Diabetes Prediction
With Machine Learning using Classification Techniques

Abstract

Diabetes is a serious chronic condition that poses a great threat to human health. Diabetes is characterized by a high blood glucose level caused by defective insulin secretion or impaired biological effects, or both. Diabetes can result in chronic damage and dysfunction of various tissues, including the eyes, kidneys, heart, blood vessels, and nerves. Therefore, the ability to diagnose and analyze diabetes rapidly and accurately is a topic worth exploring. The earlier we can detect and manage it, the better. Using their physical examination data, machine learning can help people to make a preliminary diagnosis of diabetes mellitus, and it can also be a helpful reference for doctors. The aim of this study is to predict diabetes accurately using a variety of machine learning techniques, including Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression, AdaBoost, K-Nearest Neighbor (KNN), and Gradient Boosting Classifiers.

Data

The dataset is Pima Indians diabetics data. Specifically, all patients are Pima Indian women with a minimum age of 21. This dataset contains 8 features, with 768 observations. The features include pregnancy times, plasma glucose concentration following oral glucose tolerance tests, diastolic blood pressure, triceps skin fold thickness, 2-hour serum insulin, body mass index, diabetes pedigree function, and age.

Algorithms

- ### Expolarity Data Analysis

  In this part, we have created plots to explore the relationship between the target and variables and between the variables themselves.

- ### Pre-processing

  1- Imputed missing data with mean, checked for null.
  2- Split data into train and test.
  3- Scaling data

- ## Model Building

  Several models were developed using different Machine learning techniques such as Support Vector Machine, Random Forest, Decision Tree, Logistic Regression, K-Nearest Neighbor, Ada Boost and Gradient Boosting Classifier.
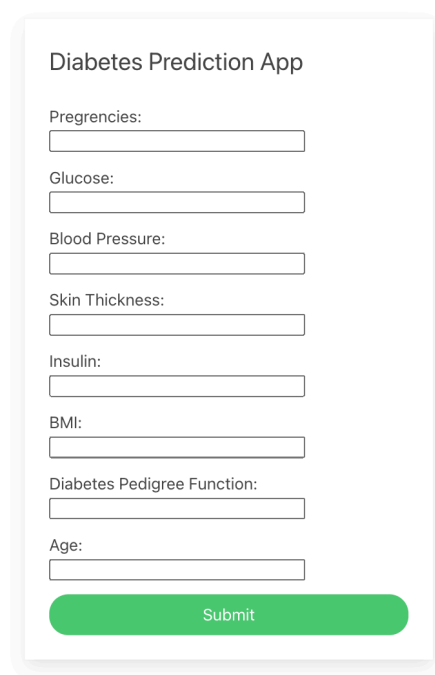
- ## Model Selection and Evaluation

  The evaluation of our models was based on Recall metric .

Tools

- Jupyter Notebook will be used to create and document live code and visualization
- Pandas for data manipulation
- Sklearn for model building
- Matplotlib and Seaborn for plotting
- Matplotlib and Seaborn for plotting

Communication

In addition, we made a dashboard to show the result of the preferred predictive model.

Diabetes Prediction App

Pregrencies:

Glucose:

Blood Pressure:

Skin Thickness:

Insulin:

BMI:

Diabetes Pedigree Function:

Age:

Submit

Congratulations

You tested **NEGATIVE** for the diabetes.

You are a fit and healthy person!
Eat Healty and stay fit.

Sorry to say that

You tested **positive** for diabetes. don't worry consult a doctor soon! you will be alright

Eat Healty ! do Yoga ! Be Happy ! .