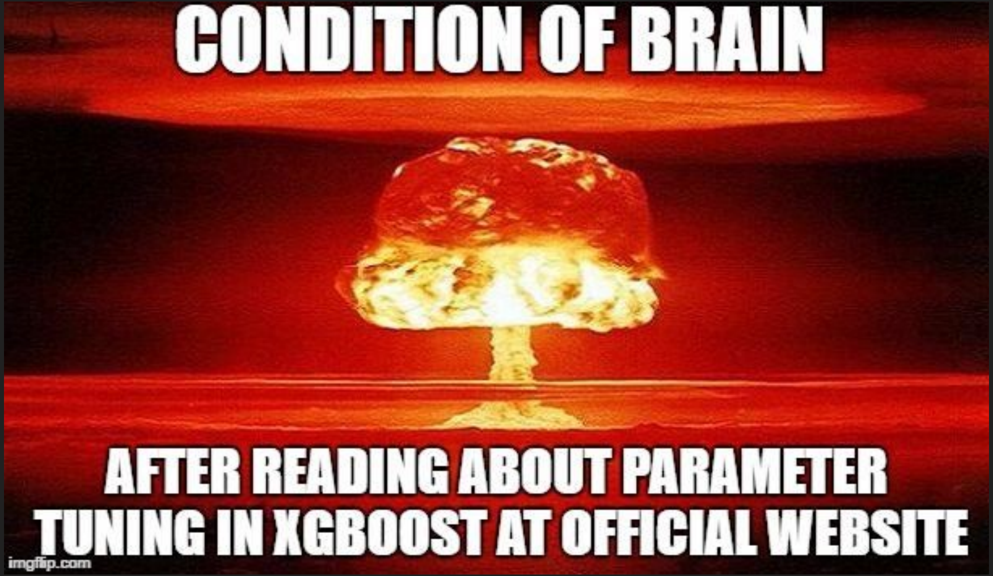
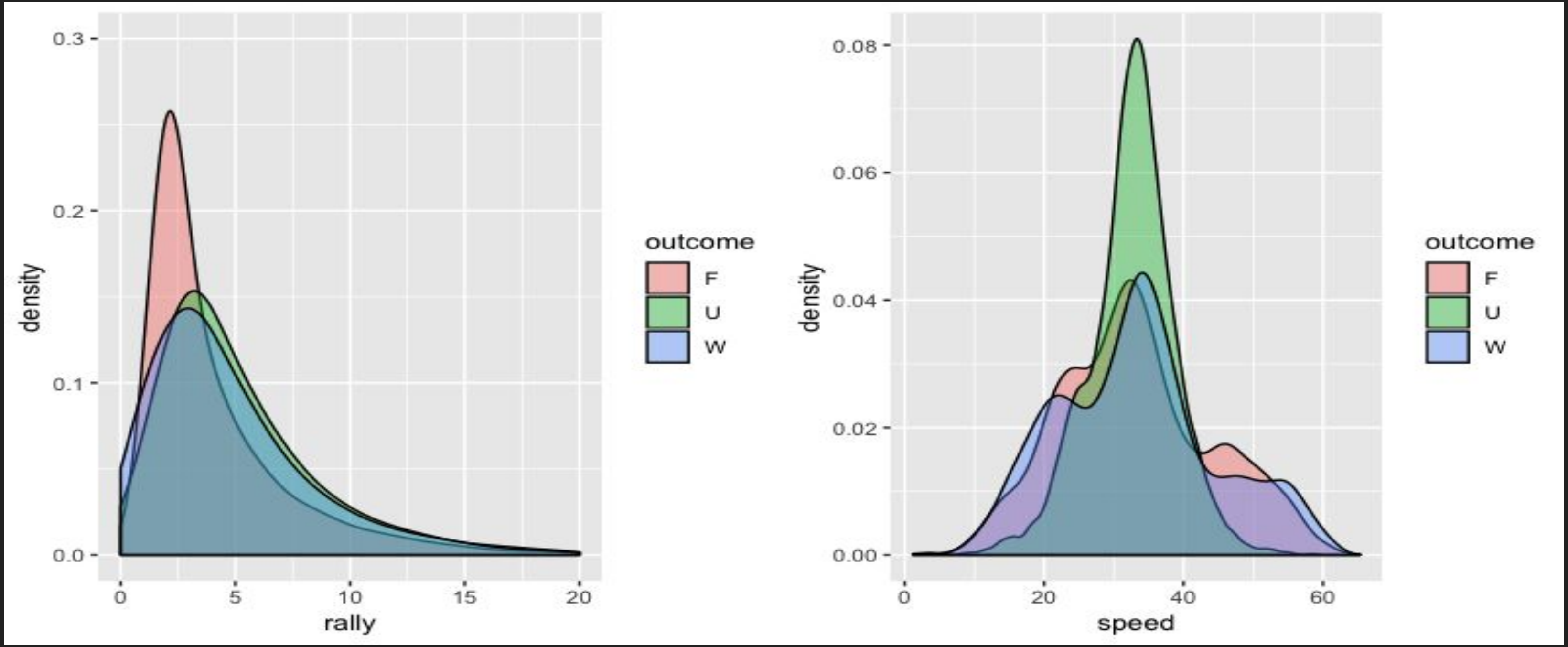
Analytics in Tennis


