Infosys Springboard

Artificial Intelligence Internship 5.0

GLUCOSENSE: AI-POWERED

DIABETES DETECTION FOR EARLY INTERVENTION

Transforming Healthcare with AI-Driven Diagnosis

AGENDA

1. Introduction to our project - GlucoSense
2. Dataset Overview
3. Exploratory Data Analysis (EDA)
4. Model Building and Evaluation
5. Deployment Process & Tech Stack
6. Results and Future Scope

INTRODUCTION TO OUR PROJECT:

GlucoSense

GlucoSense

An AI-powered system that analyzes patient health data to detect early signs of diabetes, enabling timely intervention & personalized healthcare.

OBJECTIVE:

Our AI-driven model detects early signs of diabetes with high accuracy, enabling timely medical intervention and improving patient outcomes through advanced machine learning.

APPROACH:

We analyze patient health data to identify patterns & risk factors, leveraging AI for accurate detection and personalized healthcare insights.

IMPACT:

Reliable medical data powers our model, ensuring precise predictions that support healthcare providers in making informed, proactive decisions.

CHALLENGES IN DIABETES DETECTION & AI-POWERED SOLUTIONS:

Diabetes detection faces challenges like late diagnosis, inconsistent monitoring, and data overload. AI-powered solutions enhance early prediction and personalized treatment for better patient outcomes.

GLUCOSENSE: AI-POWERED DIABETES DETECTION SYSTEM

Problem Statement

Traditional diabetes diagnosis relies on delayed detection, leading to late interventions and increased health risks.

Solution

Our AI-powered model analyzes patient health data to detect early signs of diabetes, enabling timely and accurate diagnosis for proactive healthcare.

Advantages

- **Early Detection**: Reduces complications through timely intervention.
- **Improved Accuracy**: AI-driven insights enhance diagnostic precision.

PROJECT WORKFLOW

1 DATA COLLECTION

Gathering diverse diabetes-related datasets to analyze health indicators and risk factors.

2 DATA PREPROCESSING

Cleaning, transforming, and handling imbalanced data for better model accuracy.

3 EXPLORATORY DATA ANALYSIS (EDA)

Identifying patterns, correlations, and trends in diabetes prediction.

4 MODEL BUILDING

Developing AI-driven models to predict diabetes risk with high precision.

5 MODEL EVALUATION & DEPLOYMENT

Validating model performance and integrating it for real-world application.

DATASET OVERVIEW: AN OVERVIEW OF ALL DATASETS

Feature	Pima Indians Diabetes	Diabetes Health Indicators	Early-Stage Diabetes Risk	Diabetes Prediction
Size	768 records	253,680 records	520 records	100,000+ records
Feature Count	8	22	16	9
Data Type	Mostly numerical	Mixed	Mostly categorical	Mixed
Source	UCI Repo	Survey	Survey	Medical Records
Class Imbalance	Balanced	Imbalanced	Balanced	Balanced
Preprocessing Required	Minimal	Heavy	Minimal	Moderate

Reasons for Choosing the Diabetes Prediction Dataset

LARGE SAMPLE SIZE

It provides over 100,000 records, ensuring better generalization for ML models.

DIVERSE FEATURES

Includes essential health indicators like HbA1c, glucose, and BMI while keeping complexity manageable.

BETTER GENERALIZATION

Its size, feature diversity, and distribution make it ideal for building robust ML models for diabetes prediction.

MEDICAL RECORDS-BASED

More reliable compared to survey-based datasets (e.g., Early-Stage Diabetes Risk dataset).

MODERATE PREPROCESSING

Requires some cleaning but is not overly complex compared to datasets like the Diabetes Health Indicators Dataset

DATA COLLECTION & PREPROCESSING

1 HANDLING MISSING VALUES:

Applied imputation techniques like mean, median, and mode to fill missing values, ensuring data completeness for model training.

2 ENCODING CATEGORICAL VARIABLES:

Categorical features were converted into numerical representations using one-hot encoding to enhance model compatibility.

3 FEATURE ENGINEERING:

Standard scaling was applied to Glucose_Level and Insulin_Level, while min-max scaling was used for BMI and Age. This ensured uniformity across features and improved model performance.

