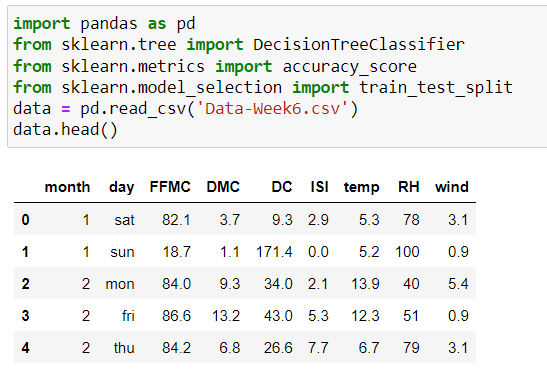
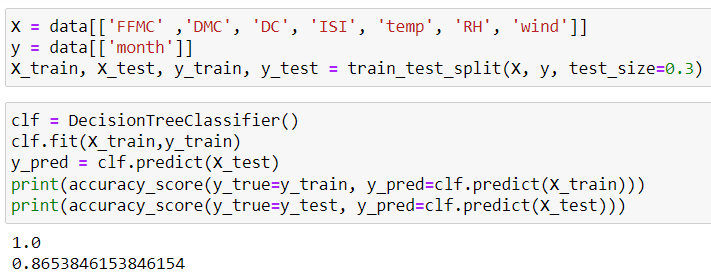
**Name:** Aravind Kumar Kaspe **Banner ID:** 001291145

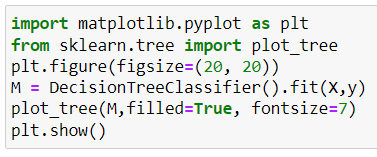
**ASSIGNMENT 5**

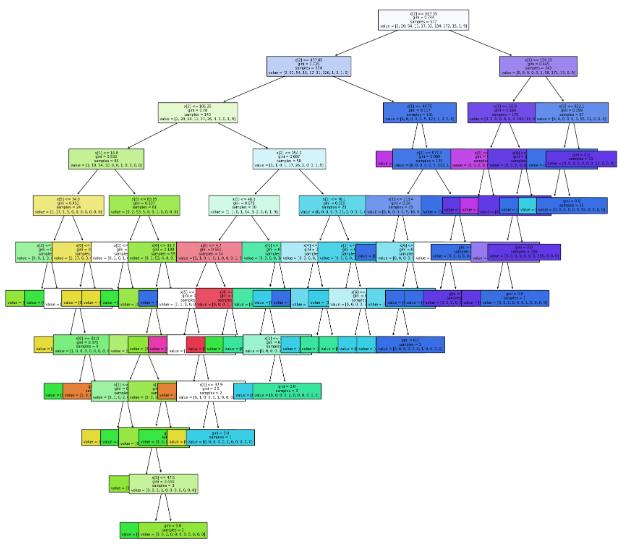
**CSCI 5930 – Homework 4: Decision Tree**

1. **Make the decision tree for Forest Fire data.  
   X: FMC DMC DC ISI temp RH wind  
   Y: month**

****

****

****

****

The highest readings of fire risk indicators such as FFMC, DMC, DC, and ISI are recorded during the most severe fire months. The simultaneous presence of hot, dry, windy conditions, coupled with more dry vegetation allowed for the intensity of fires in those days.

1. **Interpret the decision tree based on the Gini of each node.**

A decision tree visually represents a chain of decisions made based on the features of the data to categorize the data points.  
  
Initial Step: The first one will be the DC, namely, whether it is lower than or equal to 667.35. When we look at this Gini coefficient, it should be 0.744 to show a large variety of classes.  
  
Subsequent Decisions: In the left path, the next decision to be made depends on whether DMC is (< = 477.85). In embryo 4 the Gini score is 0.725 meaning that guys who have a higher impact on the chosen strategy could be considered as less differentiated in terms of their strategy. On the right side of the tree, the complete node of DC is reprocessed and connected with the cut-point (<=156,25), followed by the Gini coefficient of 0.445.  
Further Exploration: It is furthered downwards where other factors like wind and ISI are used in decision-making. The Gini impurity measure of each decision node is associated with the split quality, that is, the degree of class separation involved.  
  
Final Outcomes: A solid color histogram bar explains the probability of category at the bottom bar. As the color gets purer, the gini-impurity inherent in the color decreases, or the other way around.

1. **Find the month with the highest forest fires? Could you find the reason from Tree?**

The number of forest fires is recorded at the highest in August with 184 forest fires observed. Such information is according to the decision tree analysis. The decision tree is later found to display August as a consistently high value at the initial node and this high value trickles down the subsequent nodes in the decision tree. The prominence of the month can be owed to its portrayal in many of the nodes which are in turn this is a result of the different properties.