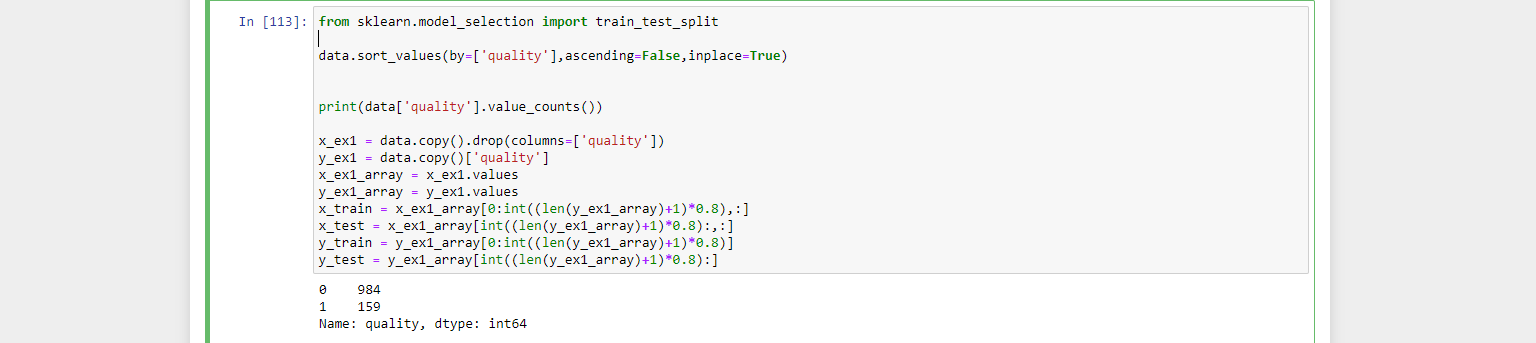
Introduction

This dataset is related to red variants of the Portuguese "Vinho Verde" wine. The dataset describes the amount of various chemicals present in wine and their effect on it's quality.

Our task is to find issues and suggest solution in the analysis and prediction of our codes.

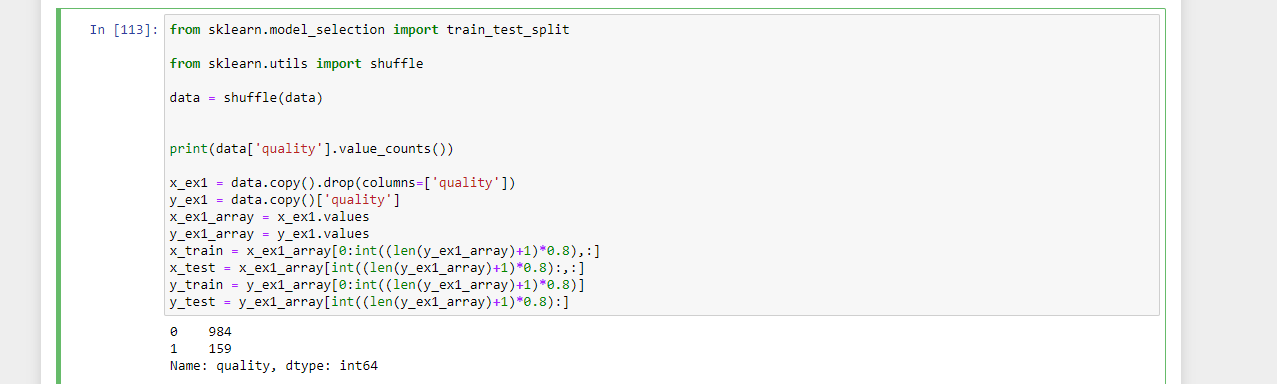
1.

