**Machine Learning Report**

**Mohamed Aziz Karoui**

**Introduction**

Hello this is my Machine Learning Report and here I will explain step by step all the things I did to my dataset.

First, I chose a Dataset called “Indian-Liver-patients.csv” which’s a Dataset that describe multiple patients from India up to a number of 590 with different criteria to determine if they are a Liver patient or not.

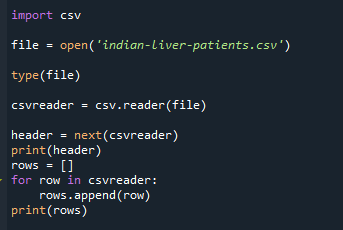
And in this project, we will try to analyze the dataset and come up with different observations and conclusions about the patients but first I got to present the attributes that we will use in this project and they are:

* Age of the patient: Which’s the age of a patient and is a number in interval of 0 to 100
* Gender of the patient: Which’s either Male or Female
* Total Bilirubin: Integer which’s a Medical Notion
* Direct Bilirubin: Integer which’s a Medical Notion
* Alkphos: Integer which’s a Medical Notion
* Alkaline Phosphotase: Integer which’s a Medical Notion
* Sgpt Alamine Aminotransferase: Integer which’s a Medical Notion
* Sgot Aspartate Aminotransferase: Integer which’s a Medical Notion
* Total Protiens: Integer which’s a Medical Notion
* Albumin: Integer which’s a Medical Notion
* Albumin and Globulin Ratio: Integer which’s a Medical Notion
* Liver Patient or Not: Which’s the conclusion of our dataset and it’s “1” for None Liver Patient and “2” for Liver Patient

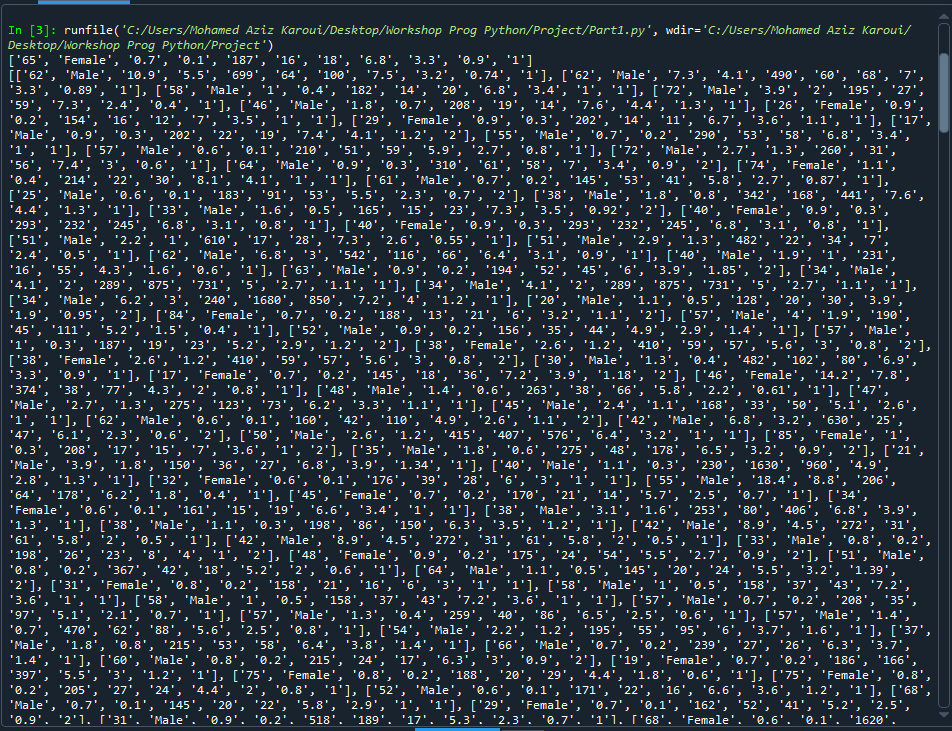
**Part 1 – Preparing Data**

1. Show the Original Data in its current form and we will use:

“import csv” to get the csv library then I used “open()” to get my dataset in a variable called file then “type()” to show the data