4 OUTLIER DETECTION & TREATMENT:

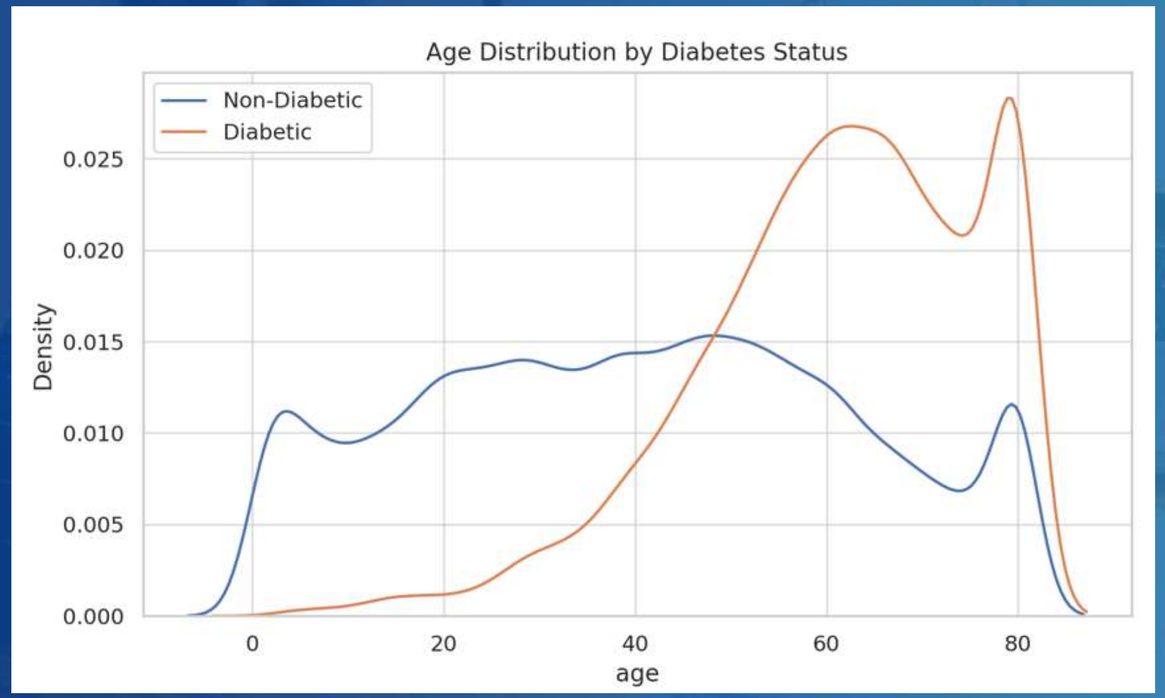
Outliers in BMI were handled using the IQR method. This prevented extreme values from distorting model performance.

EXPLORATORY DATA ANALYSIS (EDA):

1. Data Quality and Integrity Check
2. Univariate Analysis
3. Bivariate and Multivariate Analysis
4. Comorbidity Analysis
5. Gender and Health Outcome Disparities
6. Anomaly Detection and Risk Stratification
7. Feature Engineering Opportunities

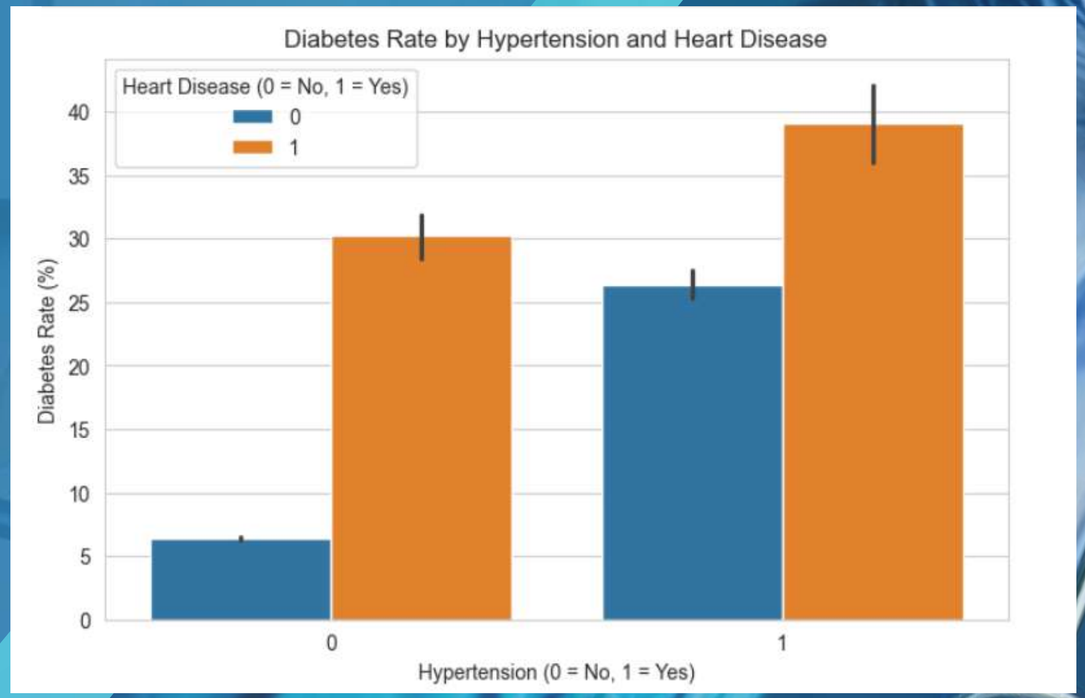
EXPLORATORY DATA ANALYSIS (EDA)

