Elements of AIML

Assignment - 1



Name - ANANYA KALIA

SAP ID - 500122859

Batch – 11

Roll No. – R2142230337

**Topic: Prediction of Diabetes in**

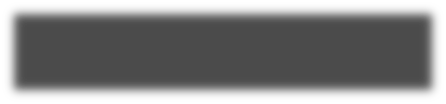
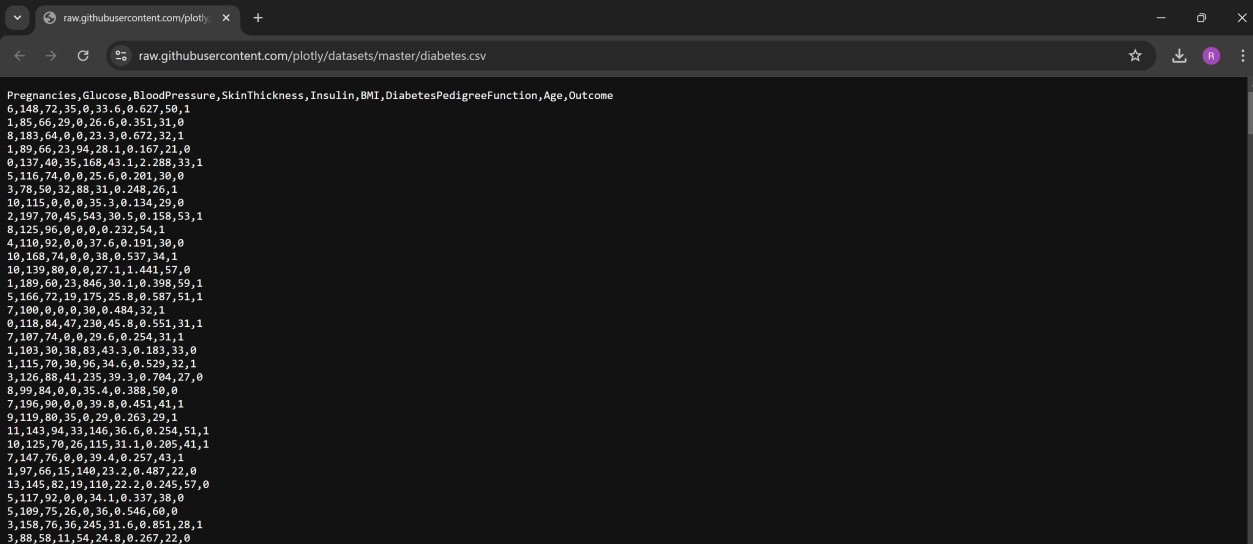
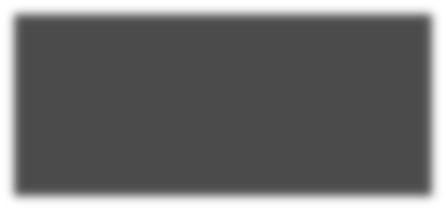
**Females(P\_D\_F)**

**STEPS OF THE CODE:**

**Step 1: Data Acquisition --** Look for a problem and a dataset that you can use

Problem :- Predict diabetes risk based on health metrics

Dataset :- Diabetes.csv (Imported From GitHub)



**Step 2 --** Define the methodology and the objectives of your work

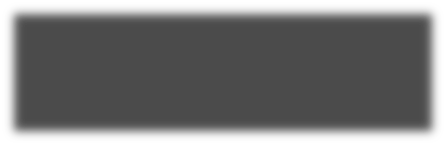
**Methodology**

1. **Data Preprocessing**: The diabetes dataset is loaded, cleaned, and preprocessed. Missing or zero values in key medical features are replaced with median values. SMOTE is applied to address class imbalance, and all features are standardized using StandardScaler
2. **Model Training & Evaluation**: Four models—Logistic Regression, Random Forest, SVM, and KNN—are trained using 5-fold crossvalidation to evaluate accuracy and precision. The model with the best performance is chosen for real-time predictions.
3. **User Interaction & Deployment**: The best model is trained on the full dataset, allowing for user input of health metrics to predict diabetes risk. The model and scaler are saved as .pkl files for easy deployment.

