Mahavir Education Trust's
SHAH AND ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Computer Engineering

DIABETES RISK PREDICTOR

GLUCOSAGE

submitted in partial fulfillment of the requirements of the degree of bachelor in engineering by

NAME	CLASS	ROLL NO.
Dhwani Darji	TE-9	46
Anuja Agare	TE-4	1
Aniket Jadhao	TE-4	20
Ojas Hedau	TE-4	19

Under the Guidance of
Ms. Pallavi Khodke

INTRODUCTION

In an era defined by technological innovation and data-driven healthcare solutions, the quest to tackle the global burden of diabetes has reached a pivotal juncture. Diabetes, a chronic metabolic disorder characterized by elevated blood sugar levels, exerts a profound impact on individual health outcomes and healthcare systems worldwide.

Within this landscape of evolving challenges and opportunities, we proudly present our Diabetes Risk Predictor—a transformative tool poised to revolutionize the approach to diabetes prevention and management.

PROBLEM STATEMENT

The increasing prevalence of Type 2 diabetes, poses a significant health challenge globally. To address this issue, there is a critical need for proactive interventions and personalized risk assessment strategies. The development of Diabetes Risk Predictor-GlucoSage aims to leverage advanced data analytics and machine learning techniques to accurately identify individuals at high risk of developing diabetes.

OBJECTIVES

Early Identification

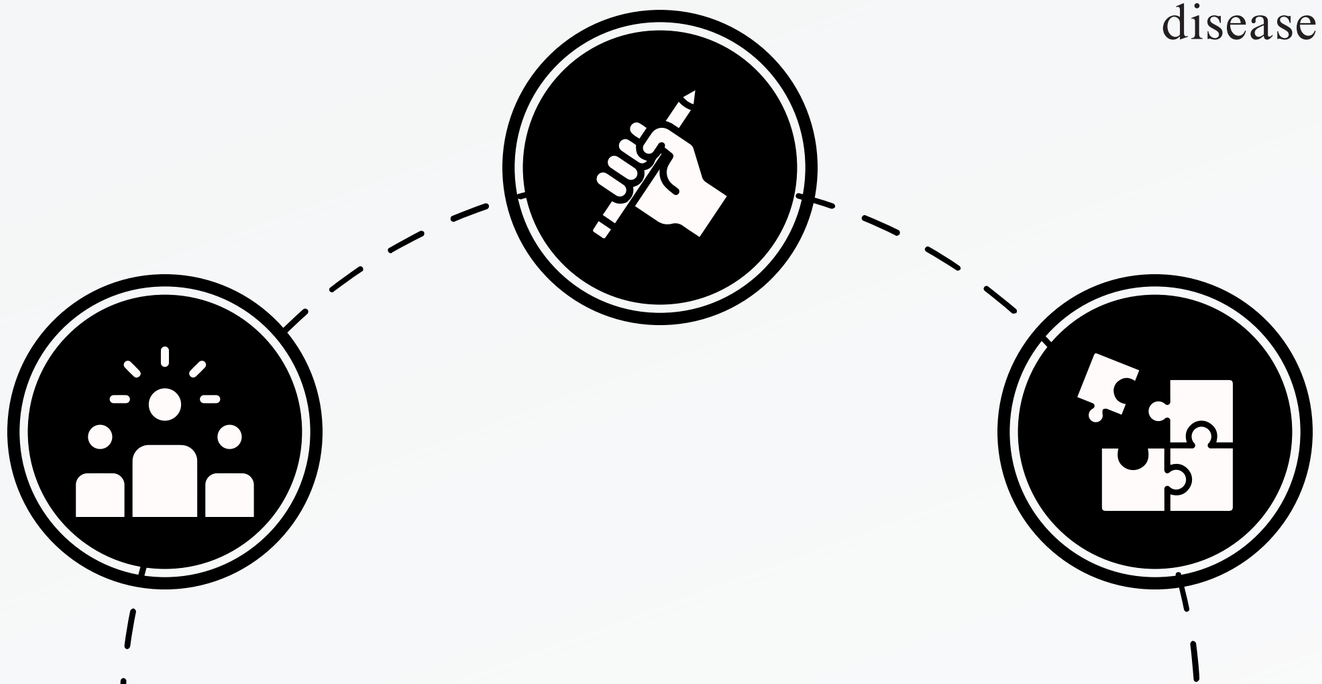
Develop a predictive model that precisely identifies individuals at risk of developing Type 2 diabetes

Personalized Risk Assessment

Provide users the foresight necessary to intervene early and effectively mitigate the onset of diabetes and its associated complications.

