Assignment 2 - Logistic Regression Model for Diabetes Prediction

This assignment focuses on building a **Logistic Regression model** to predict diabetes using a dataset ('updated dataset.csv'). The workflow includes data preprocessing, feature encoding, model training, evaluation, and prediction using a sample input.

Code Steps:
1. Import necessary libraries: numpy, pandas, sklearn, pickle.
2. Load dataset and handle missing values using a safe_fillna() function.
3. Rename columns for better readability and consistency.
4. Prepare features (X) and target (y). Encode categorical features with one-hot encoding.
5. Split data into training and testing sets (80/20).
6. Standardize features using StandardScaler.
7. Train a Logistic Regression model (max_iter=1000).
8. Evaluate the model using:
   • Accuracy Score
   • Confusion Matrix
   • Classification Report
9. Save the trained model, scaler, and feature names using pickle.
10. Perform prediction on a new sample input and print whether it's 'Diabetic' or 'Not Diabetic'.

Expected Output:
■ Data loaded successfully
■ Model training completed
■ Accuracy: Displayed as a decimal (e.g., 0.86)
■ Confusion Matrix: Printed as a table of true/false positives/negatives
■ Classification Report: Displays precision, recall, f1-score
■ Model, Scaler & Features saved successfully
■ Prediction for sample input: "Diabetic" or "Not Diabetic"

Conclusion:
This assignment demonstrates the complete machine learning pipeline for diabetes prediction. It involves preprocessing data, training a Logistic Regression model, evaluating its performance, and making new predictions. The model, scaler, and feature set are serialized for future use.