Introduction

Hello My name is Inzamamul Alam Munna. I am a global intern at hash Analytics. I am from Bangladesh The land of nature. I am a Post Graduate Student of University of Dhaka Department of CSE. I am a Machine Learning Researcher in my university. Also I am a Collaborator at Facebook Developer Circle (Dhaka). I have a vast experience in Data Analyst and Machine Learning. I have few work experience in some familiar company.   
  
Today our Work at Assignment -04 in hash analytics is Understand the decision tree algorithm and create a DataFrame by importing a new IRIS and make prediction

Explanation

1. I first download the datasets from the website.

1. **Number 1-7** Line I actually import the important module for this assignment.
2. **Number 8** Line I read the dataset using pandas. datasets is mainly **.csv** format. And indexing the dataframe according to **Id Column.**
3. **Number 9** Line I rename the Id to **Null** because of accessing easier afterwards.
4. **Number 10** Line I select the features attribute by calling their **Column Names**.
5. **Number 11** Line Select the target column which should I predict and also based on train.
6. **Number 12** Line I need to processing the target variable so using **get\_dummies** because of not using **one hot encoder.** Because It cannot process string values directly. If your nominal features are strings, then you need to first map them into integers.

But, **pandas.get\_dummies** is kind of the opposite. By default, it only converts string columns into one-hot representation, unless columns are specified. So choosing **pandas.get\_dummies** is the right option here.

1. **Number 13** Line I split the test and train data using the library function and make it train size = 0.7 and test size = 0.3. and give the random state parameter because of not getting different results.
2. **Number 14-15** Line I make an object of Decision Tree Classifier by library function and put the criterion as entropy and give the max depth, max features and min sample leaf value for better prediction. Random state makes by running the code multiple time cannot change the result.
3. **Number 16** Line Predict the test data which we make.
4. **Number 17-27** Line I print the accuracy, confusion matrix, and classification report for my model which I predicted in **Iris** Dataset.

Result:

I got **96%** train accuracy and 95.5% Test accuracy in my model. I can increase my accuracy by normalize the value of the attribute because they are highly co-related. I know but I stuck at because of my family problem My father’s elder sister Husband (Fufa we called in our language) has died so I go to my village so for early submission I cannot make it happen. Sorry for the Inconvenience.