**Akintoye Adesomoju**

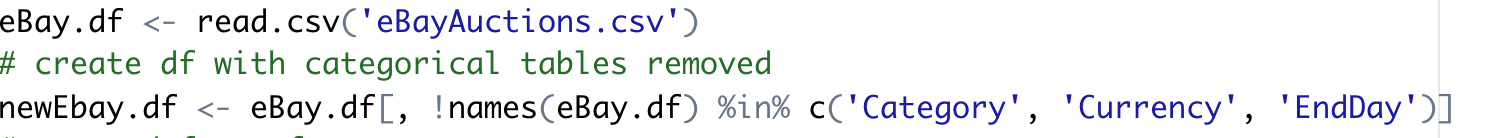
**CIS4930 Big Data Analytics**

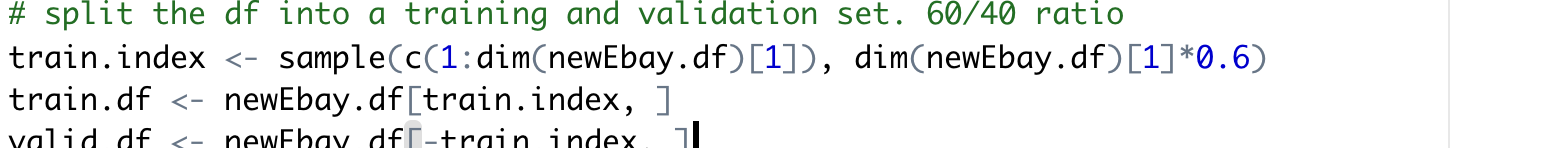
**Homework 4: Competitive Auctions on eBay.com**

Due by 7/10/2022 (Sunday), 11:59 pm

**Problems**

The file **eBayAuctions.csv** contains information on 1972 auctions that transacted on eBay.com during May-June in 2004. The goal is to use these data in order to build a model that will classify competitive auctions from non-competitive ones. A *competitive auction* is defined as an auction with at least 2 bids placed on the auctioned item. The data include variables that describe the auctioned item (auction category), the seller (their eBay rating), and the auction terms that the seller selected (auction duration, opening price, currency, day-of-week of auction close). In addition, we have the price that the auction closed at. The goal is to predict whether an auction will be competitive or not.

1. Import the dataset. Remove **Category**, **Currency**, **EndDay** variables from the imported dataset because we already have their corresponding dummy variables. 
2. Split the data into training and validation datasets using a 60%-40% ratio.



1. Fit a classification tree. Use **Competitive** as the target variable and the rest of the variables as predictors. (As mentioned in the notes, you don’t have to exclude one dummy variable from each dummy group for a categorical variable). To avoid overfitting, set the **maxdepth=6**.
   1. Report the tree - plot the tree and **copy and paste the resulting diagram**. You don’t have to care too much about the aesthetics of the diagram. Diagram

      Description automatically generated
   2. List the decision rules. For example, if variable1<0 AND variable2<2, class=0. *Calendar

      Description automatically generated with medium confidence*The decision rules are presented above, the most common (25%) being:   
      If OpenPrice < 2.4 & ClosePrice >= 2 then the model will report it as a competitive bid
   3. Report the prediction confusion matrix of validation data.

Text, table

Description automatically generated with medium confidence

* 1. Which predictors are used by the tree? Text

     Description automatically generated

1. Are the rules practical for predicting the outcome of a new auction? (Hint: Can you use the rules to classify a new auction before the auction ends? In other words, do you know the values of all predictors used in the rules before the auction ends? Some of them may not be known before the end of the auction. What are those variables?).

In short,which variables should **NOT** be included in the predictor set?Explain why.

In its current form, the rules are not practical in predicting the outcome of the new auction. The main reason is because we would not have the closing price as we had in the past. We would not be able to use that as a predictor

1. Fit another classification tree using the same setting in question 3. This time, use only the predictors that can be used for predicting the outcome of a new auction **before** the auction ends.
2. Report the tree - plot the tree and **copy and paste the resulting diagram**. You don’t have to care too much about the aesthetics of the diagram.

Diagram

Description automatically generated

1. List the decision rules. For example, if variable1<0 AND variable2<2, class=0. A picture containing calendar

   Description automatically generated
2. Report the prediction confusion matrix of validation data.

Table

Description automatically generated

1. Which predictors are used by the tree?

Text

Description automatically generated

1. Compare the overall performance (e.g., accuracy or error rates) of the two decision trees (from Q3 and Q5). Which model has better predictive performance? Explain why.

In the overall performance we can see that the initial decision tree had better predictive performance when compared to the latter. The reason for this is because ClosePrice has a huge impact on the accuracy of the model. The model will have a hard time predicting without the closing price.