Documentation of the project

**Predict Diabetes Outcomes Based on Patient Data Project**

Team members

HASSAN MOSTAFA HASSAN 2100776

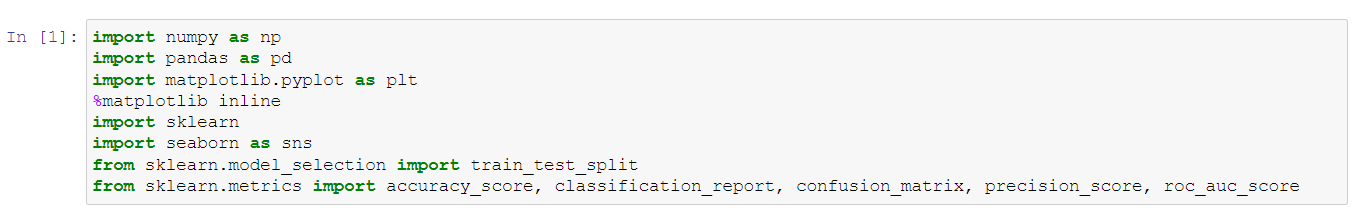
KAREEM AYMAN ELEMARY 2100837

AHMED HUSSIEN ALM ELDIN 2100336

**Problem statement**

Predict if a person will have diabetes in the future based on diabetes prediction dataset by understanding its features and done machine learning algorithms on it.

First import necessary libraries



First We Done Preprocessing

Read the dataset and show some of its records

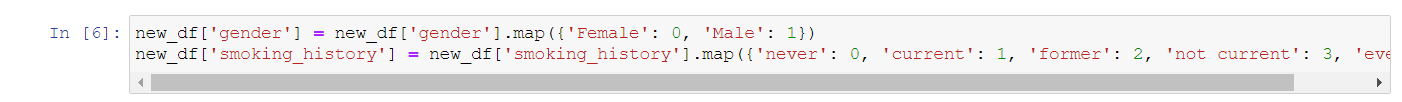
صورة تحتوي على نص, رقم, الخط, خط

تم إنشاء الوصف تلقائياً

Get the shape of the data and info about itصورة تحتوي على نص, لقطة شاشة, رقم, برمجيات

تم إنشاء الوصف تلقائياً

We have column values like gender and smoking history as a string object and this is not applicable with our data fitting operation, so we have to convert it to float numbers.



And here is the dataset after we done that

صورة تحتوي على نص, رقم, الخط, خط

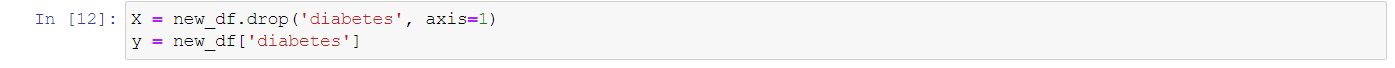
تم إنشاء الوصف تلقائياً

We want to have the number of people that have diabetes and number of people that has not and draw a graph for this.صورة تحتوي على نص, لقطة شاشة, رسم بياني, برمجيات

تم إنشاء الوصف تلقائياً

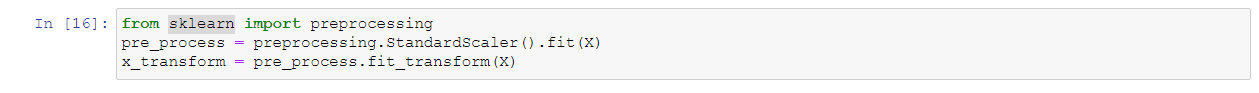
We need to remove null values in the dataset

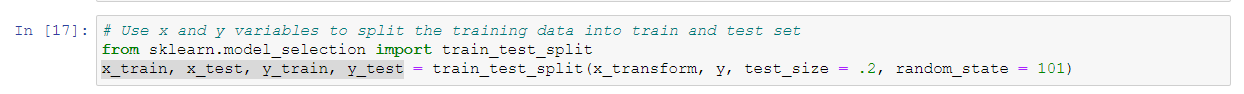
Now we need to begin to split the data to feed our model with it.



