

Faculty of Computer Engineering

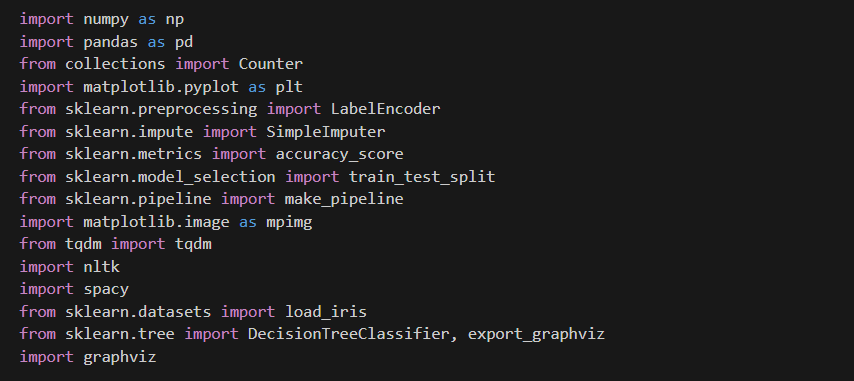
First Project (decision tree)

Artificial Intelligence Lesson

Pooria Rahimi - 99521289

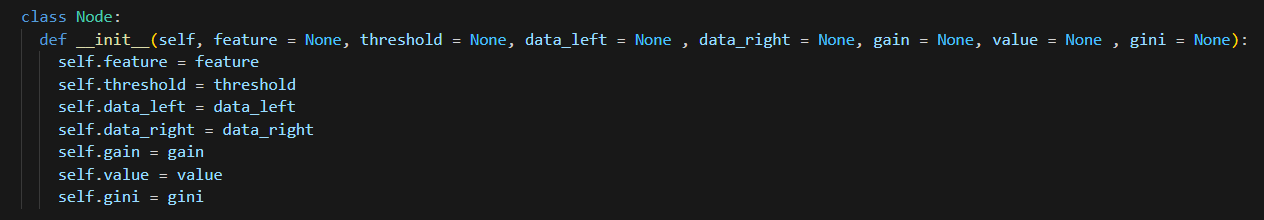
1403-1402

* At first, to do this project, we need a series of libraries that we have to import first, which is according to the picture below:



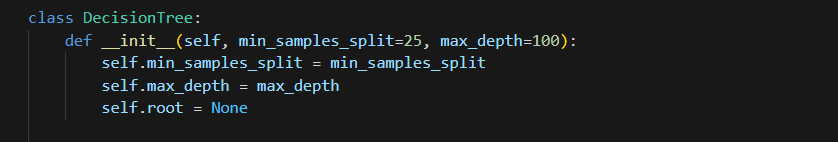
Step1 : Then we form the node class.

* The node class represents each node in the tree, which can take different values ​​depending on the location of that tree (leaf or middle node).



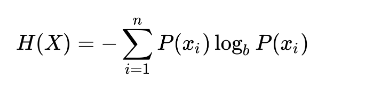
Step2 : Then it is time to implement the decision tree class.

* The decision tree class includes the general features of the tree and its different functions.

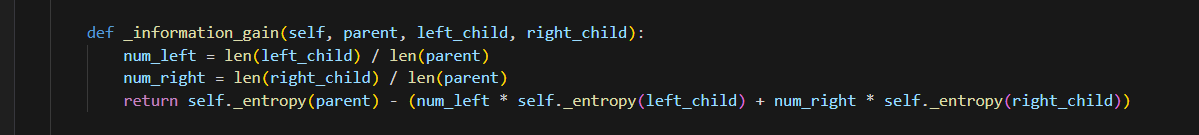


Step3 : Then it is time to implement the Entropy function.

