Diabetes Prediction Using Machine Learning Project Overview This project predicts whether a patient is diabetic using the Pima Indians Diabetes Dataset. It includes data preprocessing, visualization, model training, and evaluation. Dataset Total Samples: 768

Features: 8 medical attributes

Target: Outcome (0 = Non-diabetic, 1 = Diabetic) Data Preprocessing Steps 1. Import libraries

2. Load dataset

3. Dataset overview

4. Identify zero values

5. Replace zeros with NaN

6. Handle missing values

7. Remove duplicates

8. Label encoding

9. Feature-target split

10. Feature scaling

11. Train-test split Exploratory Data Analysis - Histograms

- Correlation heatmap

- Scatter plot

- ROC curve

- SVM learning curve Machine Learning Models 1. Logistic Regression

2. KNN

3. Decision Tree

4. Random Forest Model Performance Summary Logistic Regression: 75%

KNN: 73%

Decision Tree: 69%

Random Forest: **77%** (Best) Conclusion Random Forest achieved the highest accuracy and performed best overall. Requirements pandas

numpy

matplotlib

seaborn

scikit-learn