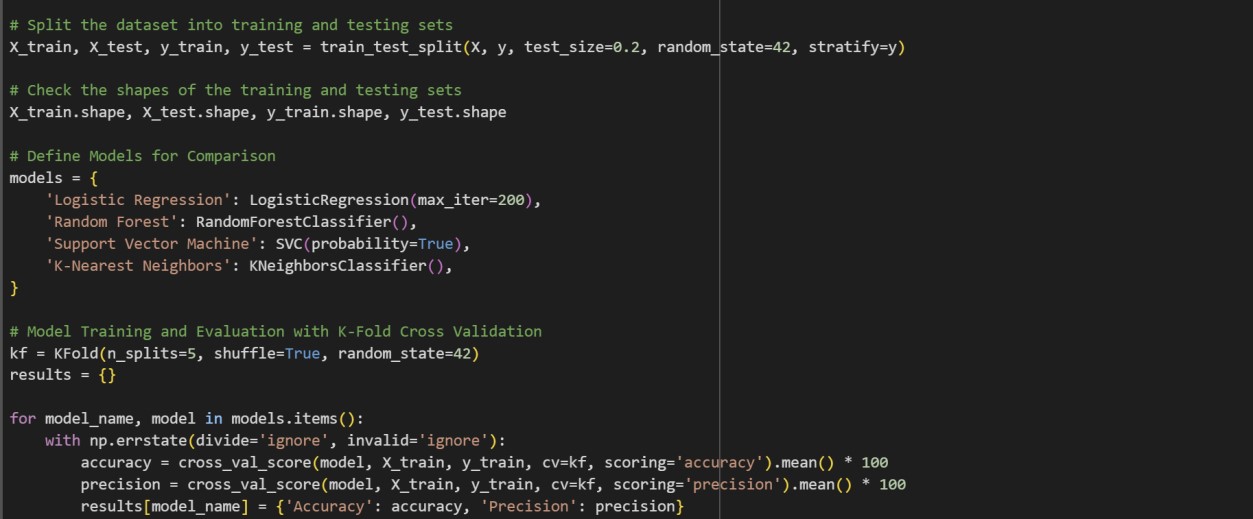
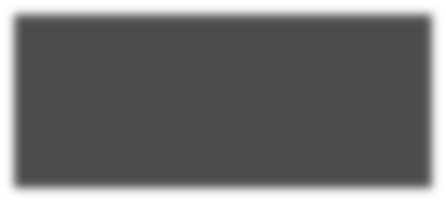
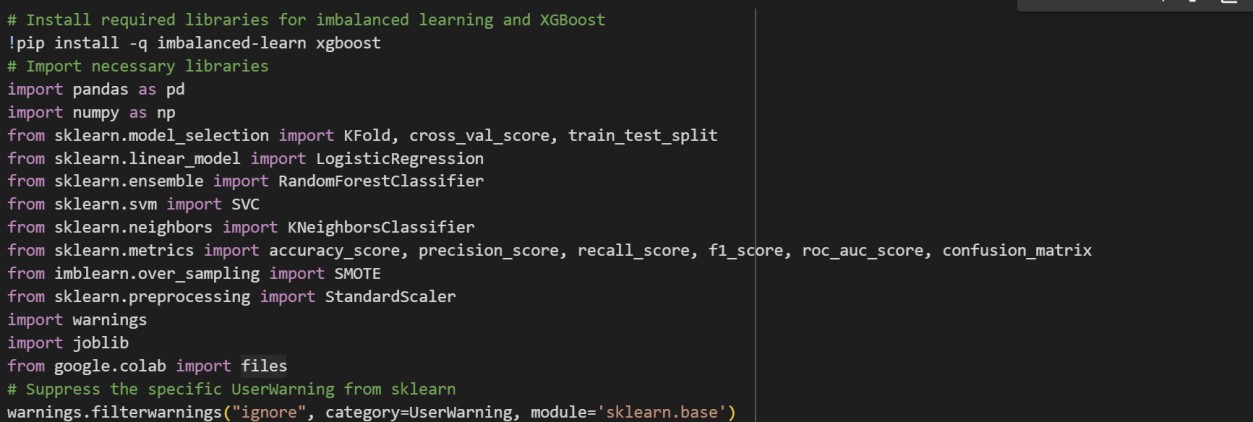
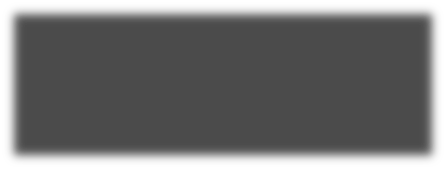
**Objectives**

1. **Develop an Accurate Diabetes Prediction Model** using balanced data and optimal scaling.
2. **Select the Best Algorithm** for robust predictions by comparing multiple models.
3. **Enable Real-Time Predictions** through user inputs.
4. **Deploy a Scalable Tool** that supports healthcare applications for early diabetes detection.

