*Does Roger Federer has a chance to beat Rafa Nadal at Roland Garros?*

# ACE.AI TENNIS MATCH SCORE PREDICTOR

Carson Chen
June 10, 2019

*Machine learning based prediction engine for professional men's tennis games*

# PROBLEM STATEMENT & DEMO

- It's always fun to predict how a tennis match goes

- It's somewhat predictable but never lack of surprises

- You might make some extra $$$ if you are smart enough to beat the odds!

**LINK TO ACE.AI**

# DATA

All 44,700 ATP (Association of Tennis Professionals) men's singles matches between 2000 and March 2018

- From Kaggle.com

- Match detail columns such as type of court, location, type of tournament, stage of tournament, score in sets

- Winner and loser detail columns such as names, ranks, handedness

Additional generated features:

- H2H (Head-to-head) records between every pair of players (such as 6W1L)

- Overall game winning percentage by type of court

# MODEL

- Boosted trees model using XGBoost (selected), random forest model (explored)

- Reshuffled all matches for the dataset to be balanced, 50% wins

- 75-25 Train test split, test model-predicted results against actual results

- <u>H2H Records</u>, <u>surface winning percentages</u>, and <u>ranks</u> are most significant predictors


- Cross-validated Accuracy: 76.0%

- Display probability rather than binary (Win vs Lose)

# INSIGHTS



Data:

- Most players do not play many ATP games at all

Model:

- A lot of games are less of a toss-up than you think!

- Model behaves less reliably for new players / uncommon match-ups

# THANK YOU

Developer: Carson Chen

Northwestern MSiA Class of 2019

Github: https://github.com/carsonzchen/MSiA423_ACE.AI

QA: Shreyas Sabnis

Data Source: Edouard Thomas on Kaggle.com

Data: https://www.kaggle.com/edouardthomas/atp-matches-dataset#confidence_data.csv

Special Thanks: Chloe and Fausto