Preventive Interventions

Facilitate targeted preventive interventions by accurately assessing an individual's risk of developing diabetes, thereby reducing the incidence of the disease.



PURPOSE

Our motive is to provide personalized risk assessments based on scientific data, empowering users to make informed decisions about their lifestyle and healthcare choices.

EMPOWERMENT THROUGH KNOWLEDGE

We aim to offer convenient and accessible alternative, unlike traditional process GlocoSage allows users to assess their risk from the comfort of their own home, at any time that suits them.

CONVENIENCE AND ACCESSIBILITY

LITERATURE REVIEW

Author(s)	Year	Title	Key Findings
Farida Mohsen, Hamada R. H. Al-Absi, & Zubair Shah	2023	A scoping review of artificial intelligence-based methods for diabetes risk prediction	• Promising Performance of AI Models • Variety of Data Modalities and Techniques
Apratim Sadhu & Abhimanyu Jadli	2021	Early-Stage Diabetes Risk Prediction: A Comparative Analysis of Classification Algorithms	• Analysis of various classification algorithms and their accuracy score • Obtained the required dataset and necessary information about it
Gary S Collins, Susan Mallett, Omar and Ly-Mee Yu	2011	Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting	• Impact of inadequate Methods Usage • Impact on Reliability

METHODOLOGY

The Dataset

The dataset that we have used is taken from the UCI repository. It contains 520 instances and 16 attributes with a few missing values which have been pre-processed by ignoring the tuples with incomplete values. The dataset is build up of following attributes:
Age, Sex, Polyuria, Polydipsia, Sudden Weight loss, Weakness, Polyphagia, Genital thrush, Visual, blurring, Itching, Irritability, Delayed Healing, Partial Paresis, Muscle Stiffness, Alopecia, Obesity, Class.

After pre-processing the dataset a total of 520 instances remained. Out of these 520 instances, 320 are positive and 200 are negative values The two class variables(positive or negative) are used to find whether a patient has a risk of diabetes or not.

Overview of Machine Learning Models Tested:

By evaluating various research papers during the literature survey, the machine learning model that best suited our problem statement are as mentioned below:
Support Vector Machine (SVM) and Random Forest Classifier.

Support Vector Machine (SVM)

- SVM is a supervised learning algorithm used for classification and regression tasks.
- It works by finding the hyperplane that best separates the classes in the feature space, maximizing the margin between the classes.

Random Forest Classifier

- A random forest is essentially an ensemble of a number of decision trees and outputting the mode of the classes as the prediction.
- we have used max-depth of 13 and 100 estimators to classify the data points.

Analysis and comparison

Aspect	Random Forest Classifier	Support Vector Machine
Model Complexity	Ensemble method with multiple decision trees	Uses hyperplanes to separate data points
Performance	Robust and resistant to overfitting	Powerful with strong theoretical guarantees
Interpretability	Provides feature importances	Produces less interpretable models
Handling Imbalanced Data	Naturally robust to class imbalance	May require additional techniques for imbalanced data

Results And Observations

Model	Accuracy	F-Score	ROC
Support Vector Machine	94.4715%	0.9529	0.9866
Random Forests	98.0778%	0.9790	0.9979

Hence, as per the results obtained we can firmly believe that Random Forest Classifier is one of the most effective algorithms against binary-based classification datasets.

Description of Frontend-Backend Project Setup:

