

# Patterns of Play

Predicting tennis match outcomes  
and player styles

Dmitri Tarasov  
Hunter Hobbs  
Nivetha Kesavan  
Pratik Revankar





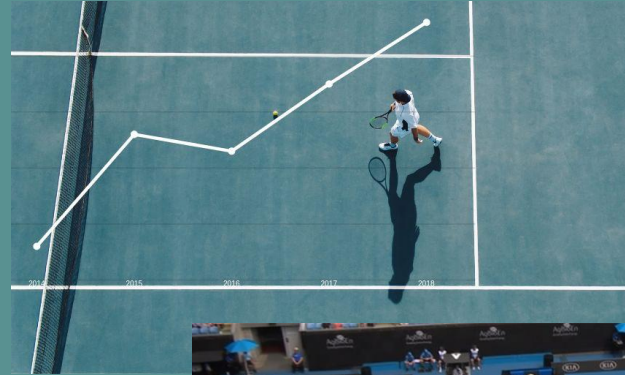
# Project Description

Using Tennis data from the Open Era, we aim to analyse match statistics for players in the WTA and ATP circuits, to predict future match outcomes amongst players and highlight some key rivalries. Using this data, we also aim to analyse player game styles and techniques, and model an “ideal player” based on user inputs.



# Goals

1. Predict player head-on match results, using Markov-Chain predictions.
2. Analyse player style (eg: offensive, defensive, baseline, volley)
3. Aggregate and highlight top rivalries in the ATP/WTA tour for past 10 years (or since the Open Era)
4. Model a player based on input stats (“Build your Ideal player”)





# Prior Work

1. Clarke and Dyte [1] used ATP rankings to predict the player's chance of winning in a head to head match. They used Logistic Regression model to predict the winner of the head to head based on the difference in points rating
2. Klassen and Magnus [2] used rankings and points won by players on service to predict the match outcome. Their technique could predict the outcome at the beginning of the match as well as during the match using a graph technique based on the current score and current server.
3. Liu [3], Newton [4] modelled the probabilities of a player winning a set and match based on the probability of a player winning a point while serving.
4. Knottenbelt et al. [5] used a hierarchical Markov model to predict the probability of each player winning a match. The model uses the match statistics of an opponent that has been encountered by both players and it computes the probability of each player winning a serve and the match



## Prior Work

5. McHale and Morton [6] developed an alternative player ranking based on the past match results, number of points won by each player in a match, how recent the match was played, etc. This new ranking based on absolute performance was used to predict the winner of a match and it performed better than using ATP rankings.
6. Zhang [7] created a network model of ATP matches where each player was a node in the network and based on the network, structural roles were discovered. Structural role discovery revealed the prominent attribute in any individual player's game.
7. Leeuw et. al [8] performed tactical analyses for a specific player using Subgroup Discovery. Using subgroup discovery they found the characteristics of a successful serve point of a specific player and they used it to identify the difference between generic successful serve point and serve point of the specific player.

# Data

1. ATP World Tour Tennis Data
2. Serve and Volley Data - ATP World Tour
3. WTA World Tour Tennis Data
4. Tennis Betting Data
5. Tennis ATP Results since 2001
6. The Match Charting Project - Jeff Sackmann



# Proposed work

1. Check sparsity of each data set
2. Cleanup the data
  - a. Datahub: fill in missing data with data from other datasets, map name to playerid.
  - b. Jacksackmann: combine the data from different years to one dataset.
  - c. Serve\_and\_volley: combine with other per match data, get player slug-id table from dataset
  - d. Wta-matches: combine with other per match data
  - e. Atp-matches: combine with the Jeff Sackmann data.
  - f. Alldata: combine with other per match data
  - g. Opendatasoft: combine with other per match data
  - h. Match-Charting-Project: fill in empty values.



# Proposed work

3. Augment data
  - a. Combine the data from the 5 different sources for player matches.
  - b. Combine the data from the player match events per player.
  - c. Augment charting data with data about player handedness and backhand.
4. Aggregate data
  - a. Combine data of match results from different datasets into one dataset.
5. Analyze the data
  - a. Make a hidden markov model by computing the probabilities of state transitions
  - b. Perform cluster analysis with the data about player moves to see if players can be separated on style based on player moves.
  - c. Perform cluster analysis on the player play and match aggregate to see who would be the best rivals.



# Tools

1. AWS
2. Sklearn
3. Tensorflow
4. Pandas
5. Matplotlib

# Evaluation

1. Predicting match outcomes
  - Have a holdover set to evaluate
2. Player styles
  - Manual evaluation to see if the clusters have distinctive player styles.
3. Player rivals
  - Manual evaluation of clusters



# References

1. Clarke, S. R. and Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research*, 7(6):585–594
2. Klaassen, F. J. and Magnus, J. R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454):500–509
3. Y. Liu. Random walks in tennis *Missouri Journal of Mathematical Sciences*, 13 (3) (2001)
4. Newton, P. K. and Keller, J. B. (2005). Probability of winning at tennis i. theory and data. *Studies in applied Mathematics*, 114(3):241–269
5. William J. Kottenbelt, Demetris Spanias, Agnieszka M. Madurska : A common-opponent stochastic model for predicting the outcome of professional tennis matches, *Computers & Mathematics with Applications*, Volume 64, Issue 12, 2012, Pages 3820-3827, ISSN 0898-1221
6. McHale, I. and Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2):619–630.
7. Zhang, S (2019) [Modelling ATP Tennis as a Network](#)
8. Leeuw, Arie-Willem & Hoekstra, Aldo & Meerhoff, Rens & Knobbe, Arno. (2019). *Tactical Analyses in Professional Tennis*.