AGE DISTRIBUTION BY DIABETES



- This density plot shows the age distribution of diabetic and non-diabetic individuals.
- Diabetes cases rise with age, peaking around 60-70 years.
- Younger individuals are mostly non-diabetic, while diabetes becomes more common after 40 years.

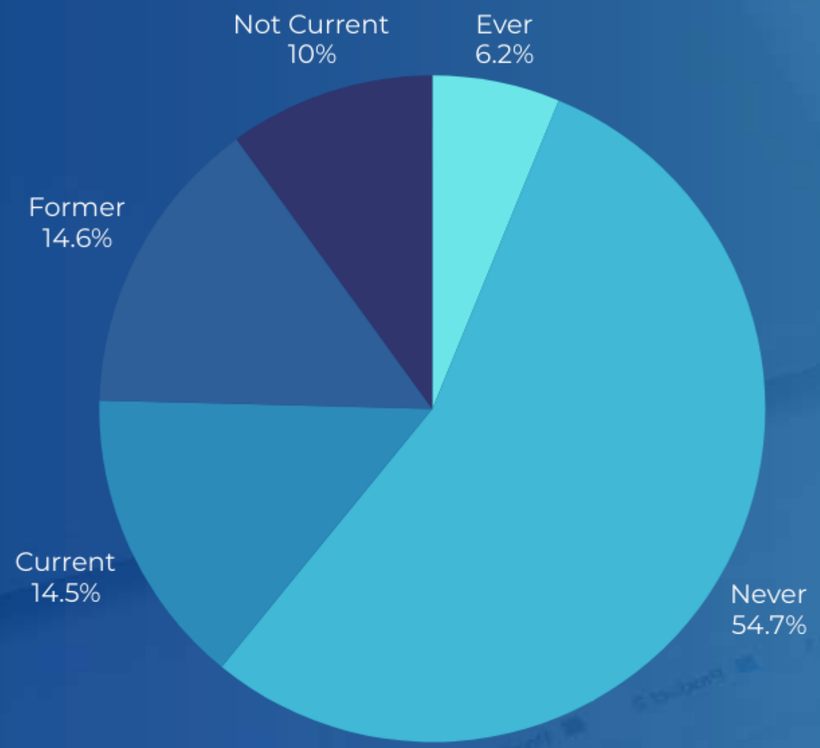
HYPERTENSION & HEART DISEASE



- This bar chart shows the diabetes rate based on hypertension and heart disease.
- Diabetes is lowest in those without either condition.
- The rate increases with hypertension or heart disease and is highest when both are present.