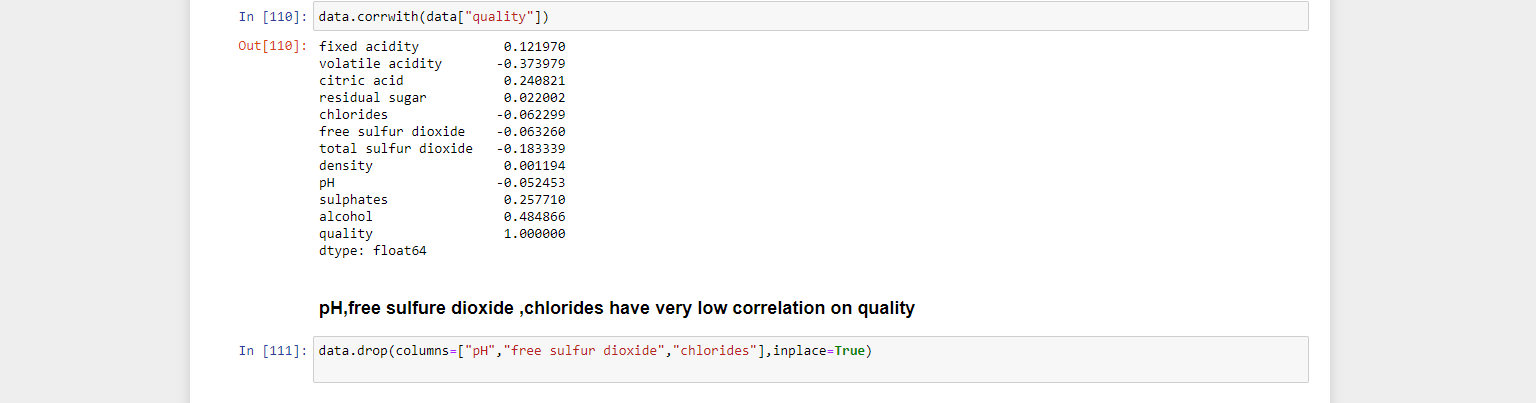


in the example above, data is sorted descending order based on target values.

sorting the rows according to the target value is a huge mistake, when we split the data 80 % for train and 20 % for test for example, the model will end up training part of the data more the other based on the sorted values.

Feeding the data sorted values will make the model train more on only part of the dataset(part of the target values) and the same for testing we will test only part of the dataset.

So instead we have to shuffle the data to make the training equally on all the target values and avoid bias for one target.

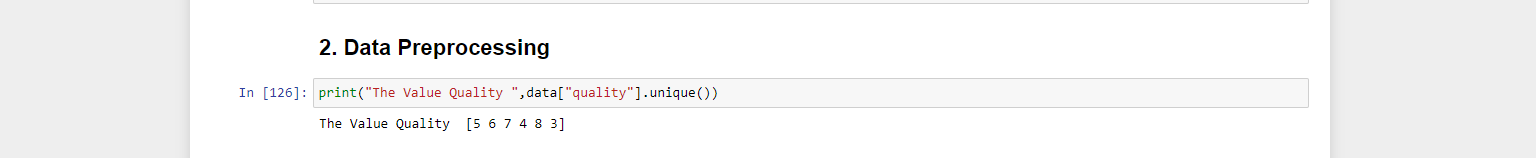
2.

In the code above we will see the correlation between all the features with the target value

as we can see pH, free sulfur dioxide, chlorides have very low correlation on quality.  
removing low correlation feature will help the model to focus on the most important feature during

training, as you know when the features of a model increase the harder the training of the model.

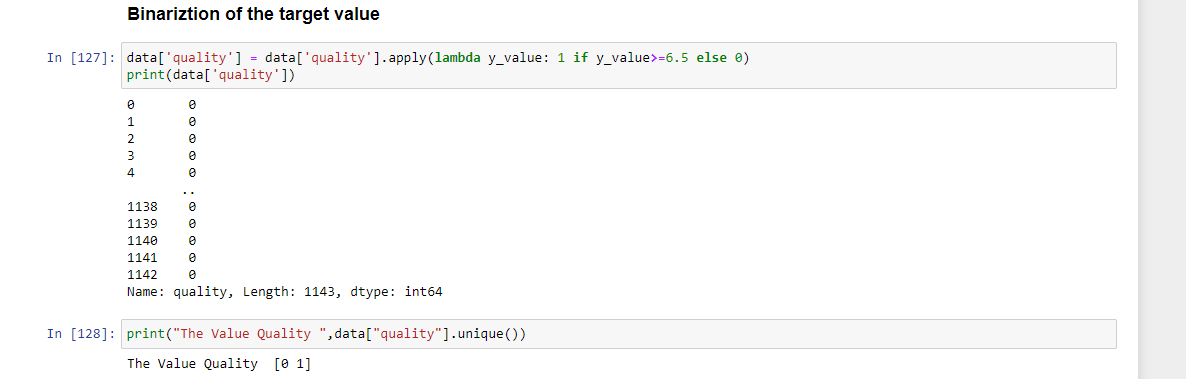
3.



as we can see we have 6 target variable for identifying the quality of the wine.

