**Overview:**This project aims to build a predictive system that can determine whether a person is likely to have diabetes based on common medical inputs such as glucose level, blood pressure, BMI, and age. It leverages machine learning algorithms, data preprocessing, and model evaluation techniques to produce reliable predictions, especially suitable for early diagnosis support.  
  
**🎯 Objectives:**

* Develop a machine learning model to classify diabetes risk.
* Handle imbalanced data using SMOTE for fair prediction accuracy.
* Enable real-time predictions through a simple interface.
* Evaluate models with performance metrics like accuracy, precision, and recall.

**🧠 Key Features:**

* Data Cleaning: Handles missing or unrealistic zero values.
* Standardization: Uses StandardScaler to normalize features.
* Model Comparison: Trained Logistic Regression, SVM, Random Forest, and XGBoost.
* Final Model: XGBoost with SMOTE (Synthetic Minority Over-sampling Technique) for improved diabetic case detection.
* Prediction Function: Takes custom input and returns real-time output as “Diabetic” or “Not Diabetic.”
* Performance: Achieves ~84% accuracy with high recall for diabetic cases.

**📂 Tools and Libraries Used:**  
Python, Pandas, NumPy, Matplotlib, Seaborn  
  
Scikit-learn for ML models and preprocessing  
  
XGBoost for the final predictive model  
  
imbalanced-learn (SMOTE) for data balancing  
  
**🩺 Dataset:**  
*Source*: Pima Indian Diabetes Dataset.  
  
*Features*: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, Age.  
  
*Target*: Outcome (1 = Diabetic, 0 = Not Diabetic).  
  
**✅ Final Output:**A trained model that can predict diabetes using 8 medical inputs.  
  
*Accuracy*: ~84%  
  
Built-in prediction function for real-time results  
  
Ready to be integrated into web applications or dashboards.