**DECISION TREE**

In implementing and analyzing of Machine learning on our dataset part, firstly we used Decision Tree Algorithm. The aim of the using Decision algorithm is defined like this ; "The idea of a[decision tree](https://www.mindtools.com/dectree.html) is to divide the data set into smaller data sets based on the descriptive features until you reach a small enough set that contains data points that fall under one label."(What Is A Decision Tree Algorithm?, 2017, September 3). For this purpose, we shrink our data set based on the best four leagues in Europe and additionally we kept Turkey’s data too. You can reach this newly created dataset from here. We removed the unnecessary columns such as Sports, Country, Year, Season, Team, Games ,HomeRatio, AwayGoalsDiff. We manually (using Excel functions) add columns that will meet our demands such as HomeWinRatio, HomeLossRatio, HomeDrawRatio, AwayWinRatio, AwayLossRatio, AwayDrawRatio, TotalPoint, SuccessRate. We removed Country column but we created new column that is called CountryID to give ID’s to these 5 columns. At the below you can see the ID’s of that countries. Under showing ID’s you can see how we calculated the columns that we added newly.

England 🡪 1

France 🡪 2

Italy 🡪 3

Turkey 🡪 4

Spain 🡪 5

|  |  |
| --- | --- |
| **Column** | **How to Calculate** |
| HomeWinRatio | HomeWin/HomeWin+HomeDraw+HomeLoss |
| HomeLossRatio | HomeLoss/HomeWin+HomeDraw+HomeLoss |
| HomeDrawRatio | HomeDraw/HomeWin+HomeDraw+HomeLoss |
| AwayWinRatio | AwayWin/AwayWin+AwayDraw+AwayLoss |
| AwayLossRatio | AwayLoss/AwayWin+AwayDraw+AwayLoss |
| AwayDrawRatio | AwayDraw/AwayWin+AwayDraw+AwayLoss |
| TotalPoint | 3\*(HomeWin+AwayWin)+1\* (HomeDraw+AwayWin) |
| HomePoints | 3\*HomeWin+1\*HomeDraw |
| AwayPoints | 3\*AwayWin+1\*AwayDraw |
| SuccessRate | If **(**(HomeWinRatio > AwayWinRatio) && (HomePoints>AwayPoints )  return 1;  Else  return 0; |

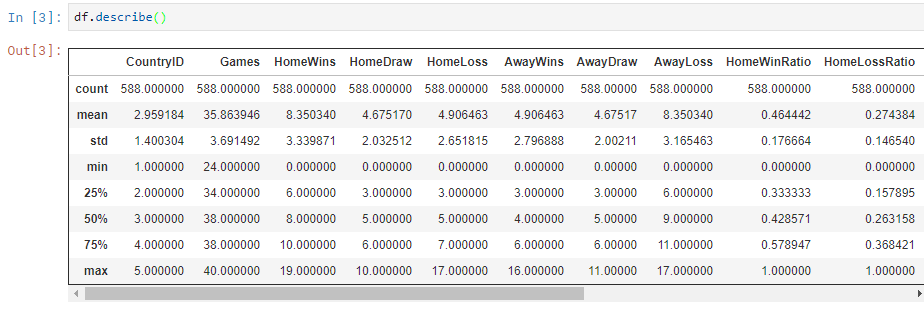
We coded in JupyterLab . First lets look which libraries and modules we used ;

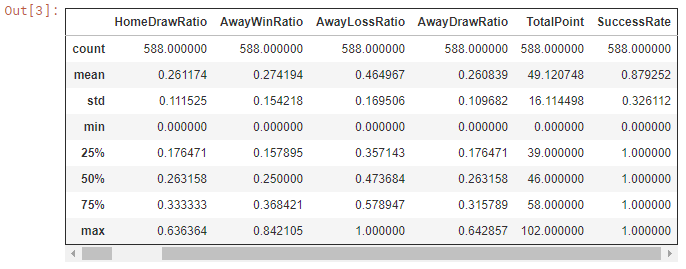


Our first part is reading our dataset and naming it as “df”.



With describe function we get a summary of our dataset. And our results at the below.





 Because of the aim of the target is predicting the value of the target this is why first we are dropping our column “Success Rate” and then labeling that column as our target.

