**Predicting ATP Singles Game Outcome—Sprint 1 Review**

**Team 13**

Dev: Wei Li

PO: Saurabh Tripathi

QA: Michael Pauleen

**Highlights:**

In the aspect of project logistics, we have successfully set up Sphinx documentation, built web app infrastructure locally, and set up AWS test environment.

In the aspect of data analytics, we have collected and merged data into a single csv file.

**Review progress:**

**Demo/analysis:**

**Data description:** ATP Singles Match Data

**Time range:** 2000.Jan – 2017.Sept

**Data size:** 50,000 game-level observations

**Data fields:**

* Tournament details

|  |
| --- |
| tourney\_id |
| tourney\_name |
| surface |
| draw\_size |
| tourney\_level |
| tourney\_date |
| match\_num |

* Player details

|  |  |
| --- | --- |
| winner\_id | loser\_id |
| winner\_seed | loser\_seed |
| winner\_entry | loser\_entry |
| winner\_name | loser\_name |
| winner\_hand | loser\_hand |
| winner\_ht | loser\_ht |
| winner\_ioc | loser\_ioc |
| winner\_age | loser\_age |
| winner\_rank | loser\_rank |
| winner\_rank\_points | loser\_rank\_points |

* Per-match stats for both players

|  |  |
| --- | --- |
| w\_ace | l\_ace |
| w\_df | l\_df |
| w\_svpt | l\_svpt |
| w\_1stIn | l\_1stIn |
| w\_1stWon | l\_1stWon |
| w\_2ndWon | l\_2ndWon |
| w\_SvGms | l\_SvGms |
| w\_bpSaved | l\_bpSaved |
| w\_bpFaced | l\_bpFaced |
| score |  |
| best\_of |  |
| round |  |
| minutes |  |

**What are the inputs to your model?**

Relevant and newly created features based on available fields

**What is the output?**

Whether Player 1 win or lose

**What information will be presented to the user?**

List of available players in the database, tournament surface, tournament level

**What information will the user give (if any)?**

User will select the 2 players, tournament surface, and tournament level. The model then will output historical match statistics between 2 players and the predicted match result (win/lose odds).

**What will the training and test data be?**

Training: historical data from 2000-2016

Test: historical data in 2017

**Draw a mock up of the website**

|  |  |
| --- | --- |
| Select your Player 1 | Xxxxxx |
| Select your Player 2 | XXXXX |
| Select Tournament Level | A/C/D/F/G/M |
| Select Tournament Surface | Carpet/Clay/Grass/Hard |

(May breakdown historical match stats by tournament levels and tournament surface)

**Lessons learned**

Original charter was not well-defined to guide project planning and execution and therefore additional time was required to modify project charter.

**Recommendations**

Modified project charter as follows:

**Charter:**

To investigate the applicability of machine learning methods to the prediction of professional tennis matches by training supervised learning models using tournament and players’ information and players’ in-game performance to propose betting strategies.

**Vision:** to investigate the applicability of machine learning methods to the prediction of professional tennis matches and to improve understanding of deterministic features in predicting ATP Singles’ game outcome

**Mission:** predict game outcome by training supervised learning models (potentially logistic, SVM, neural network) using tournament and players’ information and players’ in-game performance based on a set of user selected features on tournament and players

**Success criteria:** a set of metrics for measuring overall model performance on predicting match results