**Diabetes Prediction with Machine Learning**

**SDG 3: Good Health and Well-being**

**Problem Statement**

Diabetes affects over 400 million people globally, with many cases undiagnosed until

complications arise.

This project addresses SDG 3 by developing a machine learning model that can predict the likelihood of diabetes using basic diagnostic data helping promote early detection and better healthcare outcomes.

**Machine Learning Approach**

Type: Supervised Learning

Model Used: Random Forest Classifier

Dataset: Open-source dataset with 768 records and 8 features

Features include: Glucose, BMI, Insulin, Blood Pressure, Age, etc.

Tools: Python, Pandas, Scikit-learn, Matplotlib, Seaborn

**Workflow Summary**

Data Exploration: Visualized distributions and correlations

Preprocessing: No missing values; dataset was clean

Train/Test Split: 70% training, 30% testing

Model Training: Random Forest with default parameters

**Evaluation:**

Accuracy: ~85%

Confusion Matrix: Balanced classification

Feature Importance: Glucose and BMI were top predictors

**Results**

The model correctly predicted diabetes with ~85% accuracy

It showed good balance between false positives and false negatives

Could be integrated into a basic screening tool for community health workers

**Ethical & Social Considerations**

Bias: Dataset may not reflect global population diversity

Fairness: AI models should support, not replace, medical professionals

Impact: Enables early screening in low-resource settings, reducing health system burden

**SDG Contribution**

This project contributes to Sustainable Development Goal 3 by using AI to assist in preventive healthcare, promoting health equity, and reducing the long-term effects of chronic diseases through early intervention.