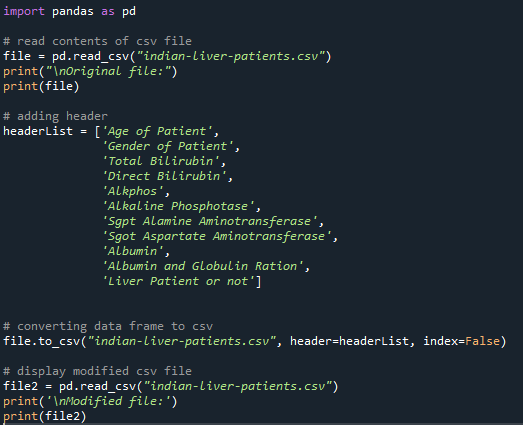


And this is the result of our script which’s all the data in our file

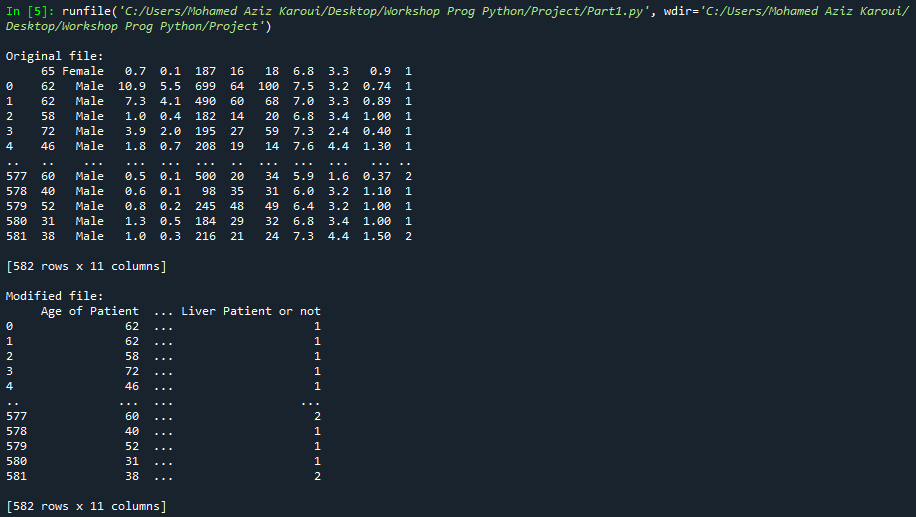


1. Add to each column the name of its attribute and show it

And we can do this thanks to function “.to\_csv()” where we can add our header there

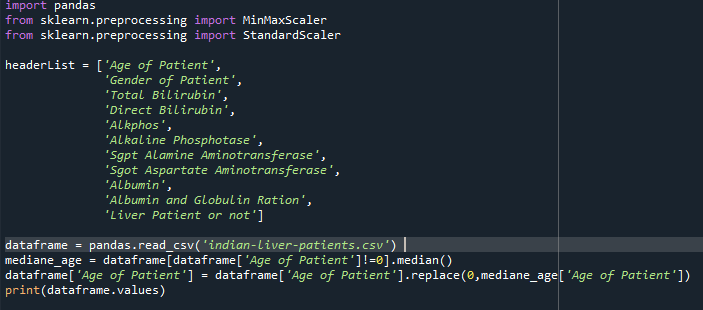


And this is the result of our script which shows the data before and after the changes and we can see after the changes there are the header names of columns of each attribute



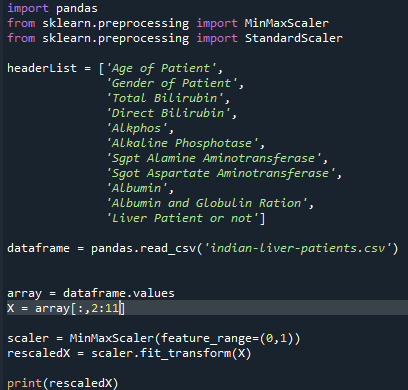
1. Calculate the median age of participants and replace the null numbers

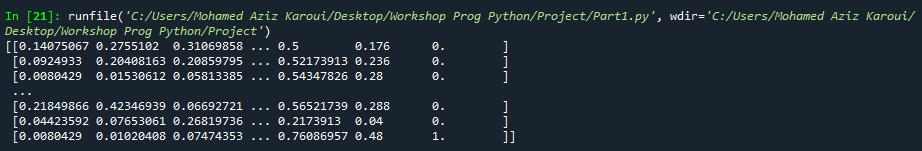
In this script we will use sklearn as library to calculate the median and change all the null numbers to the median



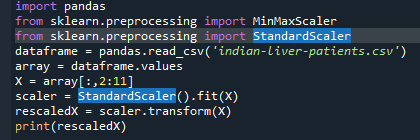
This procedure of replacing by median is only viable in case of age cause there’s no one with the age of 0 while the other 0 in medical terms are significative numbers.