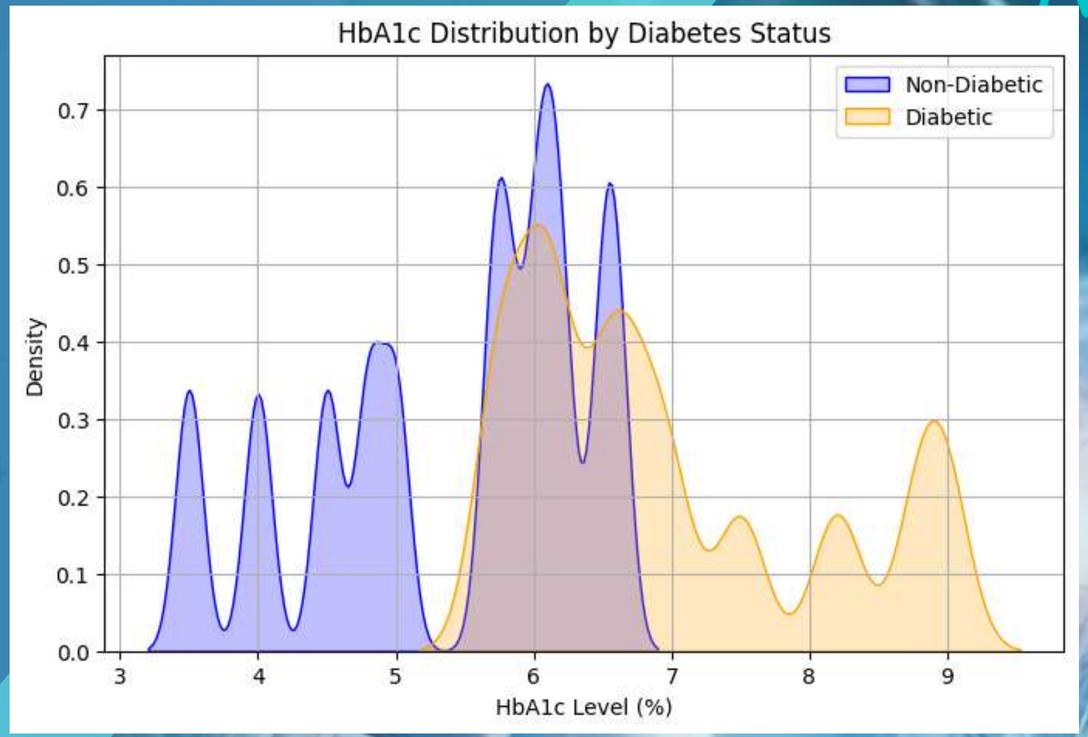
EXPLORATORY DATA ANALYSIS (EDA)

SMOKING HISTORY



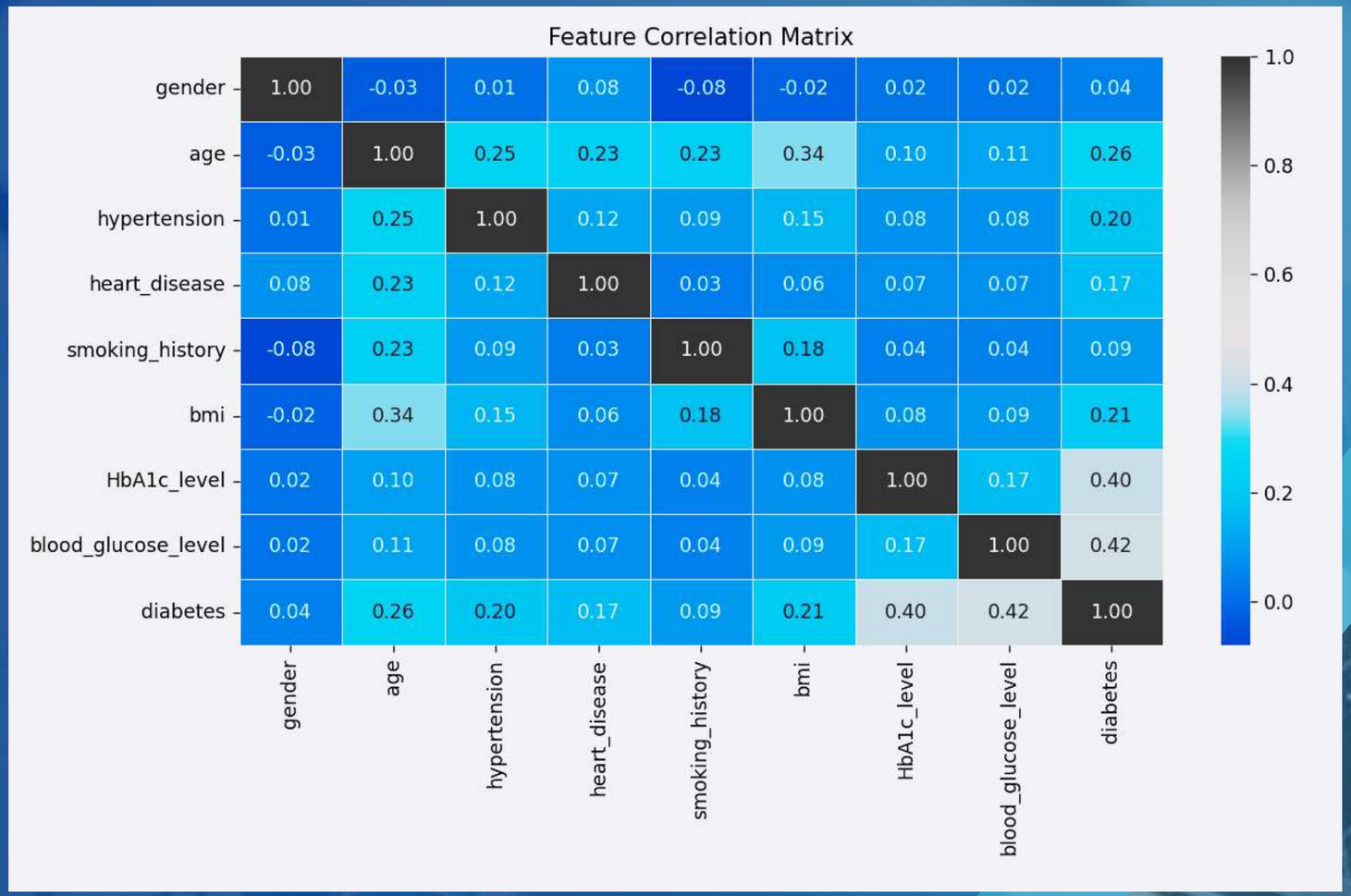
- Never smoked – Majority at 54.7%.
- Former & current smokers – Around 14.5% each.
- Others – Not current (10%), past smokers (6.2%).