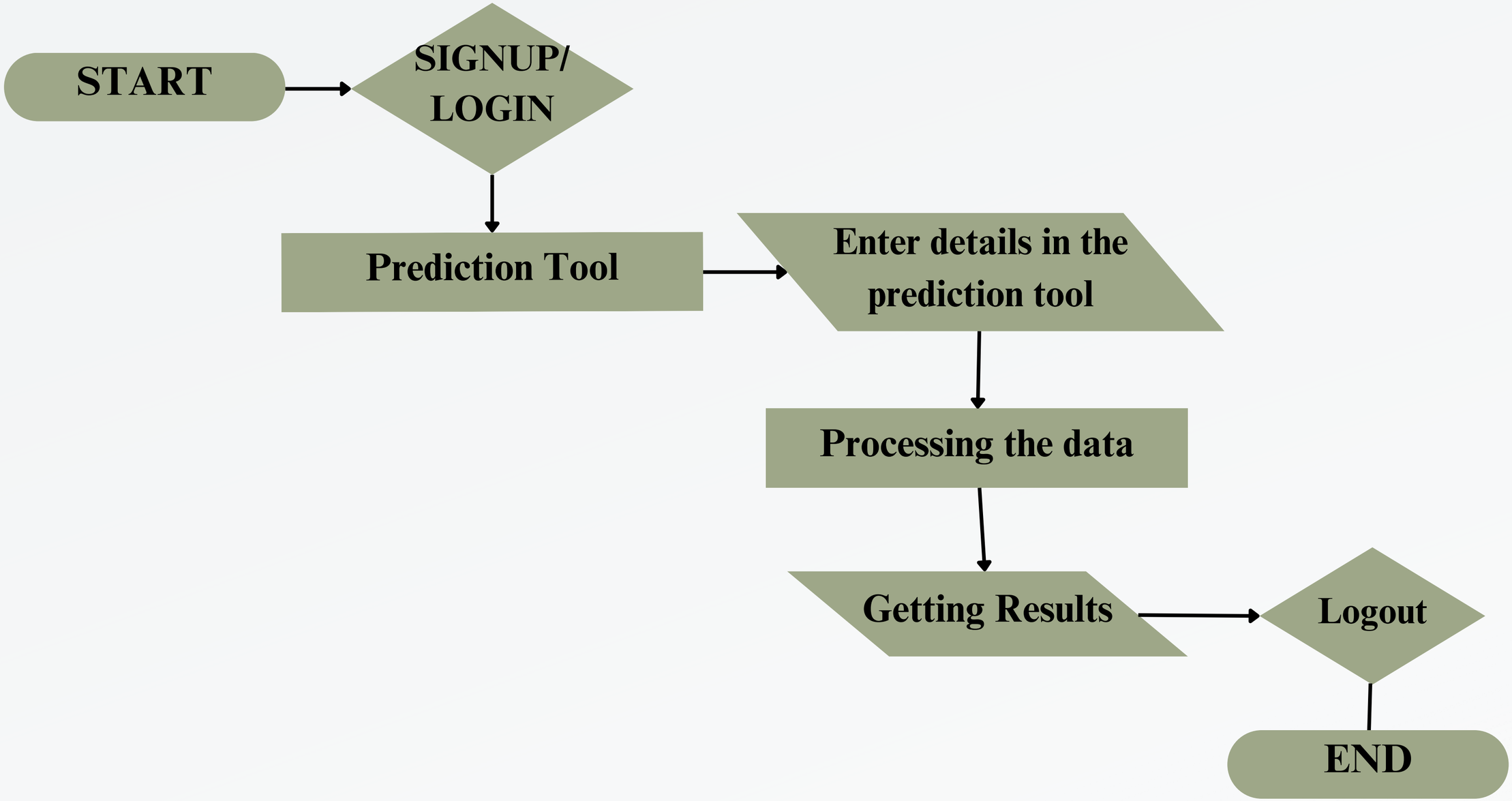
Front end: HTML, CSS, Javascript, Bootstrap

Back end: Node.js, Express.js

Database: MongoDB

Model language: Python

FLOW CHART


START
SIGNUP/
LOGIN
Prediction Tool
Enter details in the
prediction tool
Processing the data
Getting Results
Logout
END

CONCLUSION

Diabetes can be a reason for reducing life expectancy and quality. Predicting this chronic disorder earlier can reduce the risk and complications of many diseases in the long run. In this project, an automatic diabetes prediction system using various machine learning approaches has been proposed.

FUTURE SCOPE

In the future, the proposed trained model will be used to build a web app with a user-friendly interface. Additionally, the proposed framework will be applied to other medical contexts to verify their generality and versatility to predict the disease classes.

REFERENCES

[1] Apratim Sadhu, Abhimanyu Jadli ‘Early stage diabetes risk prediction: A Comparative Analysis of Classification Algorithms’ In 2021 International Advanced Research Journal in Science, Engineering and Technology

[2] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, Konstantinos Moustakas ‘Machine Learning tools for Long Term Type 2 Diabetes Risk Prediction’ , July 2021

[3] Mller, Andreas C., and Sarah Guido. Introduction to Machine Learning with Python: a Guide for Data Scientists. O’Reilly, 2017.

[4] https://github.com/sherwinvishesh/Diabetes-Prediction/blob/main/

[5] https://chat.openai.com/

THANK YOU