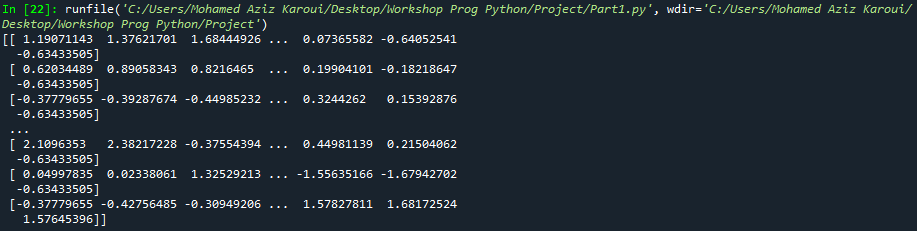
1. Scale uniformity





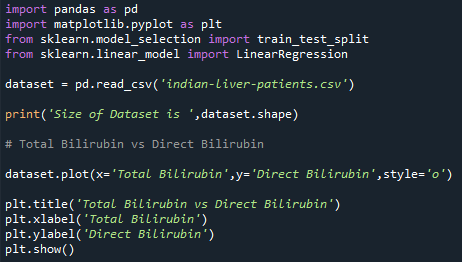
1. Normalization

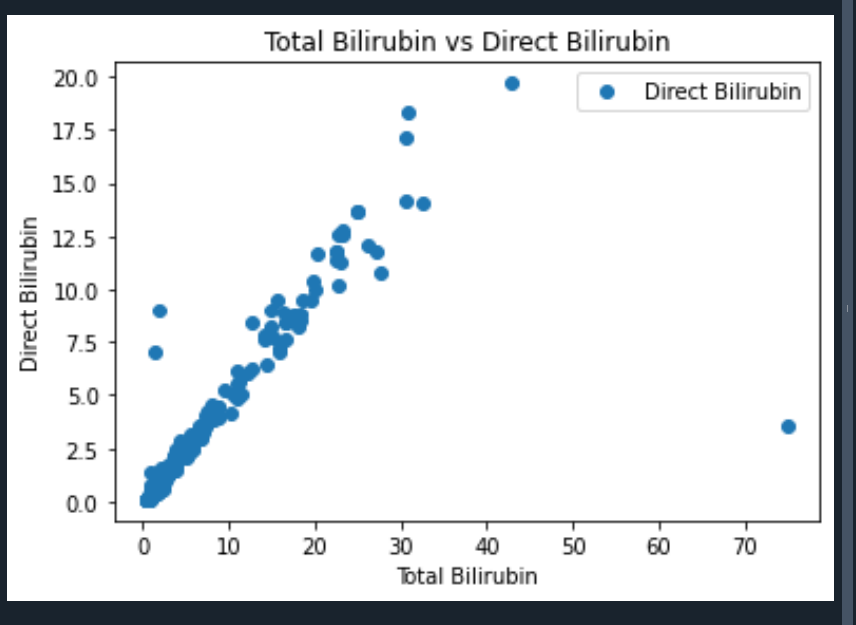




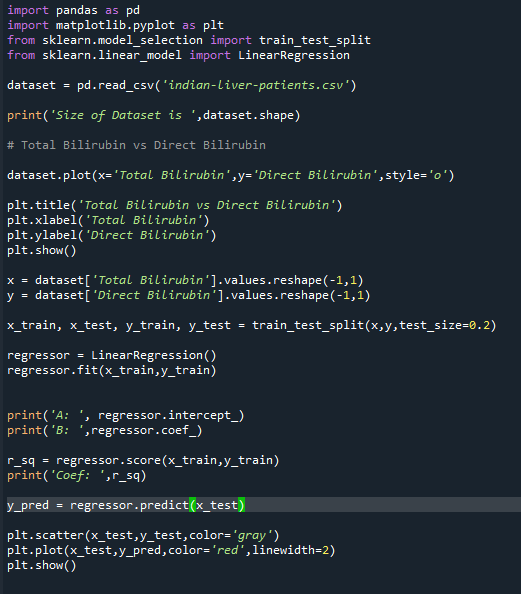
**Part 2 – Simple Linear Regression**

1. First of all we will compare Total Bilirubin vs Direct Bilirubin in each patient and draw it in a graph plot