We also used describe() function to show more details about the dataصورة تحتوي على نص, الخط, رقم, لقطة شاشة

تم إنشاء الوصف تلقائياً

Now we make sure that we preprocess the data using sklearn preprocessing functions

Now we split the data to x\_train, x\_test, y\_train, y\_test to train the machine learning models

Now comes the machine Learning by feeding the logistic regression model with our data to train with it صورة تحتوي على نص, الخط, خط, لقطة شاشة

تم إنشاء الوصف تلقائياً

Now wr check our model training accuracy and the testing accuracy

صورة تحتوي على نص, لقطة شاشة, الخط, خط

تم إنشاء الوصف تلقائياً

Now we get classification report to our model that contains precision, recall and f1-score

صورة تحتوي على نص, لقطة شاشة, الخط, خط

تم إنشاء الوصف تلقائياً

Now we print the true positives false positives, true negatives and false negatives from the confusion matrix

صورة تحتوي على نص, لقطة شاشة, الخط, رقم

تم إنشاء الوصف تلقائياً

We also draw a grap to show the details about the confusion matrixصورة تحتوي على نص, لقطة شاشة, برمجيات, أيقونة الحاسوب

تم إنشاء الوصف تلقائياً

we draw a training error graph for our model

صورة تحتوي على نص, خط, رسم بياني, لقطة شاشة

تم إنشاء الوصف تلقائياً

We print area under the receiver operating characteristic curve (AUC-ROC).



Decision Tree Code

First we run the decision tree classfier on the data

صورة تحتوي على نص, الخط, خط, لقطة شاشة

تم إنشاء الوصف تلقائياً

Then predict the target variable for the training data and getting the training accuracy and the testing accuracy

صورة تحتوي على نص, لقطة شاشة, الخط, خط

تم إنشاء الوصف تلقائياً

Now we get classification report to our model that contains precision, recall and f1-score

صورة تحتوي على نص, لقطة شاشة, الخط, خط

تم إنشاء الوصف تلقائياً

Now we print the true positives false positives, true negatives and false negatives from the confusion matrix

صورة تحتوي على نص, لقطة شاشة, برمجيات, رقم

تم إنشاء الوصف تلقائياً

We also draw a grap to show the details about the confusion matriصورة تحتوي على نص, لقطة شاشة, برمجيات, أيقونة الحاسوب

تم إنشاء الوصف تلقائياًx

We draw a graph showing the relasionship between predicted error and max depth to show how error decreases every over depth inceases.

صورة تحتوي على نص, لقطة شاشة, رسم بياني, خط

تم إنشاء الوصف تلقائياً

We print area under the receiver operating characteristic curve (AUC-ROC).



Multi Layer Perceptron

First we run the Multi Layer Perceptronclassfier on the data

صورة تحتوي على نص, الخط, خط, لقطة شاشة

تم إنشاء الوصف تلقائياً

Then predict the target variable for the training data and getting the training accuracy and the testing accuracyصورة تحتوي على نص, الخط, لقطة شاشة, خط

تم إنشاء الوصف تلقائياً

Now we get classification report to our model that contains precision, recall and f1-score

صورة تحتوي على نص, الخط, لقطة شاشة

تم إنشاء الوصف تلقائياً

Now we print the true positives false positives, true negatives and false negatives from the confusion matrix

صورة تحتوي على نص, لقطة شاشة, الخط

تم إنشاء الوصف تلقائياً

We also draw a grap to show the details about the confusion matri

صورة تحتوي على نص, لقطة شاشة, برمجيات, رسم بياني

تم إنشاء الوصف تلقائياً

Now we draw a grap to show the relationship between number of iterations and training error

صورة تحتوي على نص, خط, رسم بياني, تخطيط

تم إنشاء الوصف تلقائياً

We print area under the receiver operating characteristic curve (AUC-ROC).



Git Hub Link

<https://github.com/HassanMostafaS/Ai-Project>