HBA1C LEVELS DISTRIBUTION



- Overlap at 6.0–6.5% – Critical range for diagnosis.
- Diabetics have higher HbA1c – Peaks above 7%.
- Non-diabetics vary more – Multiple peaks below 6%.

FEATURE ENGINEERING POTENTIAL



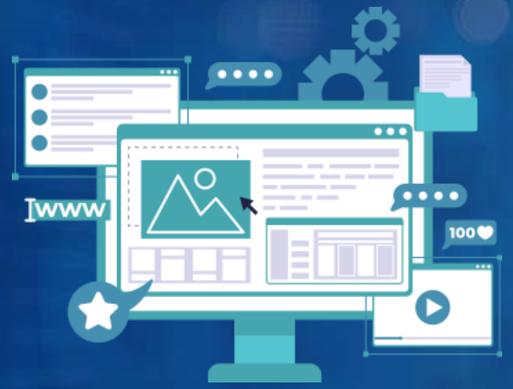
HbA1c & Blood Glucose: +0.75
(Strong Positive Correlation)
Higher HbA1c is strongly associated with higher Blood Glucose levels.

Age & Hypertension: +0.4
(Moderate Positive Correlation)
Older individuals are more likely to have hypertension.

BMI & Diabetes: +0.3
(Moderate Positive Correlation)
Higher BMI slightly increases diabetes risk.

MODEL DEVELOPMENT: TECHNOLOGIES & FRAMEWORKS

GlucoSense



TECH STACK:

- **Frontend**: React
- **Backend**: FlaskAPI, Python
- **Machine Learning Model**: XGBoost

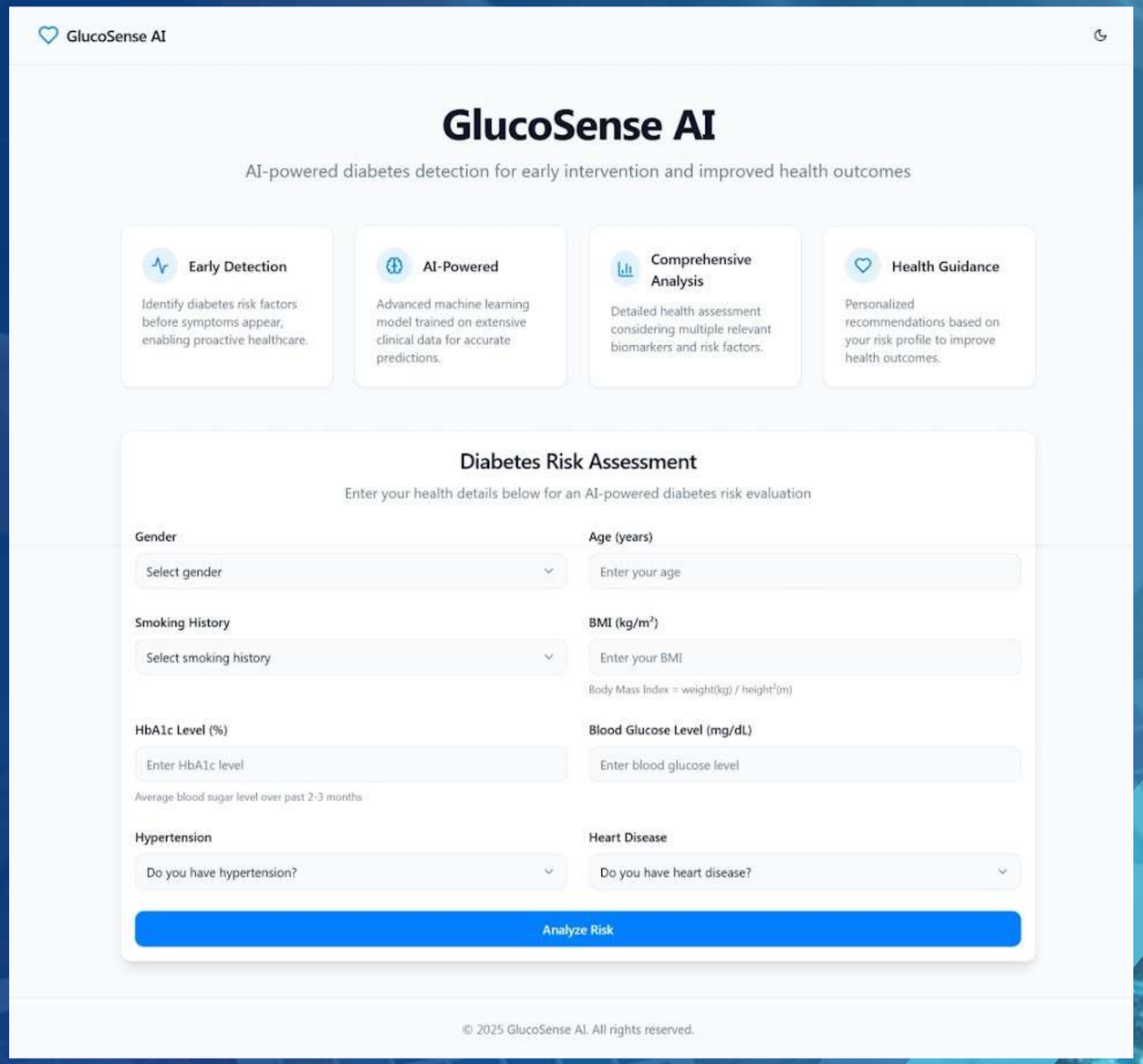
FRAMEWORK:

1. GlucoSense is a diabetes prediction web app that leverages machine learning to evaluate diabetes risk based on user inputs.
2. It features a React frontend, FlaskAPI backend, and an XGBoost model for predictions.

FEATURES:

1. User-friendly interface for inputting health metrics.
2. Real-time diabetes risk prediction using XGBoost.
3. API-based communication between frontend and backend.

DEPLOYMENT PROCESS



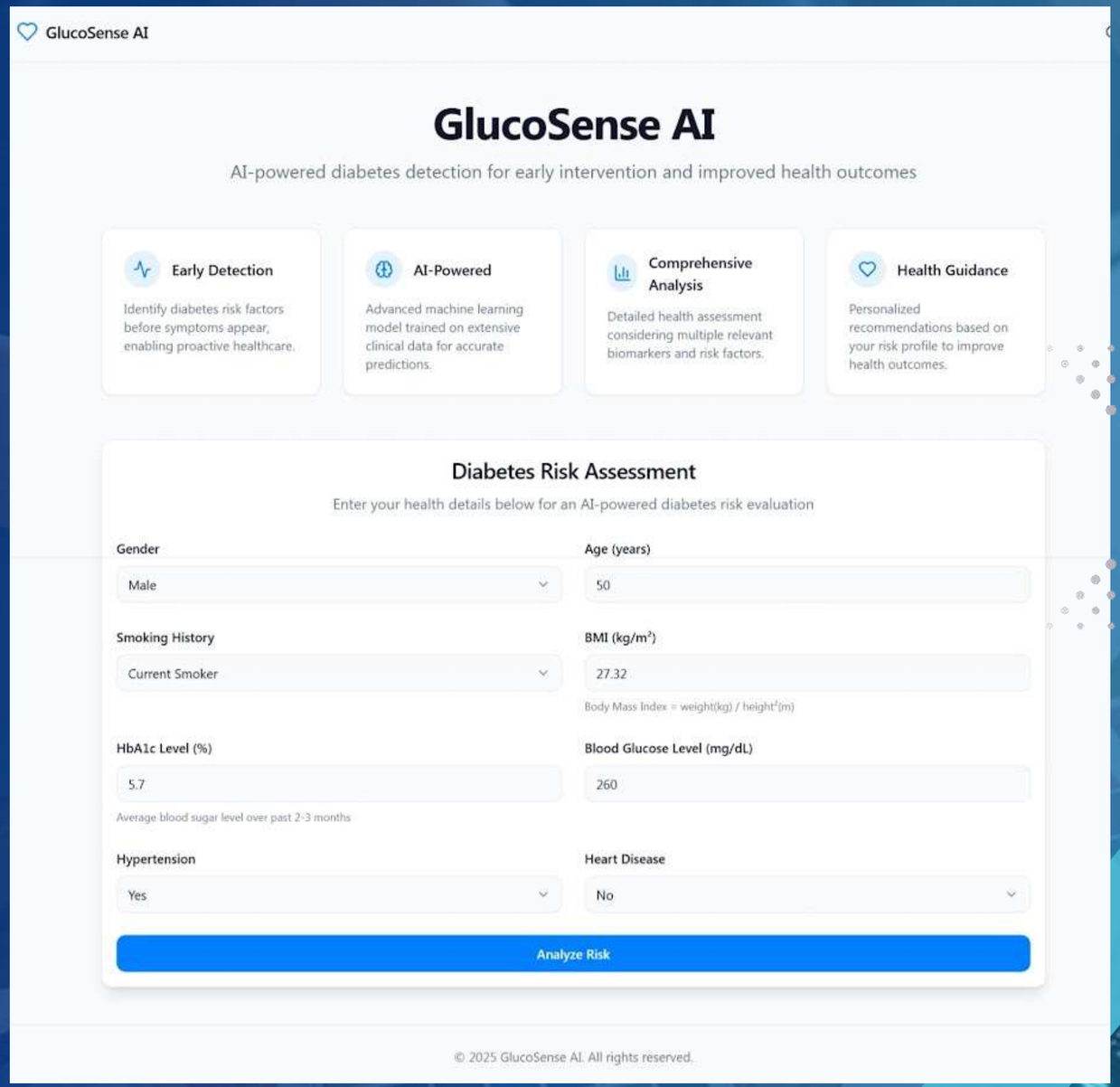
User Input: Users enter health metrics such as HbA1c, BMI, hypertension, and glucose level.