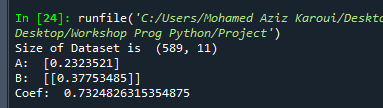


****

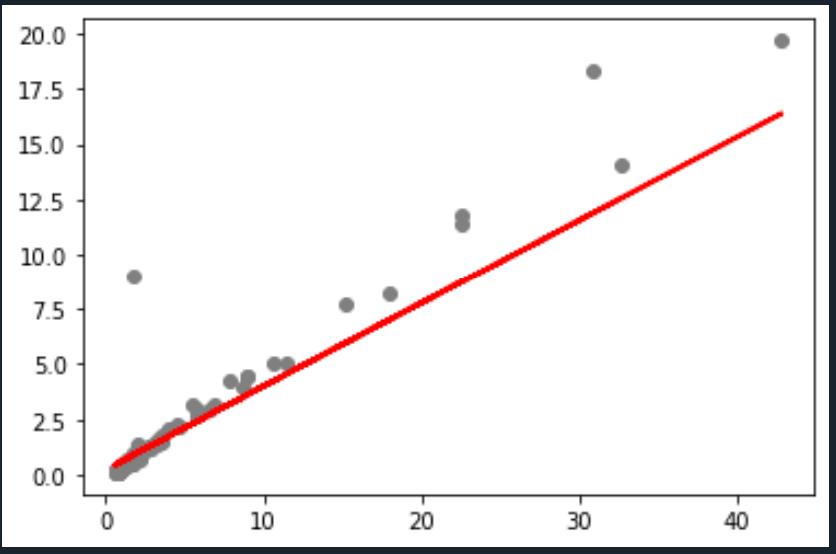
1. Then we will start doing the linear regression and this is the full script



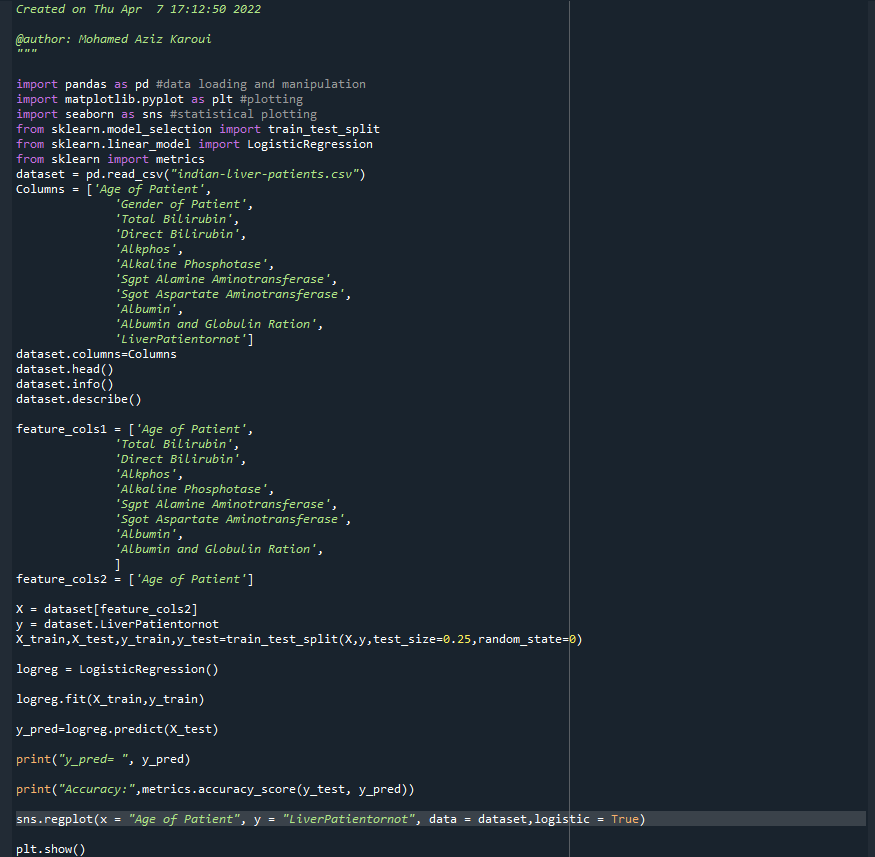
1. And we will show the slope and ordinate A and B then the Coefficient before finally show the final plot