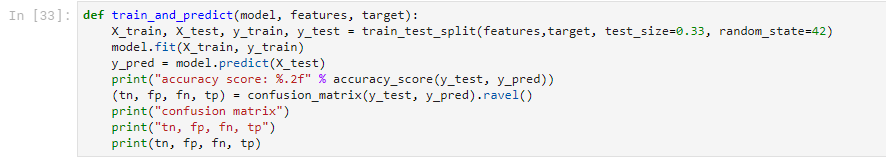


We are creating two Model. First of them is named as “naive\_model” that takes number\_state=42 that means 42 is used by random number generator.

Second model is “param\_model” that takes the parameter below and these parameters mean the maximum depth of tree will be 5, minimum 10 samples required to split an internal node,minimum 10 samples required to be at a leaf node and again 42 is used by random number generator.

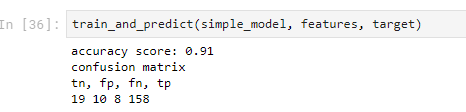


Then we used the algorithm at the below to train and predict the result on our target (SuccessRate).

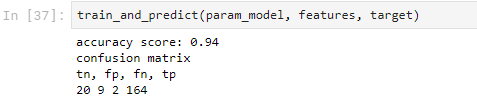


We used this algorithm on our both models.

Let’s see the results on “simple\_model”



Let’s see the results on “param\_model”



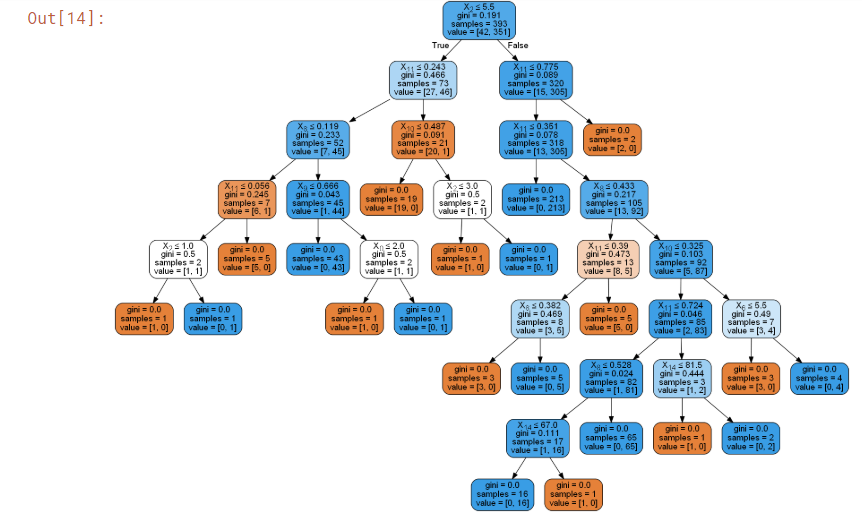
Dot\_data is used for string representation of the input tree in GraphViz dot format and this will export a tree that designed with the parameters below.



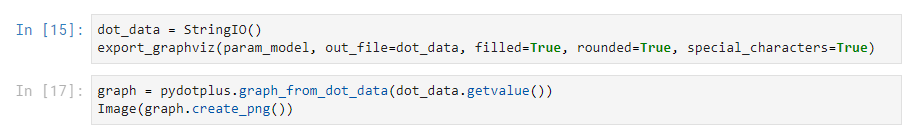
With the code at below, it will make a graph of the dot\_data and create an image of it.



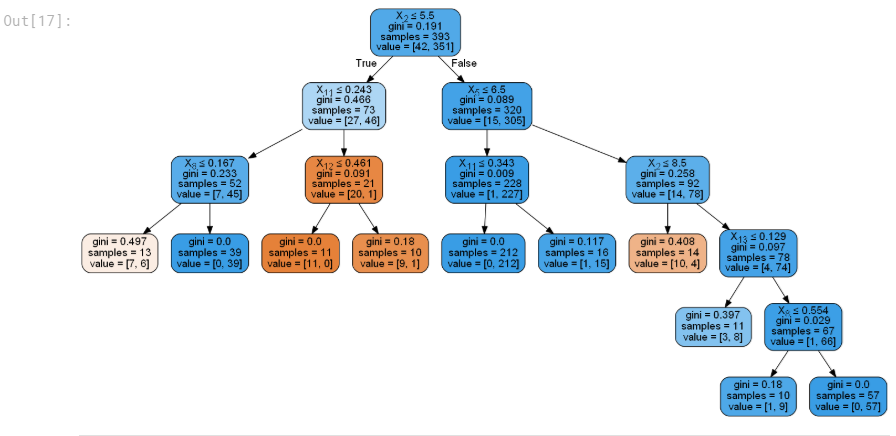
Now let’s the decision tree based on simple\_model



