Graded Questions (Tree Models):

1. Truncation and Pruning

**Graded Questions**

Please download the training data set and validation data set that are attached, to answer the following questions.

**[Training Set](https://cdn.upgrad.com/UpGrad/temp/50d07e36-7382-401a-9de8-ffbfdd464654/Training_Set.csv" \o "Training_Set.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/50d07e36-7382-401a-9de8-ffbfdd464654/Training_Set.csv" \o "Training_Set.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/50d07e36-7382-401a-9de8-ffbfdd464654/Training_Set.csv" \o "Training_Set.csv" \t "_blank)**

**Training Set -** The first two columns in the training set file represent the two attributes, Attribute 1 (a1) and Attribute 2 (a2). The third column (column C) is the class label associated with each data point in the training set. There are two class labels: 1 and 2. In total, there are 30 data points/observations in the training set. You need to build a decision tree using this file. Use all 30 points for training. Construct the decision tree step-by-step as the following questions suggest.

**[Validation Set](https://cdn.upgrad.com/UpGrad/temp/7ce4563e-24e8-45a1-8f2d-f6a844caa747/Validation_Set.csv" \o "Validation_Set.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/7ce4563e-24e8-45a1-8f2d-f6a844caa747/Validation_Set.csv" \o "Validation_Set.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/7ce4563e-24e8-45a1-8f2d-f6a844caa747/Validation_Set.csv" \o "Validation_Set.csv" \t "_blank)**

**Validation Set** - The first two columns in the validation set file represent the two attributes, Attribute 1 (a1) and Attribute 2 (a2). The third column (column C) is the class label associated with each data point in the validation set. The fourth column is the class label as predicted by the decision tree model. There are two class labels: 1 and 2. In total, there are 4 data points/observations in the validation set. This data file will be needed to check the accuracy of pruned trees.

Please note that you mustn't use DecisionTreeClassifier or write codes to solve the following questions. You can use Excel to split and filter the data set according to the test conditions specified in the questions. The DecisionTreeClassifier function uses a slightly different test on each node, and hence, the final answer to the following questions may not match.

Please use the following hyperparameters to solve the following questions:

* max\_depth = 20
* min\_samples\_split = 10
* min\_samples\_leaf = 5
* Homogeneity measure = gini

Note: **(13, 5)**, written on a node, implies that 13 data points belong to class label 1, and 5 data points belong to class label 2.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a social media post

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. Random Forest:

A screenshot of a social media post

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated