I have used the Teaching Assistant Evaluation data set which can be found at- https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation

1. The following modules are required to run my code-
pandas
numpy
matplotlib
pypot
graphviz (Not a python module)
tree
metrics
Urllib

2. After install the modules, run the code on Github which functions as follows-
a. The dataset is loaded using the URL.
b. splitting of dataset is done to form the test and train data
c. Decision tree classifier is formed using Gini criteria (with depth 1,2,3,4,5,6)
d. Accuracy is calculated for point c using accuracy_score for depth 1,2,3,4,5. Here, it is tested with train and test both.
e.Tree is visualised for depth 1,2 and 3 using graphivz.

With increase in depth from 1 to 6, accuracy goes up for both test and train data. It shows that as the depth increases, the model becomes more accurate.

Performance of the train set changes as a function of depth as depth goes up, performances also goes up. At a point, it will start falling down which is because of overfitting.

Performance of the test set changes as a function of depth as depth goes up, performance goes up but at a point it becomes constant and then increases again. (underfitting)

Link to Github -https://github.com/ShomronJacob/CS595