Diabetes Prediction Using Machine Learning & Interactive Dashboards

Introduction & Purpose

The goal of this project was to build a full-stack, AI-powered application that predicts the likelihood of diabetes in individuals using clinical attributes from the Pima Indian Heritage dataset. By combining machine learning (ML) with interactive Tableau dashboards and a Flask-based web interface, the application allows users to explore health trends and make real-time predictions.

This project highlights the practical use of AI in healthcare and demonstrates how data-driven insights and predictive analytics can be made accessible through dynamic, user-friendly tools.

Dataset & Motivation

We used the Pima Indian Diabetes dataset, which contains health metrics for female patients of Pima Indian heritage over the age of 21. Key features include glucose levels, BMI, insulin levels, age, and blood pressure.

Motivation:
Diabetes is a global health concern. Early prediction can help patients take preventive action. This dataset is widely used in health-related ML experimentation due to its medical relevance and feature diversity.

Data Cleaning & Preprocessing

- Handled missing values (represented as 0 in key features like glucose, BMI, insulin)
- Used SimpleImputer for imputing missing values
- Scaled numerical features using StandardScaler
- Added a PatientID column and set it as the index
- Created a ColumnTransformer to pipeline transformations

Machine Learning Experiment

Preprocessing

- Applied SMOTE to handle class imbalance
- Built a scikit-learn pipeline to automate preprocessing + modeling
- Train-test split with stratification
- Used cross-validation to ensure robustness

Models Tested

- Logistic Regression
- Random Forest
- Gradient Boosting
- Decision Tree
- SVM, kNN, AdaBoost (Additional experimentation)

Performance Metrics

- Accuracy
- Precision
- Recall
- F1-score
- ROC AUC

Best Model: The Random Forest Classifier yielded the highest overall performance with excellent balance between precision and recall. It was selected for deployment.

Flask Application Overview

The deployed web app features:
- Landing Page – Brief explanation and project overview
- ML Form – User can enter feature values to get a real-time diabetes prediction
- Dashboard 1 – Exploratory overview of patient demographics and features
- Dashboard 2 – Health condition comparisons and outcome statistics
- About Us, Works Cited, Report Page – Additional contextual content

[Add Screenshot of Flask App Here]

Tableau Dashboard Designs

Dashboard 1: Demographic Analysis

- Visuals: Bar charts, pie charts, histograms
- Filters: Age range, BMI group, Outcome (diabetes or not)
- Insights: Distribution of patients by age, BMI, pregnancies, and diabetes occurrence

[Add Screenshot of Dashboard 1 Here]

Dashboard 2: Feature Correlation & Health Metrics

- Visuals: Lollipop plots, heatmaps, grouped bar charts
- Filters: Insulin level, glucose, and diabetes status
- Insights: Strong patterns between glucose levels and diabetes presence

[Add Screenshot of Dashboard 2 Here]

Dashboard Questions Answered

- Are there particular health metrics that strongly correlate with diabetes?
- How do demographic factors (like age or pregnancy count) affect diabetes likelihood?
- What are the most common feature values for diabetic vs. non-diabetic individuals?

Limitations & Bias

- Data Imbalance: Originally imbalanced, addressed via SMOTE
- Feature Simplicity: Dataset lacks lifestyle or family history data
- No Gender Diversity: Only includes female patients
- Model Generalizability: May not generalize well beyond Pima Indian demographics

Conclusions & Future Work

We developed a reliable, full-stack ML application that can assist with early diabetes screening. The use of Tableau makes data trends easy to explore, while the ML model enables real-time inference for healthcare prediction.

Future improvements may include:
- Add deeper feature sets (e.g., family history, activity level)
- Use SHAP for model explainability
- Expand app with user authentication
- Improve front-end aesthetics and mobile responsiveness
- Deploy via Azure or AWS for scaling