CONDITION OF BRAIN
AFTER READING ABOUT PARAMETER TUNING IN XGBOOST AT OFFICIAL WEBSITE
imgflip.com

David Kontrobarsky,
Hai Ninh Duong,
Hua Chen Li,
Nick Henderson

Data exploration

Density plots of the most important variables

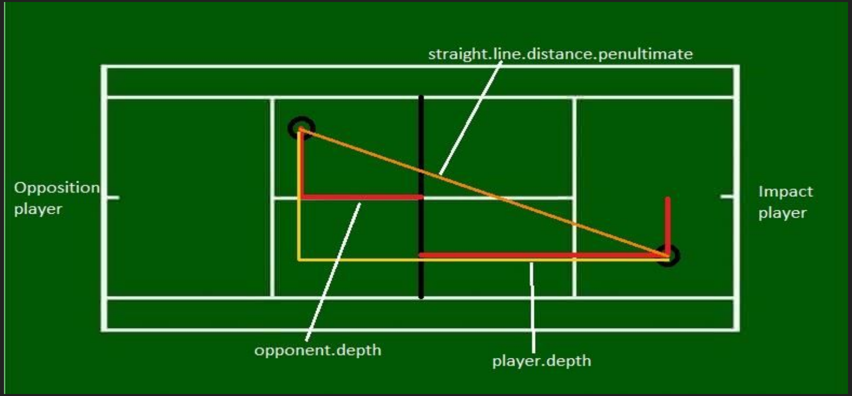


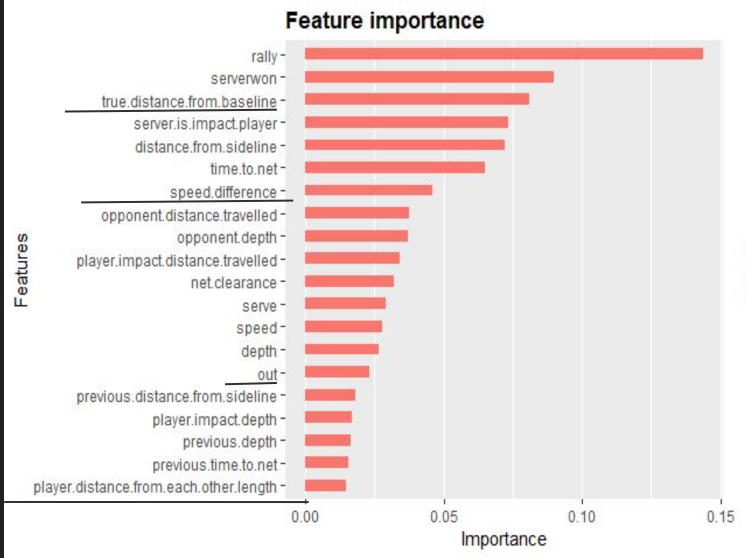
Feature Engineering & Feature selection

Some of the features we added include:

- *Whether the point was in a men's or women's match?*
- *How far away from each other were the players in total?*
- *What was the difference in speed between the penultimate and final shot?*
- *Was the shot out? (combination of two present variables)*

The highlighted variables are feature engineered variables.





Methodology

We started with random forest as a base model, and eventually found that xgboost gives the best prediction accuracy of around 92%

ENSEMBLING

We also combined the deep learning, random forest and gradient boosting models into an ensemble model, however, the prediction accuracy was not greatly improved over the individual xgboost.

Concluding Remarks and Recommendations

- Even with our best model, there is still a significant error in the classification of the points

- The model has come quite far, but there is still room for improvement in future