## **APPLIED BUSINESS ANALYTICS**



## 2024-25, Spring Semester

## Exercíses to Practice - Week 7

- 1. (Conceptual Question) Suppose you are trying to classify a variable where 96% of its observations equal 0 and only 4% equal 1. You run a decision tree classifier, and the classification table shows that 97% of the classifications are correct. Why might this large percentage still not be cause for celebration?
- 2. (Conceptual Question) What are the strengths and drawbacks of classification trees?
- 3. (Conceptual Question) Explain the CART Algorithm for Decision Trees.
- 4. (Conceptual Question) How does a Decision Tree handle continuous (numerical) features?
- 5. The file eBayAuctions.csv contains information on 1972 auctions that transacted on eBay.com during May-June 2004. The goal is to use these data to build a model that will classify auctions as competitive or noncompetitive. A competitive auction is defined as an auction with at least two bids placed on the item auctioned. The data include variables that describe the item (auction category), the seller (his/her 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 at which the auction closed. The task is to predict whether or not the auction will be competitive.

**Data Preprocessing.** Convert variable Duration into a categorical variable. Split the data into training (60%) and validation (40%) datasets.

- a. Fit a classification tree using all predictors. To avoid overfitting, set the minimum number of records in a terminal node to 50 and the maximum tree depth to 7. Write down the results in terms of rules. (Note: If you had to slightly reduce the number of predictors due to software limitations, or for clarity of presentation, which would be a good variable to choose?)
- b. Is this model practical for predicting the outcome of a new auction?
- c. Describe the interesting and uninteresting information that these rules provide.
- d. Fit another classification tree (using a tree with a minimum number of records per terminal node = 50 and maximum depth = 7), this time only with predictors that can be used for predicting the outcome of a new auction. Describe the resulting tree in terms of rules. Make sure to report the smallest set of rules required for classification.
- e. Plot the resulting tree on a scatter plot: Use the two axes for the two best (quantitative) predictors. Each auction will appear as a point, with coordinates corresponding to its values on those two predictors. Use different colors or symbols to separate competitive and noncompetitive auctions. Draw lines (you can sketch these by hand or use Python) at the values that create splits. Does this splitting seem reasonable with respect to the meaning of the two predictors? Does it seem to do a good job of separating the two classes?
- f. Examine the lift chart and the confusion matrix for the tree. What can you say about the predictive performance of this model?
- g. Based on this last tree, what can you conclude from these data about the chances of an auction obtaining at least two bids and its relationship to the auction settings set by the seller (duration, opening price, ending day, currency)? What would you recommend for a seller as the strategy that will most likely lead to a competitive auction?