Diabetes Prediction Model

Using SVM Classifier

Introduction

This project aims to develop a machine learning model using Support Vector Machine (SVM) to predict diabetes based on medical parameters. The dataset used is the PIMA Indian Diabetes Dataset. After preprocessing and feature scaling, the data is split into training and testing sets. The SVM model is trained and evaluated, achieving an accuracy of approximately 75–80% depending on the kernel and parameters. The final model can aid in early detection of diabetes, thus helping in timely intervention.

Objective

To build a machine learning model using Support Vector Machine (SVM) to predict whether a patient is likely to be diagnosed with diabetes based on medical attributes.Dataset

PIMA Indian Diabetes Dataset

Features:

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, and Age

Target:
Outcome (1 = diabetic, 0 = non-diabetic)

Steps Involved:

1. Import Libraries / dependencies (numpy, panda, train_test_split function, accuracy_score, StandardScaler, SVM)
2. Load the Dataset
3. Data Preprocessing
    - Check for nulls
    - Replace null values with ‘ ’
    - Data Standardization
    - Split into data and labels
4. Split the Data into test and train dataset
5. Train SVM model
6. Evaluate the model