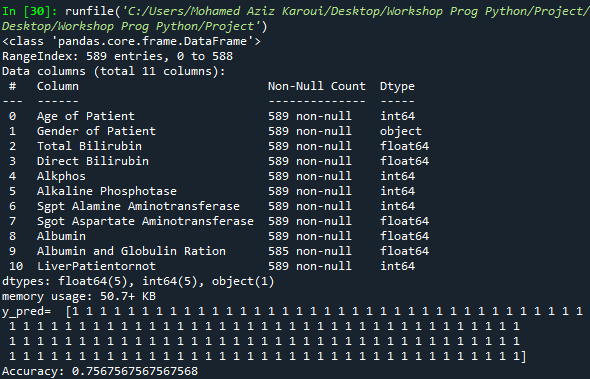


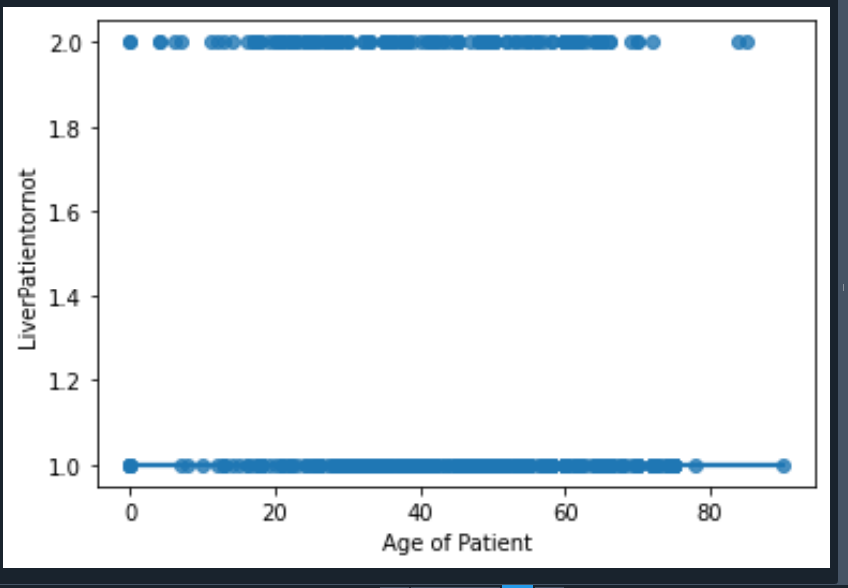
1. The Final Result Plot of the Simple Linear Regression :



