Project Overview: Predicting eBay Auction Competitiveness Using Decision Trees

1. Objective

The primary objective of this project is to develop a classification model that predicts whether an eBay auction listing is *competitive*. A competitive auction is defined as one that generates significant bidder engagement. The model leverages decision tree algorithms to provide transparent, interpretable classification rules based on various auction-related features.

2. Data Description

The dataset was sourced from a structured Excel file containing detailed information on eBay auctions. It includes both categorical and numerical predictors such as:

- Auction characteristics: Category, Currency, Duration, End Day
- Performance metrics: Seller rating, Opening price
- Outcome variable: A binary label indicating competitiveness (Competitive?)

The data was verified to be free from missing values and suitable for supervised learning tasks.

3. Preprocessing and Feature Engineering

To prepare the dataset for modeling:

- Categorical variables were transformed using one-hot encoding to generate binary dummy variables, ensuring compatibility with scikit-learn's modeling framework.
- The dataset was reviewed for class balance. The target variable (Competitive?) was found to be evenly distributed, minimizing the risk of bias in accuracy-based evaluation metrics.
- In the second modeling iteration, target leakage was addressed by removing the ClosePrice field—an outcome variable irrelevant to prediction at the listing stage.

4. Exploratory Data Analysis (EDA)

Initial exploratory analysis included:

- Bar charts to visualize class distribution of the target variable.
- Correlation matrices among numeric predictors to assess potential multicollinearity and understand linear relationships.
- Scatter plots of seller rating versus opening price, with points colored by competitiveness status, provided insights into potential predictor interactions.

To reduce skewness and highlight nonlinear relationships, logarithmic transformations were applied to numerical variables exhibiting long-tailed distributions.

5. Modeling Approach

The core modeling technique was the Decision Tree Classifier, chosen for its interpretability and ability to handle nonlinear feature interactions without scaling.

- Data was split into training and test sets using a 60/40 split ratio, maintaining class distribution through stratified sampling.
- A decision tree was trained with a minimum leaf size constraint (min_samples_leaf = 50) to reduce overfitting and improve generalization.
- Two models were built:
    - Model 1 included all predictors.
    - Model 2 excluded the ClosePrice variable to create a more realistic, pre-outcome prediction model.

6. Model Evaluation

Model performance was evaluated using the following metrics:

- Classification accuracy on the holdout (test) set.
- Confusion matrices to assess precision, recall, and misclassification rates.
- Cross-validation scores (in the extended setup) to evaluate model stability.

In both iterations, the models demonstrated good generalization capability, with balanced error rates and acceptable overall accuracy.

7. Model Interpretation

To enhance explainability:

- Textual tree representations were generated, outlining the full sequence of decision rules from root to leaf.
- Graphical tree visualizations were constructed using hierarchical diagrams, labeling both split criteria and class outcomes.

These tools provided clear insight into which features (e.g., opening price, seller rating, auction category) were most influential in determining competitiveness.

8. Key Findings

- Seller rating and opening price showed strong influence on the competitiveness of an auction.
- Listings with lower opening prices and higher seller ratings tended to be more competitive.
- Certain categories and end days also played a role, suggesting behavioral patterns among bidders.

Removing the ClosePrice field in the second model improved model fairness and applicability for real-time auction optimization.

9. Conclusion

The decision tree modeling process successfully delivered a transparent and effective tool for classifying eBay auctions as competitive or not. By carefully handling feature engineering, model tuning, and interpretation, the project achieved its goal of providing actionable, interpretable insights for auction performance forecasting.

10. Recommendations for Future Work

- Model Expansion: Incorporate ensemble methods such as Random Forest or Gradient Boosting for improved predictive performance.
- Hyperparameter Optimization: Conduct grid search across tree depth, split criteria, and pruning thresholds.
- Feature Importance Analysis: Quantify variable contributions to better inform auction design decisions.
- Real-Time Deployment: Integrate the model into an eBay listing assistant to guide sellers in optimizing auction parameters pre-launch.