* Entropy is a method of decision tree class which is implemented according to the following formula:



* The information gain method also calculates the gain of each node

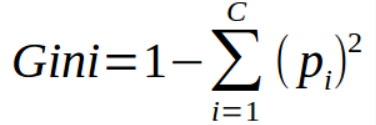


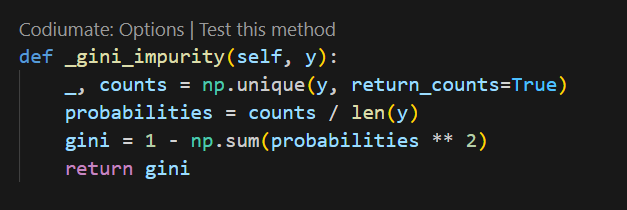
Step4 : Now, in this section, we will implement the main part of our code.

* In the main part of our program, we implement the best splite method.
* In this method, we intend to do the best division among the children.
* The working process of this function is that we separate children according to different thresholds defined for each feature and calculate information gain and gini index for each separation in order to make the best choice according to these parameters. Let's get in isolation.
* It should be noted that this function simulates the importance function in slides, which is as follows:

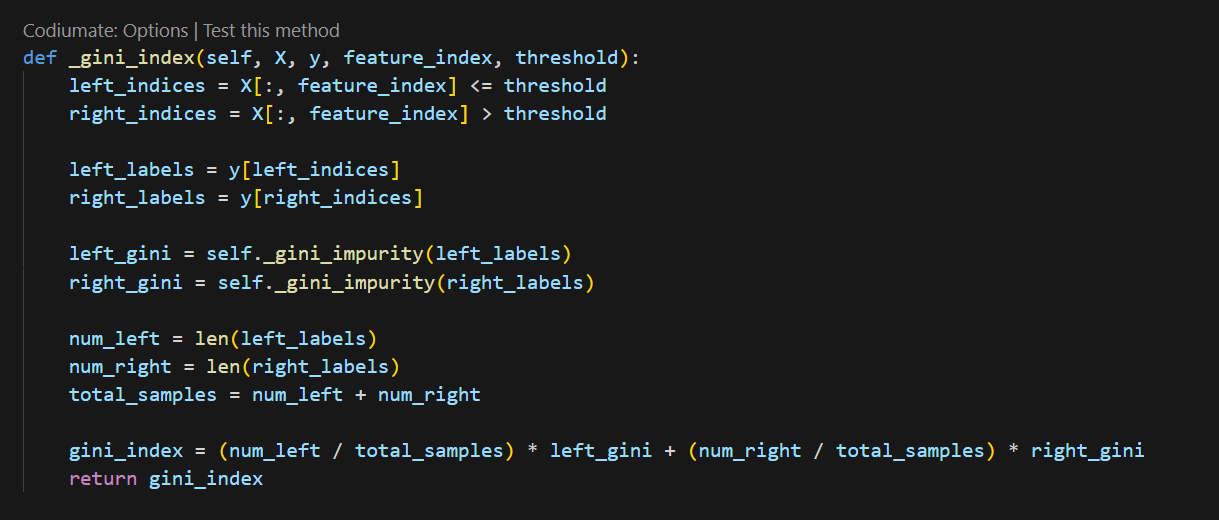


* The gini impurity function calculates the gini value according to the following formula:

