**Report**

***Objective***

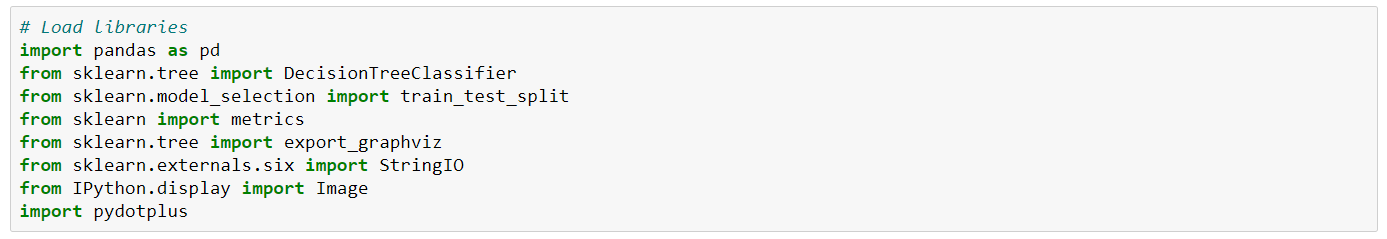
In this project, I have a dataset containing features extracted by a deep network on a 3-class classification problem. Each record corresponds to features generated for a sample. My objective is to build a decision tree model using these features. Your decision tree will be evaluated by its accuracy on the test set.

***Dataset***

I have a csv file named data.csv. This dataset contains 2048 features for each record, with their labels in the last column indicating the class (“0”, “1” or “‘2”). There are a total of 2060 records.

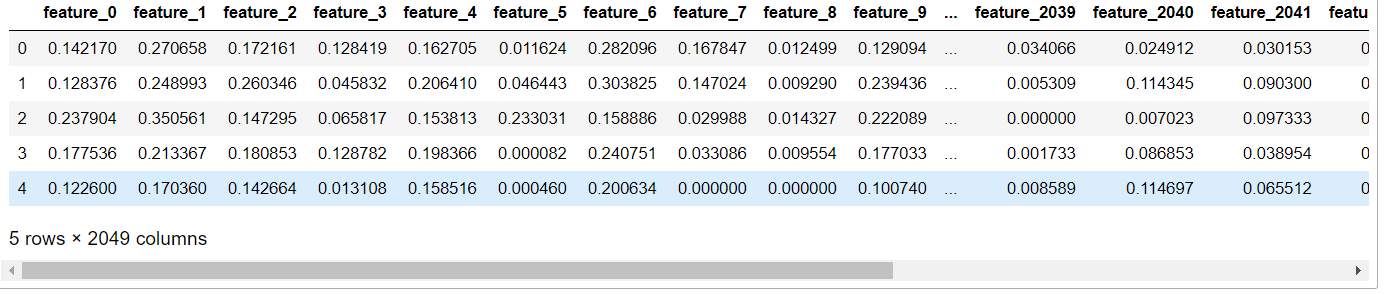
1. In Python, sklearn is the package which contains all the required packages to implement Machine learning algorithm. I will use DecisionTreeClassifier.

First, I import all the necessary modules.

