## Assignment 1 Decision Tree

## Classification tree using all predictors

A diagram of a structure

AI-generated content may be incorrect.

* Decision Rule

For competitive auctions:



For non-competitive auctions:



* Findings

Obvious Findings:

1. Both prices matter a lot in the number of bids within an auction.
2. Auctions starting with very low price attract more bidders, so are mostly competitive.
3. Auctions which see great difference between opening and closing prices tend to experience more bids, so are mostly competitive.

Unexpected Findings:

* 1. Despite the expectation that more competitive auctions might last longer, and that during weekends people shall have more time to focus on bids and attend auctions, neither Duration nor EndDay appear in the decision rule of the tree. Similarly, despite the expectation that certain types of items should be more attractive to bidders, Category does not appear in the decision rule.
  2. Auctions starting at high opening prices are more likely to be competitive when the sellers have smaller past-auction score. This is not expected as more experienced sellers are expected to catch more attention from bidders, and be more successful in increasing the competitiveness of auctions. However, it might be that the low-rated sellers use aggressive pricing to compete, and the resulting lower price attracts more bidders to enter the auction.

From the above results, if I the number of predictors need to be slightly reduce due to software limitations, or for clarity of presentation, the 2 predictors OpenPrice, and ClosePrice should be chosen to capture the most relevant influence as shown from the feature importance ranking, where OpenPrice accounts for 48% importance, ClosePrice for 47%, SellerRating for 4.66%, and all other predictors see 0 importance.

However, this model is not practical for predicting the outcome of a new auction, as the ClosePrice is not known until the auction completes. For predicting the outcome of a new auction, the predictor ClosePrice should be removed.

## Interpretation and Effectiveness of Decision Tree Splits

To visualize the decision tree boundaries over a scatter plot, the two predictors OpenPrice and SellerRating were selected, as they exhibit the highest importance metrics and together account for over 95% of the total importance in the decision tree. The tree is therefore fitted using only these two predictors. Furthermore, without log-transformation, the scatter plot is tightly clustered due to large outliers, making the decision boundaries difficult to interpret. To address this, both continuous predictors were log-transformed, consistent with the procedure applied in the EDA section.

A diagram of a diagram

AI-generated content may be incorrect.

Decision boundary visualization on a scatter plot:

A graph showing a number of red and blue dots

AI-generated content may be incorrect.

* Interpretation:

The splits are interpretable, and align with the findings of the decision tree on all predictors: sellers with low ratings are mostly classified as competitive (they might do competitive pricing to attract bidders), and items with high starting prices are mostly non-competitive.

* Effectiveness:
* The red and blue dots are mixed-up, especially in the mid-range of starting price and seller ratings. This is expected as for medium starting price or seller-rating items, other factors (e.g., category, duration, timing) may strongly influence the outcome, so using a two-way scatter plot offers limited dimension to see the impacts of these other factors.
* However, the tree performs reasonably well in separating extreme regions, classifying those with high rating and low price as competitive, and those with low rating and high price as non-competitive.

Classification Table for test set

A black text on a white background

AI-generated content may be incorrect.

The accuracy for the test set is 0.7148, which is only a little lower than for the training set: 0.7270. This suggests that over-fitting is not occurring for the current model. Among the actual competitive auctions, 77.6% are predicted correctly by the model, and among the actual non-competitive ones, 62.9% are predicted correctly. Among the predicted competitive auctions, 72.3% are actually competitive, and among the predicted non-competitive auctions, 70.3% are actually non-competitive.

Based on the fitted decision tree, OpenPrice and SellerRating alone are sufficient to predict whether an auction is likely to be competitive, achieving an accuracy of over 70%. Once these two variables are considered, additional predictors such as EndDay, Duration, and Currency do not provide further predictive power from the decision tree’s perspective. This finding is further supported by cross-validation across various decision tree parameters, including maximum depth, minimum samples per split, and maximum features (see the last tree in the codes), while maintaining a minimum of 50 records per terminal node to prevent overfitting. Intuitively, this is consistent with the inherent feature selection capability of decision trees: the model prioritizes the most informative predictors to determine splits, effectively ignoring less relevant variables once a split is made. In this case, the tree assigns greater importance to continuous variables compared with categorical ones with fewer levels (in current case 2 levels for dummies), highlighting a known limitation of decision tree models.

## Recommendations for seller strategy based on Results

1. Set an optimal opening price

* Decision tree shows that extremely high or low opening prices strongly influence competitiveness. **Medium-range prices** tend to be competitive because they attract more bidders without scaring them away.
* Sellers should s**et the opening price in a range that balances attractiveness and value**—not too high to discourage bidders, not too low to signal low value.
* Sellers could analyze past auctions for similar items to find the suitable start price.

1. Maintain or improve seller rating

* Higher seller ratings correlate with more competitive auctions.
* Buyers tend to trust sellers with good reputations, leading to more bids.
* Recommendation: Focus on good service, accurate descriptions, and timely shipping to maintain a strong rating.

1. Other factors less influential

* Variables like EndDay, Duration, and Currency had little effect on competitiveness in the tree.
* Recommendation: These can be chosen based on convenience or cost, as they are less likely to impact bidding intensity.