SPRINT 2

Predicting Tennis Outcomes

TABLE OF CONTENTS

Can tennis outcomes be predicted?

Problem

- Uncertainty and debate around who will win a tennis match

Solution

- Predictive model to see not only who will win, but what features are important

Impact

- Fans
- Analysis

Preprocessing

Filtering dataset down

- Modern tennis: 2000s onwards
- Highest level professional tournaments
- Rows where NAN values
- Raw: 73247 entries Filtered: 25585 entries

Building a numerical dataset with a target variable

- Looking at data in pandas

E D A

checking for class imbalance in target variable and simple model

- ~68% for higher ranked
- Class imbalance so will stratify test and train

Non-linearity of some features and no real demarcation

- Age and potentially height
- Looking at data in pandas

MODELING

Outcome of logistic regression

- ~68%for train and test scor
- Overall the exact same score as guessing so not great
- But also not a lot of overfitting by the model

Areas for improvement

- Adding additional features

NEXT STEPS

Non-linear models

- Random forest
- Decision trees

Adding features

- Specialist on a specific court type
- Streak / trending player
- Grand Slam champion