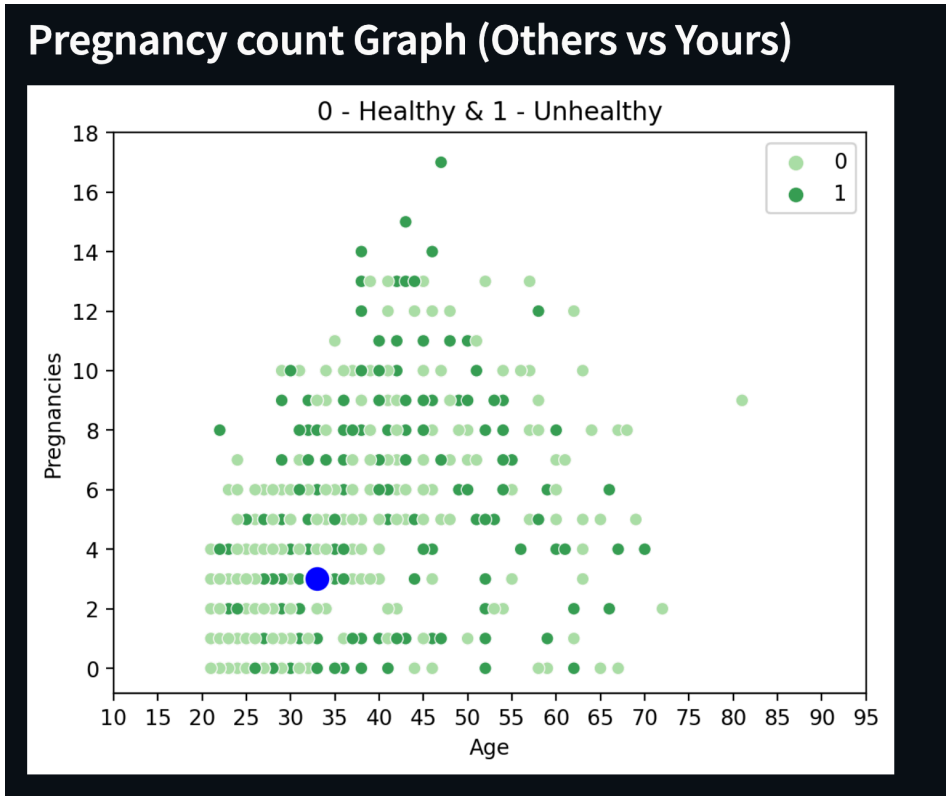


Fig-2: Visualization of Pregnancies

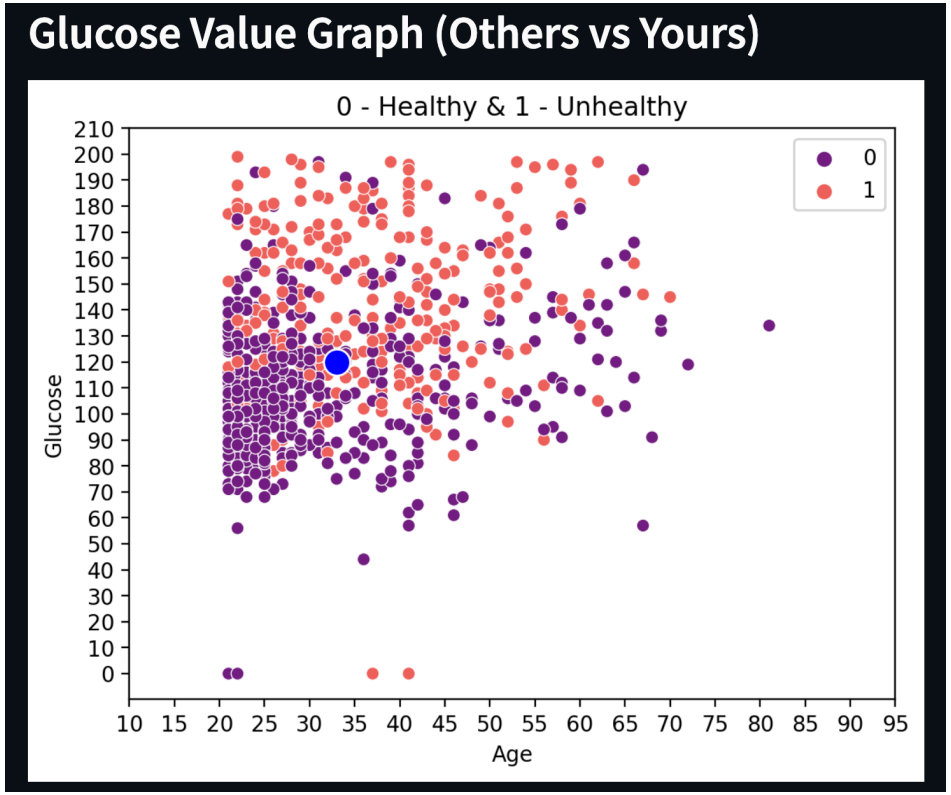


Fig-3: Visualization of Glucose

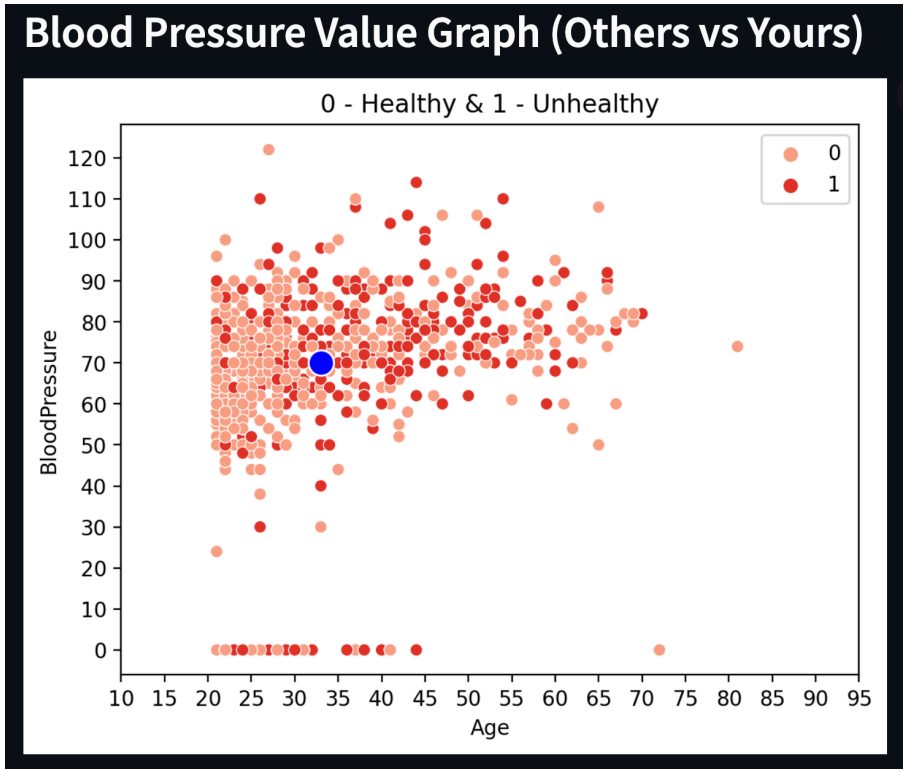


Fig-4:Visualization of Blood Pressure

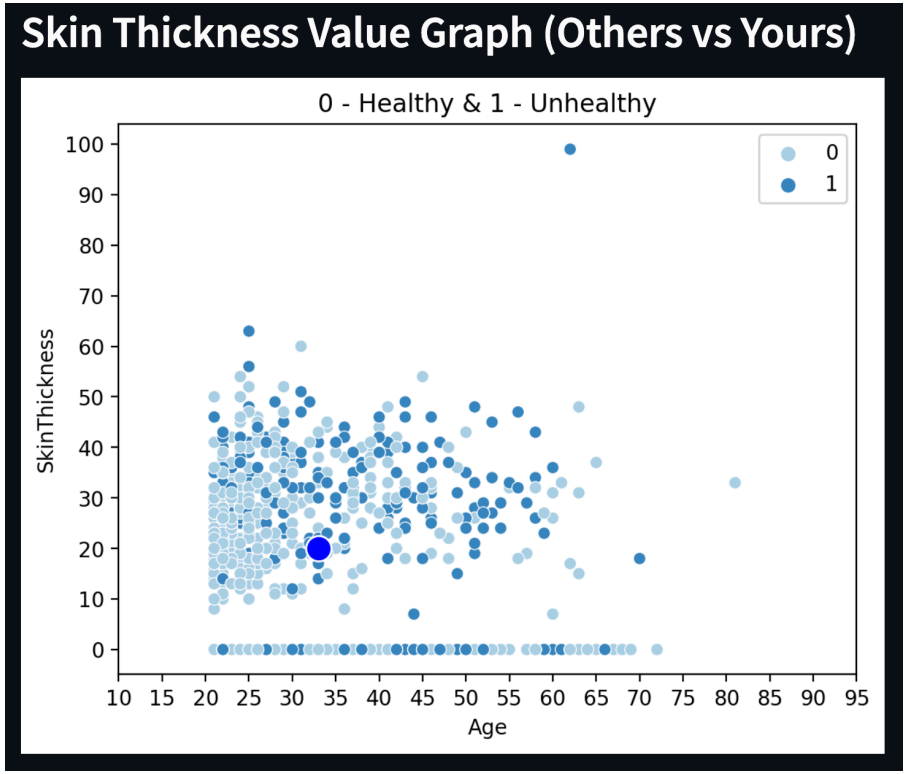


Fig-5:Visualization of Skin Thickness

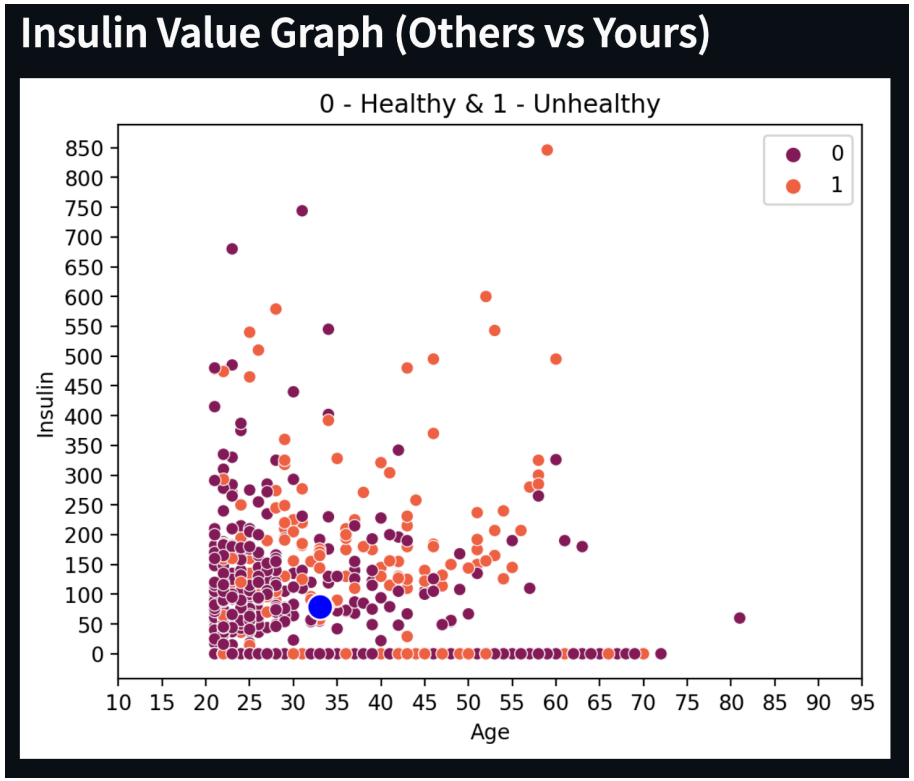


Fig-6: Visualization of Insulin

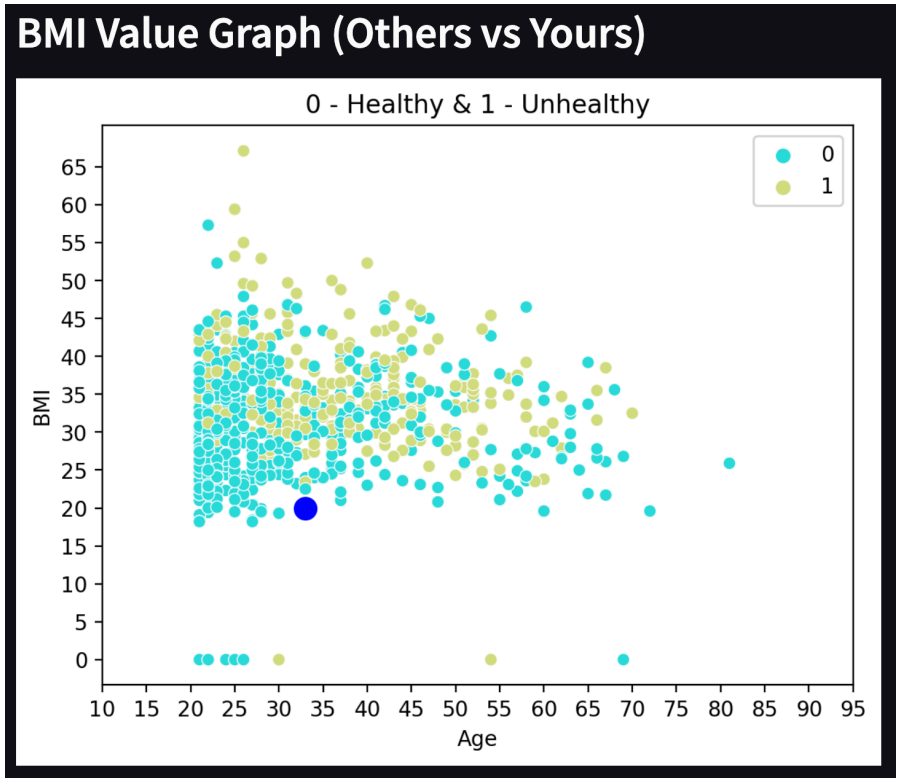


Fig-7: Visualization of BMI

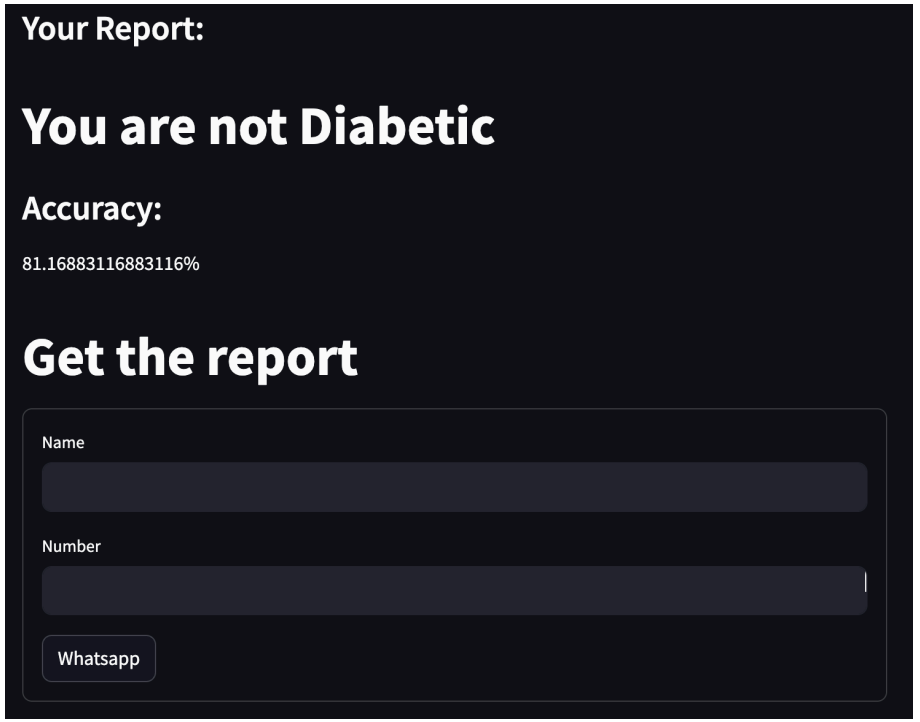


Fig-8:Reporting WhatsApp Message

Attainment of objective

