1. Induce a decision tree from the wine dataset.

Use 60% of the dataset for training and 40% for testing (use random\_state = 0). Display the resulting tree and compute the accuracy and confusion matrix for this model.

2. Induce decision trees using just a subset of the features from the dataset. Specifically, use the first 5, middle 4, and last 4 features in the dataset. Determine which has the best accuracy and present the resulting tree, accuracy, and confusion matrix for the best model.

Submit the jupyter notebook for this assignment.

—--

1. Induce a decision tree from the wine dataset ([wine.csv)](https://ateneo.instructure.com/courses/34185/files/6415993?wrap=1)
2. [Download wine.csv)](https://ateneo.instructure.com/courses/34185/files/6415993/download?download_frd=1)
3. Use 60% of the dataset for training and 40% for testing (use random\_state = 0). Display the resulting tree and compute the accuracy and confusion matrix for this model.
4. Induce decision trees using just a subset of the features from the dataset. Specifically, use the first 5, middle 4, and last 4 features in the dataset. Determine which has the best accuracy and present the resulting tree, accuracy, and confusion matrix for the best model.

Submit the jupyter notebook/s for this assignment.

You may begin with the following jupyter notebook (loads and splits wine dataset): [dt\_wine\_blank.ipynb](https://ateneo.instructure.com/courses/34185/files/6422990?wrap=1)

[Download dt\_wine\_blank.ipynb](https://ateneo.instructure.com/courses/34185/files/6422990/download?download_frd=1)

You may refer to the following:

* [DecisionTree\_iris.ipynb](https://ateneo.instructure.com/courses/34185/files/6422316?wrap=1)
* [Download DecisionTree\_iris.ipynb](https://ateneo.instructure.com/courses/34185/files/6422316/download?download_frd=1)
* [04b - Trimming Decision Tree.pdf](https://ateneo.instructure.com/courses/34185/files/6423077?wrap=1)
* [Download 04b - Trimming Decision Tree.pdf](https://ateneo.instructure.com/courses/34185/files/6423077/download?download_frd=1)
* [05 - Classification with Scikit Learn.pdf](https://ateneo.instructure.com/courses/34185/files/6423080?wrap=1)

One of the feature names in the dataset is a bit long, so start with the following code, for convenience:

# Load and pre-process wine dataset  
wine\_full\_df = pd.read\_csv("wine.csv")  
wine\_full\_df.rename(columns={'OD280/OD315 of diluted wines': 'Wine Dilution'}, inplace=True)  
wine\_classes = ["1", "2", "3"] # used for plot

wine\_df = wine\_full\_df.drop(columns=['Wine Variety']) # removing the wine label from the dataframe  
wine\_y = wine\_full\_df['Wine Variety'] # keeping only the wine label column

Submit source code or jupyter notebook(s) for this assignment.