Diabetes Prediction

Project Objective

The objective of this project is to build a machine learning model that can predict whether a person is diabetic or non-diabetic based on medical diagnostic measurements. This helps in early detection of diabetes and supports healthcare decision-making.

Problem Type

Machine Learning Type: **Supervised Learning**
Task: Binary Classification
Target Variable: Diabetes Outcome (0 = Non-Diabetic, 1 = Diabetic)

Dataset Description

Dataset Name: PIMA Indians Diabetes Dataset
Source: UCI Machine Learning Repository / Kaggle
Total Instances: 768
Number of Features: 8
Feature Type: Numerical
Target Classes: 0 → Non-Diabetic, 1 → Diabetic

Input and Output

Input (X): 8 medical numerical features
Output (y): 0 → Non-Diabetic, 1 → Diabetic

Algorithm Used

Support Vector Machine (SVM)
Chosen because it is powerful for classification tasks, works well with high-dimensional data, and provides good generalization performance in medical prediction problems.

Project Workflow

1. Load the dataset
2. Explore and analyze the data
3. Handle missing or zero values
4. Separate features and target variable
5. Perform feature scaling
6. Split data into training and testing sets
7. Train the SVM classifier
8. Evaluate model performance
9. Make predictions on new data

Model Evaluation

Evaluation Metrics:
- Accuracy Score
- Confusion Matrix
- Precision, Recall, F1-score

Accuracy is calculated on both training and testing datasets to evaluate model performance and avoid overfitting.

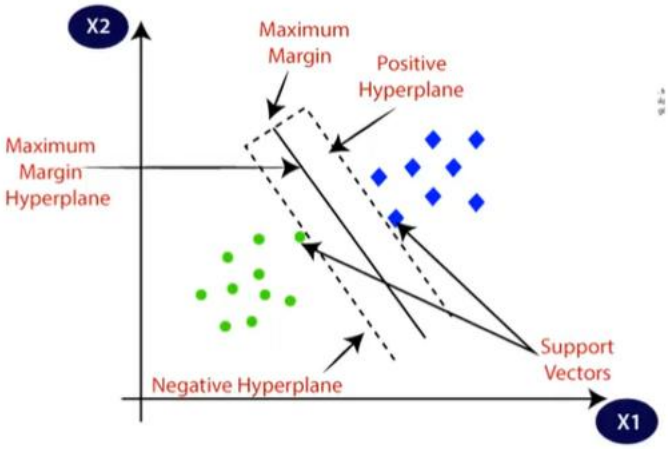
Conclusion

The SVM model successfully predicts diabetes based on medical diagnostic data. This project demonstrates a complete end-to-end machine learning workflow including preprocessing, feature scaling, model training, evaluation, and prediction using Support Vector Machines.

Features Description

Feature Name	Description
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration
BloodPressure	Diastolic blood pressure (mm Hg)
SkinThickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body Mass Index
DiabetesPedigreeFunction	Diabetes hereditary score
Age	Age of the patient

Support Vector Machine



Work Flow