- The objective is to predict whether an user is diabetic or not based on the information given. So we designed a model using random forest classifier which predicts the results with high accuracy.
- The model is then made to anticipate the result as per the patient's information by comparing. Finally, a WhatsApp message is sent to the given portable number that says a client is diabetic or not.
- As per our objective, we constructed a user-friendly model in an understandable way.

Conclusion

- All in all, the execution of the present-day diabetic forecast model shows a more prominent effect on humans' existence. In basic words, client inputs are contrasted with different information, which envisions a report with additional accuracy, and a WhatsApp message is sent saying a client is diabetic or not.In this quick world, foreseeing diabetes at the beginning is exceptionally pivotal. The current models are not less tedious, easy to use, and client-justifiable. So, our proposed model beats these impediments, i.e., it is less tedious, easy to understand, and gives highly precise outcomes.

Future Scope of Project

- Early identification is essential for optimal care of diabetes, a chronic metabolic condition that affects millions of people worldwide. Our study aims to create a prediction model that can identify people at risk of diabetes before clinical signs appear by leveraging the power of machine learning.
- **Improved Prediction Accuracy**: One of the main advantages of using machine learning algorithms like Random Forest is their ability to improve prediction accuracy over time.
- **Personalised Healthcare**: As the model becomes more accurate, it can provide personalised health recommendations based on an individual's risk factors and historical data.

THANK YOU