We can represent the 6 variable by only 2 variable only where if the target values is 7, 8 that will means the quality is good and for the values 3,4,5,6 that will means the quality is bad.

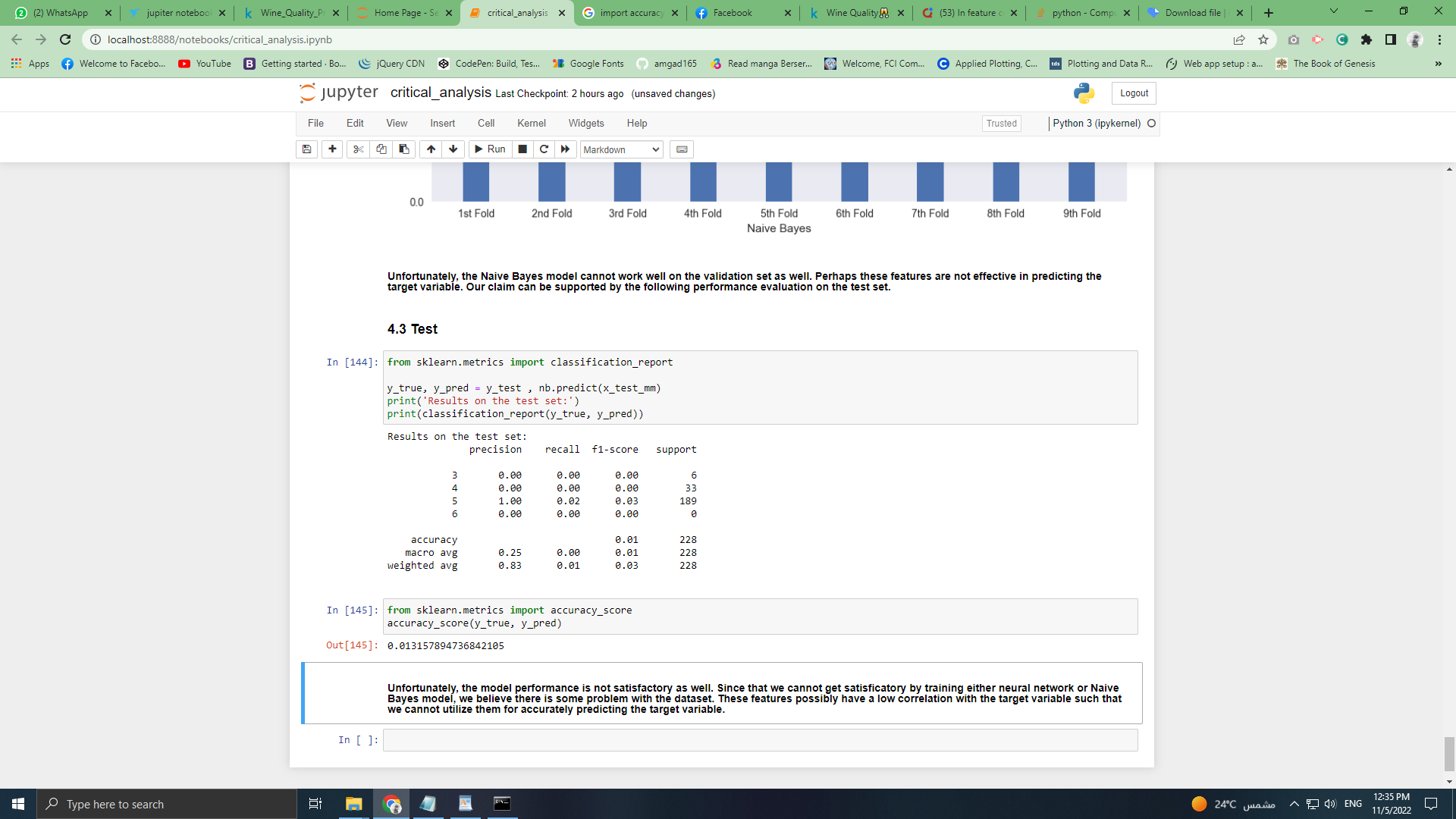
dealing with many target values affects the model accuracy too bad as our aim for the model is to identify the quality of the wine so converting our classification problem from multi-classification to binary classification is really good decision and has a great influence on the model accuracy.



Result

finally difference in accuracy

before:



After:

