**🩺 Diabetes Diagnosis Prediction App**

This project uses machine learning to predict whether a patient is diabetic based on clinical features such as glucose level, BMI, insulin, and age. It simulates how ML models can support early medical diagnosis and assist healthcare professionals in decision-making.

**📊 Dataset**

* **Source:** Pima Indians Diabetes Dataset (publicly available)
* **Features:**  
  Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age
* **Target Variable:** Outcome (0 = non-diabetic, 1 = diabetic)

**🛠 Workflow Summary**

1. **Data Cleaning**
   * Replaced biologically invalid zeros with NaNs
   * Imputed missing values using column means
2. **Exploratory Data Analysis (EDA)**
   * Visualized distributions, relationships, and correlations
   * Detected key feature influences like Glucose, BMI, and Age
3. **Feature Preparation**
   * Scaled input features using StandardScaler
   * Split data into training and testing sets
4. **Model Building**
   * Tested multiple models:
     + Support Vector Machine (SVM)
     + K-Nearest Neighbors (KNN)
     + Decision Tree (with hyperparameter tuning)
   * Final model selected:  
     DecisionTreeClassifier(criterion='entropy', max\_depth=4, min\_samples\_leaf=10)
5. **Evaluation**
   * Final model achieved:
     + **Accuracy:** 70%
     + **Recall (Diabetic Cases):** 0.62
     + **F1-Score (Diabetic Cases):** 0.59
6. **Feature Importance**
   * Top features: Glucose, BMI, Age

**✅ Why Decision Tree Was Chosen**

* Outperformed SVM and KNN in detecting diabetic cases
* High interpretability crucial in medical applications
* Balanced accuracy and recall, especially for minority class
* Easy to deploy and understand

**💾 Files Included**

| **File** | **Description** |
| --- | --- |
| Diabetes Daignosis.ipynb | Full notebook with code, visuals, and interpretations |
| diabetes\_model.pkl | Final trained Decision Tree model |
| feature\_columns.pkl | List of features used for prediction |
| README.md | Project overview and documentation |

**🧠 Skills Demonstrated**

* Data Cleaning & Imputation
* Exploratory Data Analysis
* Classification Model Building
* Hyperparameter Tuning with GridSearchCV
* Model Evaluation & Interpretation
* Feature Importance Extraction
* Model Serialization with Pickle