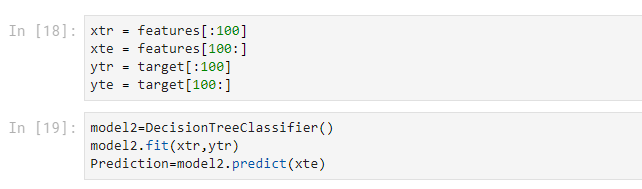
Now let’s do the same process to see our decision tree based on “param\_model”



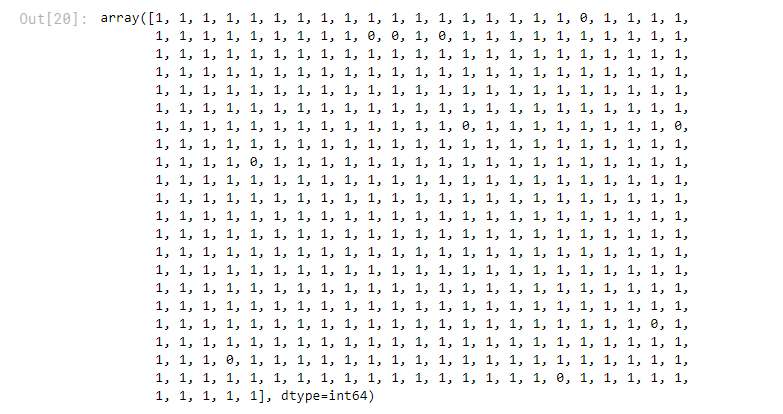
Our decision tree as an image based on param\_model



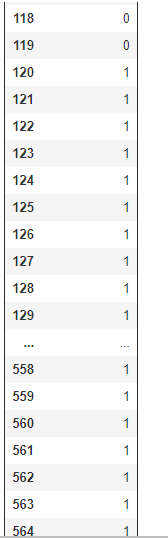
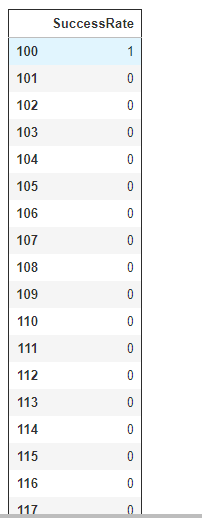
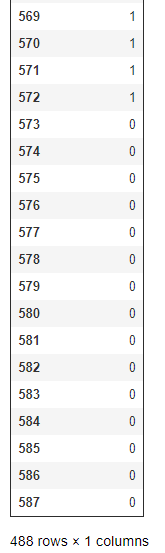
For each pair of our df features, the decision tree learns decision boundaries made of combinations of simple thresholding rules inferred from the training samples



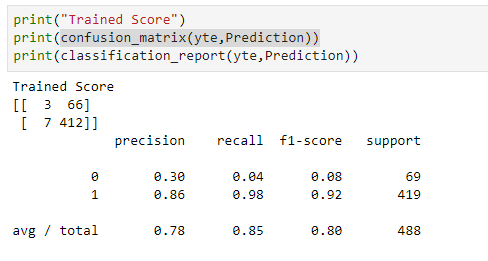
Now let’s see the prediction on our features



Now let’s see the some predictions on our target;

Now we will calculate our confusion matrix that is used to evaluate the accuracy of a classification.



REFERENCES:

What Is A Decision Tree? , Retrieved *(*2017, September 3). from [https://medium.com/@SeattleDataGuy/what-is-a-decision-tree-algorithm-4531749d2a17 (2018](https://medium.com/@SeattleDataGuy/what-is-a-decision-tree-algorithm-4531749d2a17%20(2018), May 14)