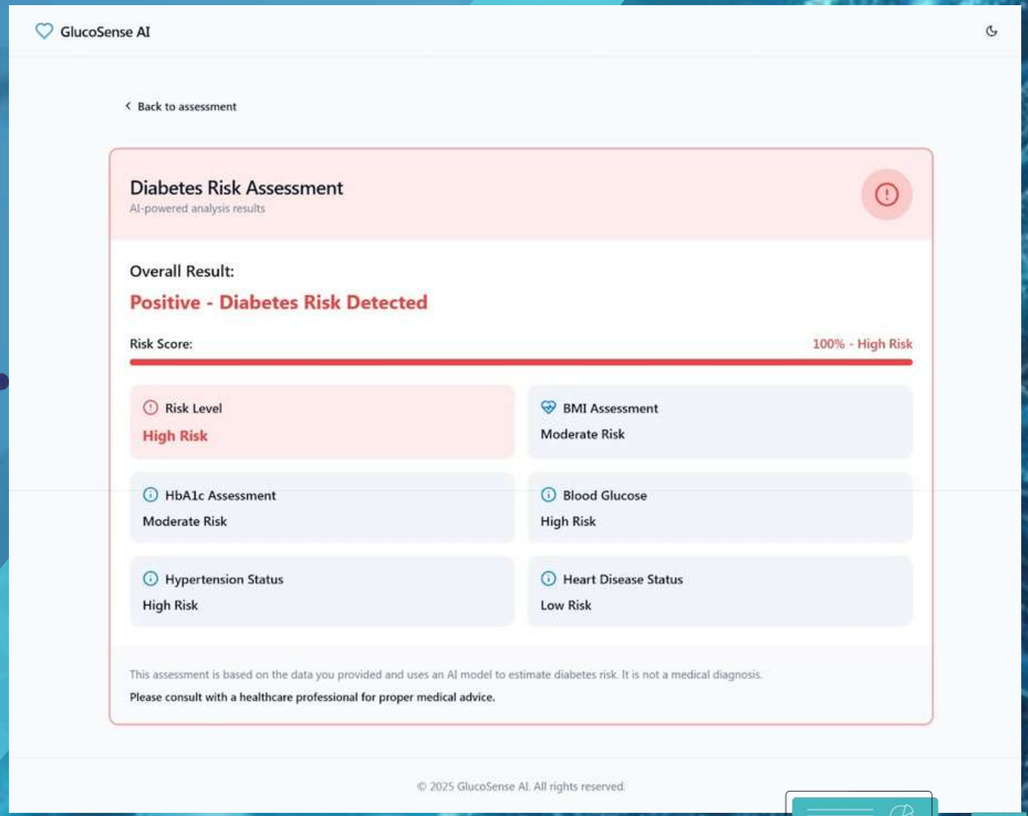
Data Processing: The entered data is analyzed using the prediction model.

Risk Assessment: The system evaluates and determines the diabetes risk.

Result Display: The predicted diabetes risk is instantly shown on the web app.

RESULTS


GlucoSense AI
GlucoSense AI
AI-powered diabetes detection for early intervention and improved health outcomes
Early Detection
Identify diabetes risk factors before symptoms appear, enabling proactive healthcare.
AI-Powered
Advanced machine learning model trained on extensive clinical data for accurate predictions.
Comprehensive Analysis
Detailed health assessment considering multiple relevant biomarkers and risk factors.
Health Guidance
Personalized recommendations based on your risk profile to improve health outcomes.
Diabetes Risk Assessment
Enter your health details below for an AI-powered diabetes risk evaluation
Gender
Male
Age (years)
50
Smoking History
Current Smoker
BMI (kg/m²)
27.32
Body Mass Index = weight(kg) / height²(m)
HbA1c Level (%)
5.7
Average blood sugar level over past 2-3 months
Blood Glucose Level (mg/dL)
260
Hypertension
Yes
Heart Disease
No
Analyze Risk
© 2025 GlucoSense AI. All rights reserved.


GlucoSense AI
Back to assessment
Diabetes Risk Assessment
AI-powered analysis results
Overall Result:
Positive - Diabetes Risk Detected
Risk Score:
100% - High Risk
Risk Level
High Risk
BMI Assessment
Moderate Risk
HbA1c Assessment
Moderate Risk
Blood Glucose
High Risk
Hypertension Status
High Risk
Heart Disease Status
Low Risk
This assessment is based on the data you provided and uses an AI model to estimate diabetes risk. It is not a medical diagnosis.
Please consult with a healthcare professional for proper medical advice.
© 2025 GlucoSense AI. All rights reserved.


HTML
CSS

FUTURE SCOPE

1 MOBILE APP DEVELOPMENT

Launching a mobile application for seamless access, real-time alerts, and user-friendly tracking of health metrics.

2 INTEGRATION WITH WEARABLE DEVICES

Connecting with smartwatches and continuous glucose monitors (CGMs) for real-time diabetes risk assessment and proactive health monitoring.

3 PERSONALIZED HEALTH INSIGHTS

Expanding features to provide tailored recommendations on lifestyle, diet, and medication adjustments based on user health profiles.

4 ENHANCED PREDICTION ACCURACY

Continuous model improvement through deep learning techniques and larger, diverse datasets to increase accuracy and reliability.

5 SELF-LEARNING & ADAPTIVE MODEL

Implementing reinforcement learning where the model continuously improves by learning from new user data and medical research trends.