



# DEVL MINI PROJECT

Department of Computer Engineering  
Second Year - A

By-

**Gaurav Jaypatre - 124B1B025**

**Pranav Chaudhari - 124B1B027**

**Tanmay Talreja - 124B1B032**

**Digvijay Birajdar - 124B1B013**



# PROBLEM STATEMENT +

Diabetes is one of the most common chronic diseases affecting millions of people worldwide. Early detection of diabetes can help prevent severe complications through timely treatment and lifestyle changes.



However, traditional medical tests are time-consuming and expensive. So, our goal is to develop a machine learning model that can predict whether a person is likely to have diabetes based on health parameters such as Glucose, BMI, Age, Blood Pressure, etc.

By analyzing these parameters, the system can help doctors and individuals make data-driven health decisions.

# OBJECTIVE

Apply data preprocessing, feature engineering, and machine learning to achieve accurate predictions.

Build an interpretable and reliable model for healthcare insights.

To predict the likelihood of diabetes in patients using medical data.



# IMPLEMENTATION



01

## Data Loading

- Imported the Diabetes Dataset using Python (Pandas).
- Explored the dataset to understand structure and key attributes.
- Identified target variable: Outcome (1 = Diabetic, 0 = Non-Diabetic).

02

## Data Preprocessing

- Detected missing or invalid values (e.g., 0 in Glucose, BMI, Insulin).
- Replaced unrealistic zeros with median values for accuracy.
- Ensured all data was clean, consistent, and medically valid.

03

## Exploratory Data Analysis (EDA)

- Visualized data using histograms, count plots, and boxplots.
- Checked for imbalanced classes (more non-diabetic cases).
- Used a correlation heatmap to find relationships between features.

04

## Feature Engineering

- Applied log transformation on Insulin to reduce data skewness.
- Performed feature importance analysis using Random Forests.
- Selected the most influential features for model training.



# IMPLEMENTATION



05

## Model Building

- Used Random Forest Classifier for classification.
- Balanced the dataset using SMOTE to handle class imbalance.
- Applied feature scaling for better accuracy and convergence.
- Tuned parameters for optimal performance.

06

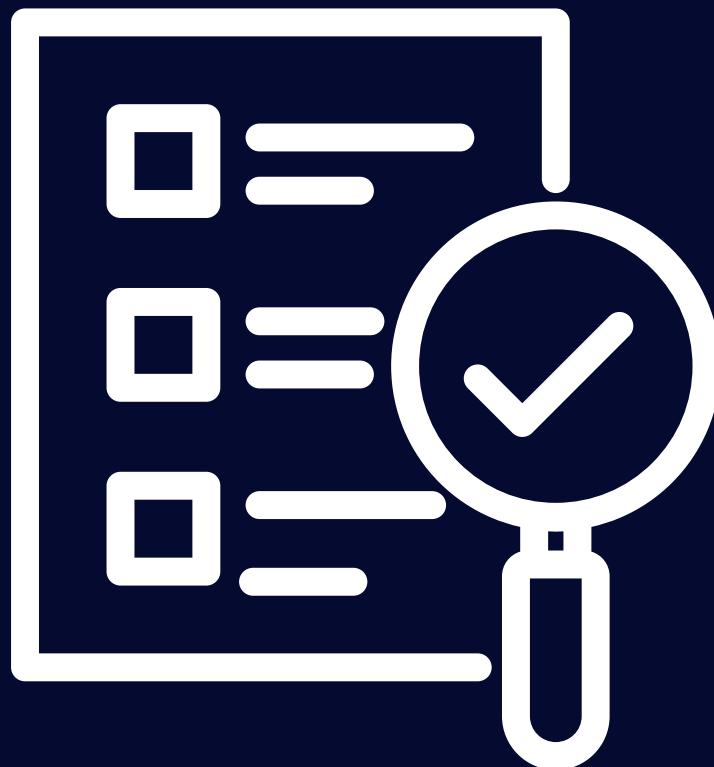
## Model Evaluation

- Evaluated using accuracy, confusion matrix, and precision-recall metrics.
- Achieved ~90–92% accuracy after scaling and SMOTE balancing.
- Model showed strong precision and recall on test data.

07

## User Prediction System

- Built an interactive interface for users to input health parameters.
- Model predicts diabetes likelihood instantly.
- Displays clear result: “Likely Diabetic” or “Not Likely Diabetic.”



# RESULTS & INSIGHTS



Number of Pregnancies: 1  
Glucose Level: 95  
Blood Pressure: 70  
Skin Thickness: 25  
Insulin Level: 80  
BMI: 26.5  
Diabetes Pedigree Function: 0.3  
Age: 29

--- Diabetes Prediction ---

Enter the following health details:  
Number of Pregnancies: 1  
Glucose Level: 95  
Blood Pressure: 70  
Skin Thickness: 25  
Insulin Level: 80  
BMI: 26.5  
Diabetes Pedigree Function: 0.3  
Age: 29

Prediction: NO, the person is not likely to have Diabetes.

Final Model Accuracy: 80.5 %

--- Project Completed Successfully ---

- Accuracy : 80.5%
- Top Influential Features: Glucose, BMI, Age, Insulin
- Model Strength: High interpretability and robustness.

Number of Pregnancies: 5  
Glucose Level: 165  
Blood Pressure: 88  
Skin Thickness: 32  
Insulin Level: 180  
BMI: 35.4  
Diabetes Pedigree Function: 0.6  
Age: 45

--- Diabetes Prediction ---

Enter the following health details:  
Number of Pregnancies: 5  
Glucose Level: 165  
Blood Pressure: 88  
Skin Thickness: 32  
Insulin Level: 180  
BMI: 35.4  
Diabetes Pedigree Function: 0.6  
Age: 45

Prediction: YES, the person is likely to have Diabetes.

Final Model Accuracy: 80.5 %

--- Project Completed Successfully ---



THANK YOU