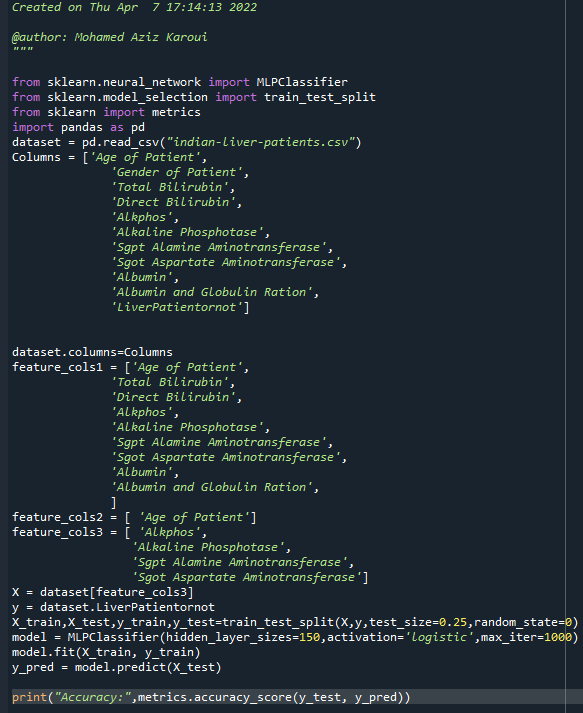
**Part 3 – Binary Logistic Regression**







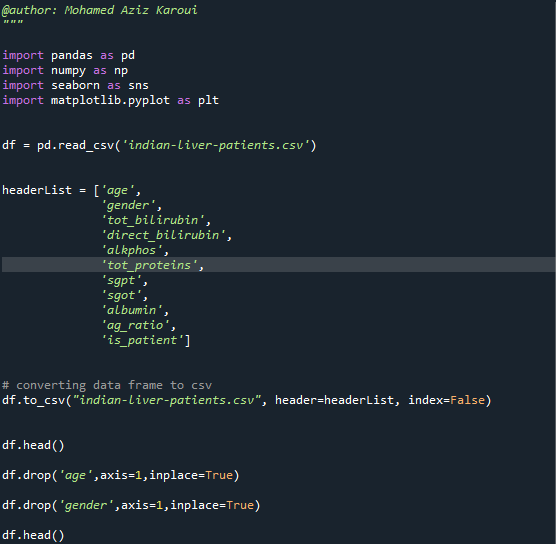
**Part 4 – Artificial Neuron Network (MLP – Multi Layer Perceptron)**

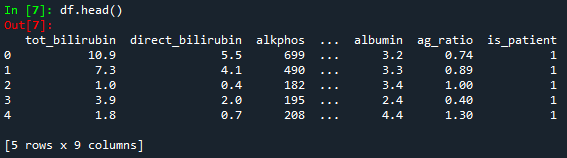
****

****

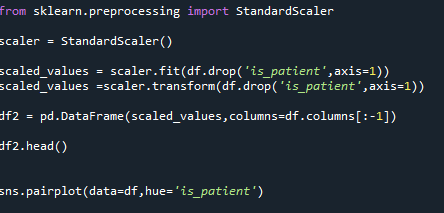
**Part 5 – K-NN (Nearest Neighbor)**

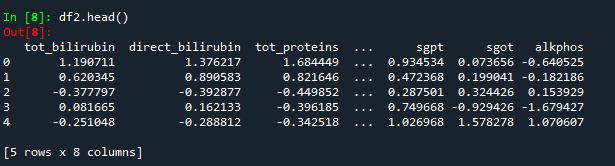
First we prepare the Data and remove age and gender