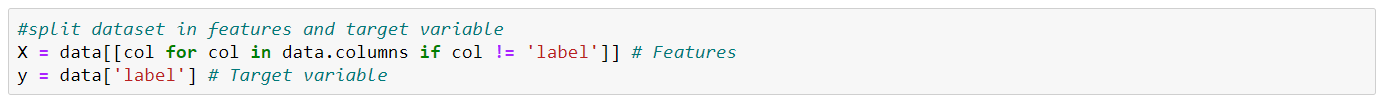


1. Now, I load the dataset using pandas' read CSV function.

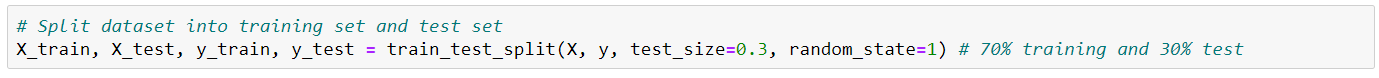




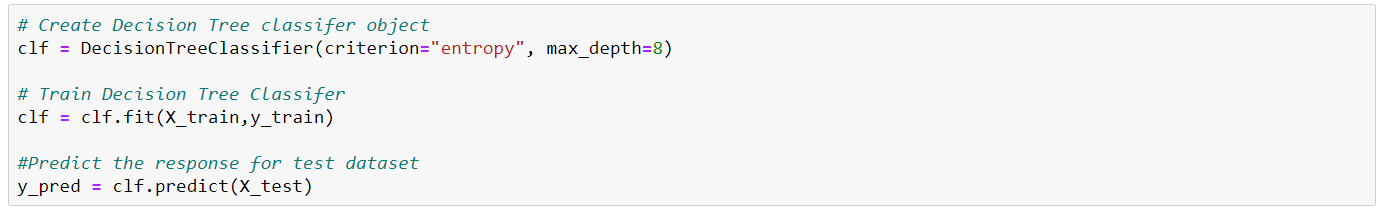
1. Here, I need to divide given columns into two types of variables dependent(or target variable) and independent variable(or feature variables).



1. To understand model performance, dividing the dataset into a training set and a test set is a good strategy. Let's split the dataset by using function train\_test\_split(). I need to pass 3 parameters features, target, and test\_set size.



1. I create a Decision Tree Model using Scikit-learn.



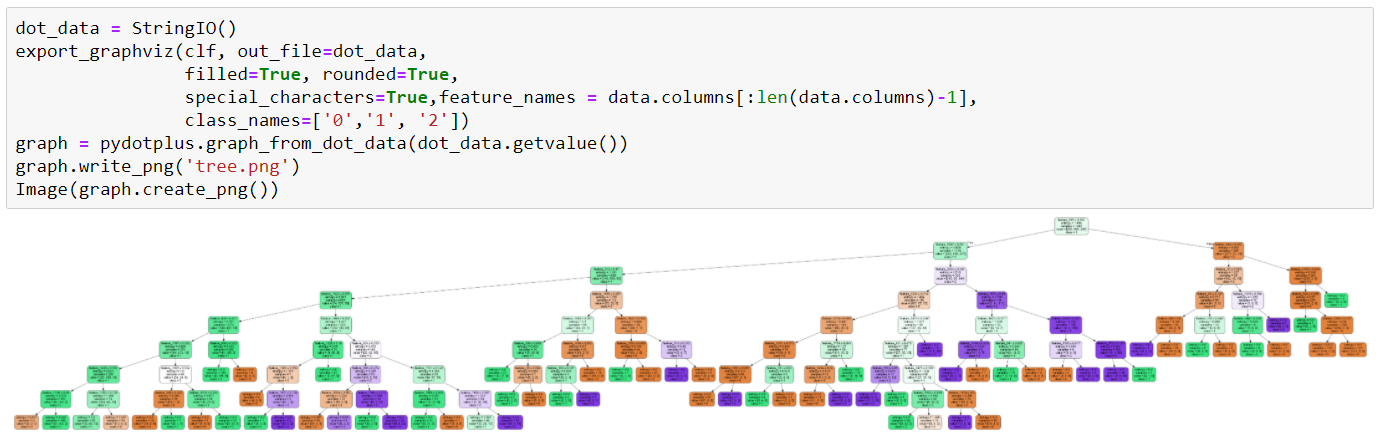
1. I estimate, how accurately the classifier or model can predict labels.

Accuracy can be computed by comparing actual test set values and predicted values.



Well, I got a classification rate of 77.83%, considered as good accuracy.

1. I can use Scikit-learn's export\_graphviz function for display the tree within a Jupyter notebook. For plotting tree, you also need to install graphviz and pydotplus.



In the decision tree chart, each internal node has a decision rule that splits the data. Gini referred as Gini ratio, which measures the impurity of the node. I can say a node is pure when all of its records belong to the same class, such nodes known as the leaf node.