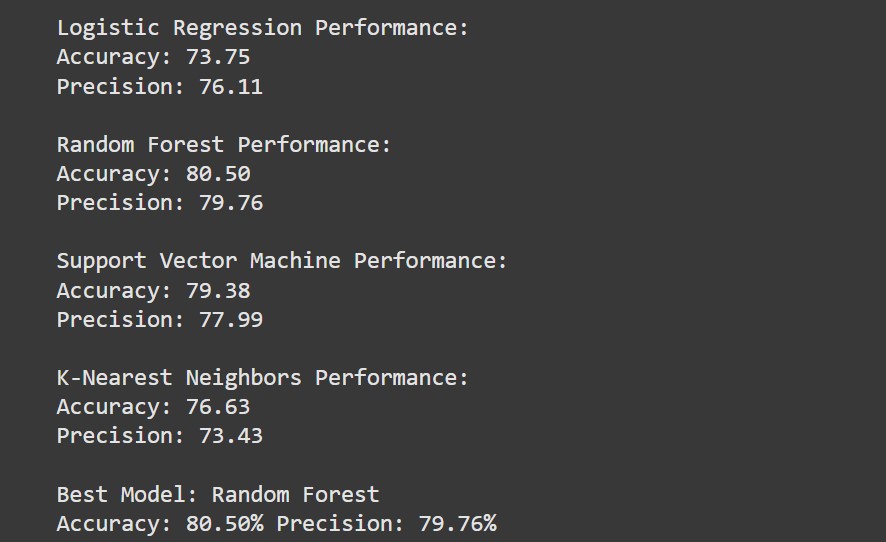
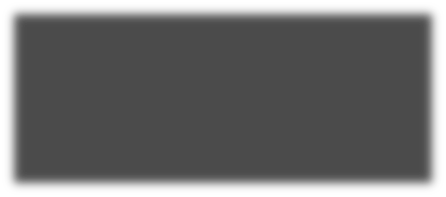
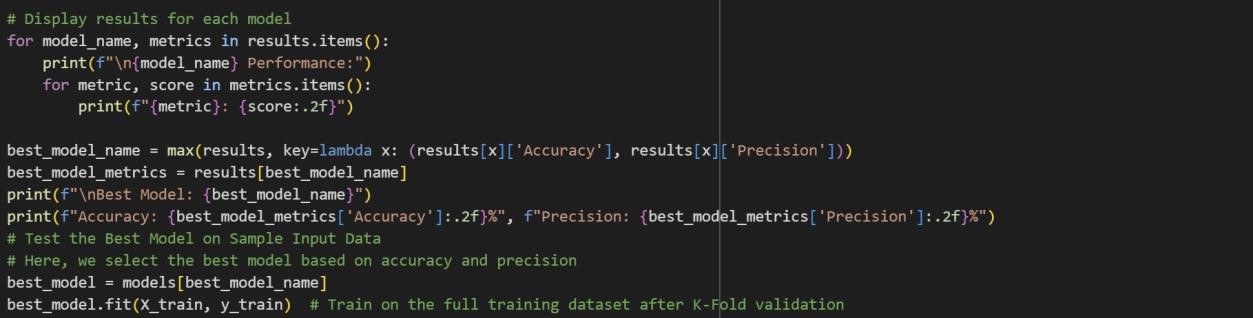
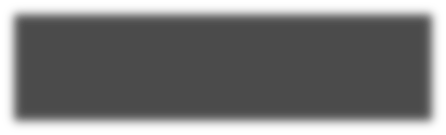
# Step 3: Data Preprocessing



**Step 4:** Use multiple ML methods and validate them using KFold Cross Validation.



**Step 5**: Comparing the results using suitable performance metrics such as accuracy, precision.



# Testing the model on sample data (user given )

This code segment takes user input for various health metrics related to diabetes risk and uses the trained model to predict the likelihood of diabetes. Here’s a brief breakdown of the steps:

1. Collect User Input: Prompts the user to enter values for key features such as pregnancies, glucose level, blood pressure, skin thickness, insulin level, BMI, diabetes pedigree function, and age.
2. Scale the Input: Converts the user input into a NumPy array, then scales it using the previously fitted StandardScaler to match the model’s training data scale.
3. Make Prediction: The model predicts whether the user has diabetes (1) or not (0) and provides the probability of diabetes.
4. Display Results: Prints the prediction (0 or 1) and the probability (as a percentage) of diabetes for the input data.

This enables real-time diabetes risk assessment based on user-provided health metrics.