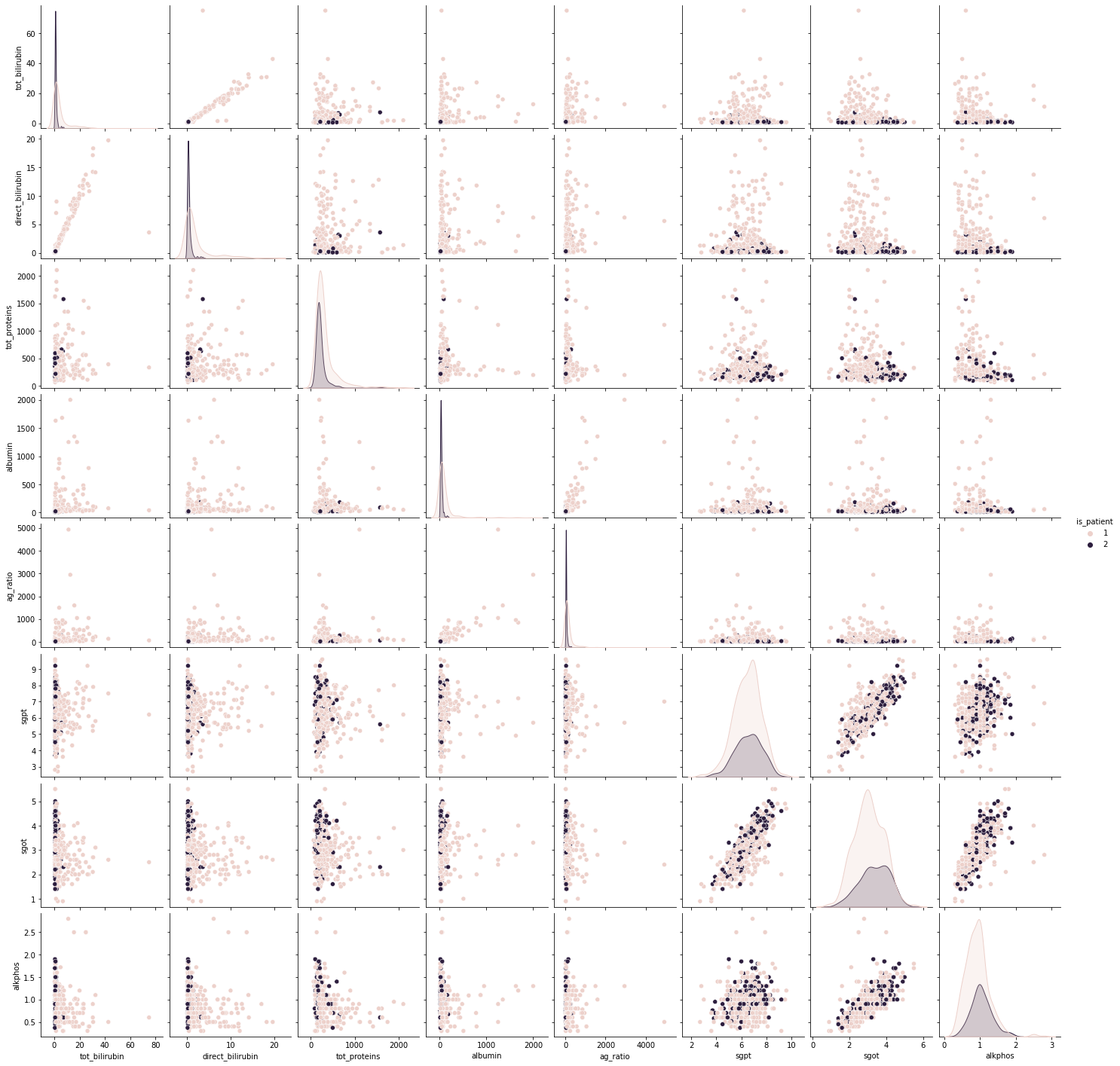




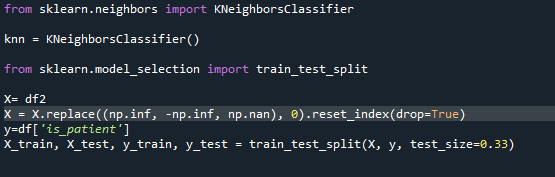
Then next we will standardize the data with mean = 0 s.d = 1 and then creating the data frame from the scaled values



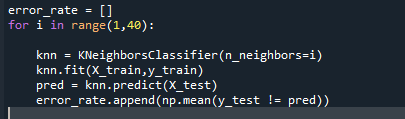


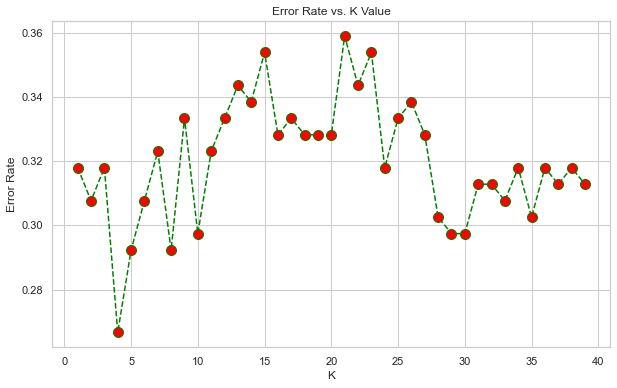


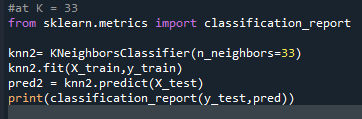
Then we import train\_test\_split for splitting dataset and use it to split our data X and Y and we use X.replace to change all NaN values to 0 so we can proceed with our tests



Then we check for K = 1 to 40 the error rate



Then we plot the error graph



And Finally with this script we can check the accuracy using classification report

