

# DSCI - 552: Machine Learning for Data Science

## Assignment - 4

### Decision Tree

Using the following dataset:

Holiday	Discount	Purchase
No	Yes	Yes
No	Yes	Yes
No	No	No
Yes	Yes	Yes
Yes	Yes	Yes
Yes	No	No
Yes	Yes	Yes
No	Yes	Yes
Yes	Yes	Yes
Yes	Yes	Yes
Yes	No	Yes
Yes	No	No
Yes	Yes	Yes
Yes	Yes	Yes
Yes	Yes	Yes
No	Yes	Yes
Yes	No	Yes
No	Yes	Yes
Yes	No	Yes
Yes	No	Yes

No	Yes	Yes
Yes	Yes	No
Yes	No	Yes
No	Yes	Yes
yes	No	No
No	No	No
No	Yes	Yes
No	Yes	Yes
yes	Yes	Yes
yes	Yes	Yes

Create a decision tree based on the "Discount" and "Holiday" variables to predict the value of "Purchase". Using weighted gini index as the splitting criteria and show the resulting leaf nodes for each branch of the tree. For the root and each node of the decision tree, calculate the gini index, sample size, and sample distribution.

Hint: Sample distribution = [a,b] where a is the number of "yes" of target value in the current sample, b is the number of "no" of target value in the current sample.