CSE 375/381 Lab 5: Decision Trees

You are provided with the "Playing Tennis" dataset. You are required to carry out the following tasks:

1. Load the data and perform basic data pre-processing needed the classification procedure. What is the appropriate data preprocessing task that is required and why?
2. Split the data to 70%/30% training/test sets.
3. Using scikit-learn, train a decision tree with the proper parameters (entropy should be used to select the splitting features)
4. Using the test dataset, calculate the accuracy
5. Use other classification evaluation metrics to assess the results
6. Draw the produced tree using the proper module in scikit-learn
7. Repeat steps 3-6 using the Gini index for splitting