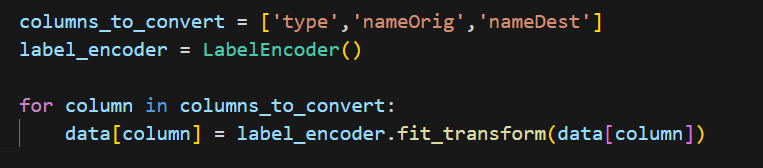


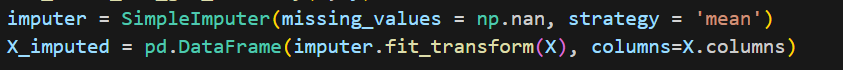


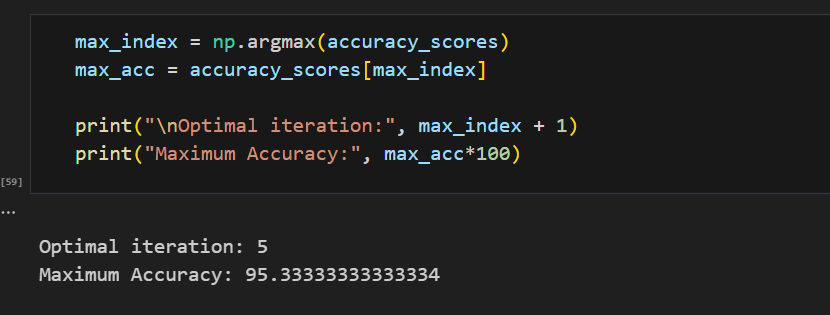
* The gini\_index method multiplies the probability of data occurrence in gini obtained in gini\_impurity, which is as follows:

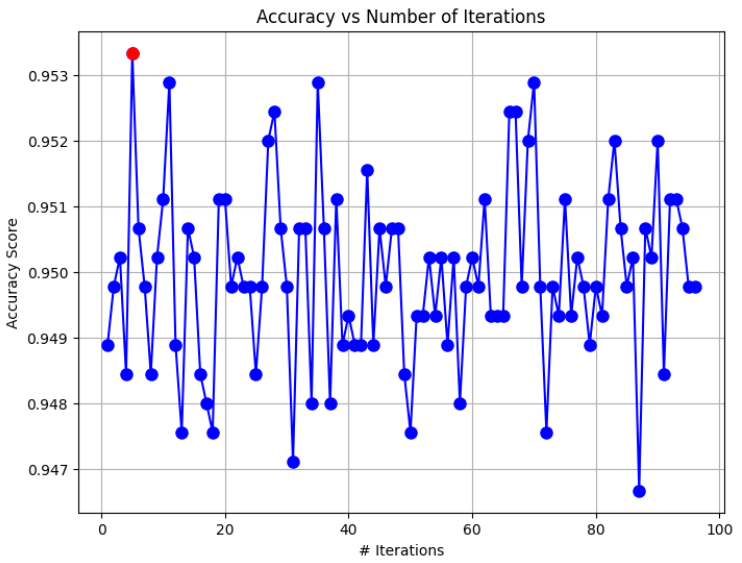


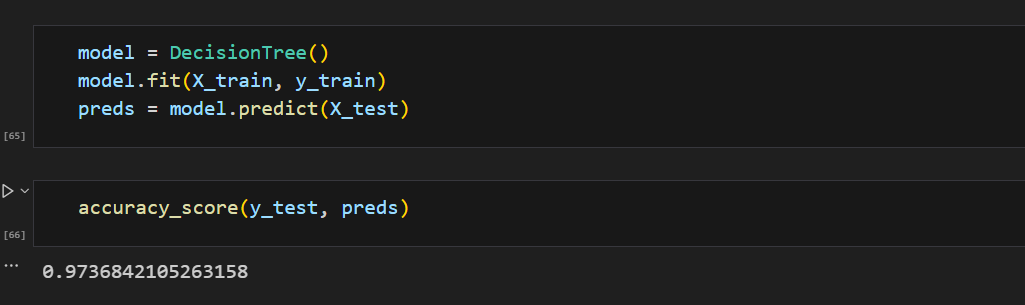
* Then, in the function implemented below, we try to normalize the data:
* For example, in this part, we label the dataset data:



* Now, we may have none data or empty data. In this section, to solve this problem, we put the average amount of available data for this feature, which is as follows:
* And finally, to test the designed model, we compare it with the model implemented in sikit learn, and the accuracy percentage of our function and the ready decision tree is as follows, and we also used the accuracy\_score function to test the accuracy, and the accuracy of the model You can see the designed and ready model below:







The End.