ETC3250 Business Analytics Project

By Tennis Pros

I. Introduction

★ **Motivation**
  ○ Creating a model that predicts the outcomes of professional tennis matches
  ○ Objective: to provide the most accurate predictions for the outcomes

★ **Data description**
  ○ 49,914 observations and 32 variables
  ○ Variables
    ■ Response: 'outcome' with three classes (F, U, W)
    ■ Predictors: categorical and numerical
  ○ Issues: Subjectivity created when classifying unforced and forced errors.

II. Methodology

★ **Models attempted:**
  ○ **<u>XGBoost</u>**
  ○ GBM
  ○ Neural Network, K-means clustering, SVM
★ **XGBoost Improvements**
  ○ Variable Conversion
  ○ Removing Outliers
  ○ Feature Engineering
  ○ Variable Importance

III. Findings and Discussion

★ **Precision:**
   ○ Forced errors: 0.839
   ○ Unforced errors: 0.897
   ○ Winner: 0.992
★ **Sensitivity:**
   ○ Forced errors: 0.872
   ○ Unforced errors: 0.871
   ○ Winner: 0.989
★ **Overall error:** 0.091
★ **Room for future improvements:**
   ○ Variable measurements
   ○ Subjective counting of outcomes
   ○ Separate models for Men's and Women's matches
   ○ Specific model for F and U errors

IV. Conclusion

★ **Summary:**
- Model performance
- Motivation - objective achieved?
- What can be done differently?

Thank you! Q&A time…