



Curtin University

WHICH SCHEDULE BEST SERVES A PROFESSIONAL TENNIS PLAYER?

Graeme Ward and Dr Stephanie Kovalchik

Player Goals

- Winning tournaments?
- Making money?
- Becoming famous?
- Being highly ranked

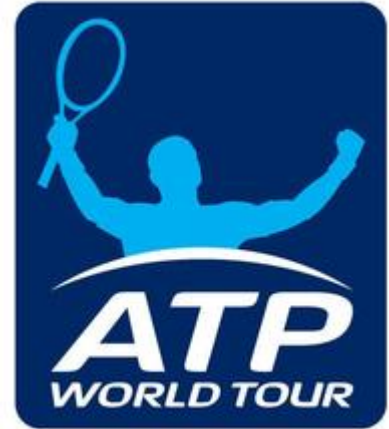


Objectives

- Identify and explore variables that characterise a schedule
- Create a model to predict the change in rank for a given schedule

Playing schedules

- 232 tournaments run in 2016
- Some requirements to fulfil
- Player chooses his own schedule



ATP Rankings

- Ranking points awarded for performance in ATP tournaments
- Best 18 tournaments in the past 52 weeks
- Players ranked by ranking points

Rank	Name	Ranking Points
1	Andy Murray	9,890
2	Rafael Nadal	7,285
3	Stan Wawrinka	6,175
4	Novak Djokovic	5,805
5	Roger Federer	4,945
6	Milos Raonic	4,450
7	Marin Cilic	4,115
8	Dominic Thiem	3,985
9	Kei Nishikori	3,830
10	Alexander Zverev	3,070

ATP Rankings on 18/06/17



Curtin University

Data

- Information on all ATP matches played by 100 of the top players in 2014 and 2015
- List of 2016 ATP World Tour tournaments



Ranking Definitions

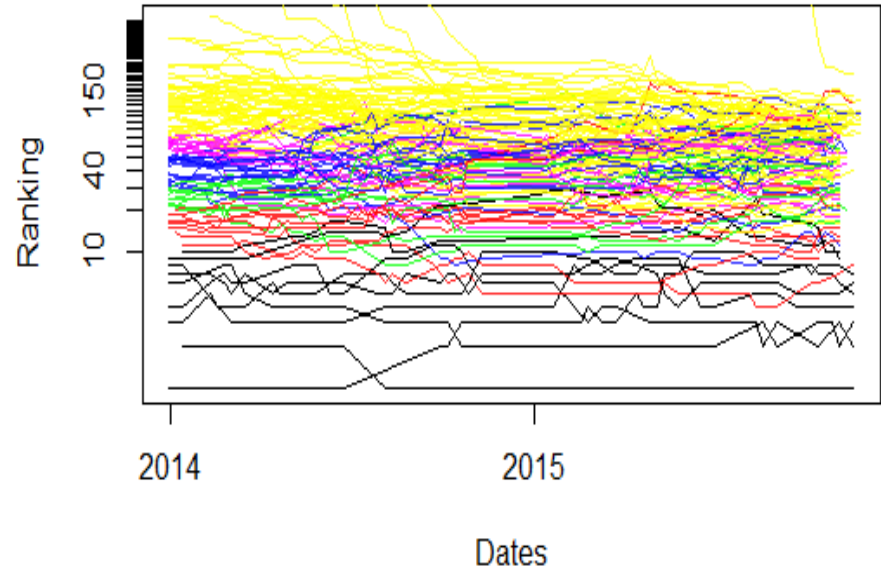
- Initial ranking used to approximate skill level
- Important as the schedule is dependent on the initial rank

Ranking transformations

adjusted ranking = $8 - \log_2(\text{ranking})$

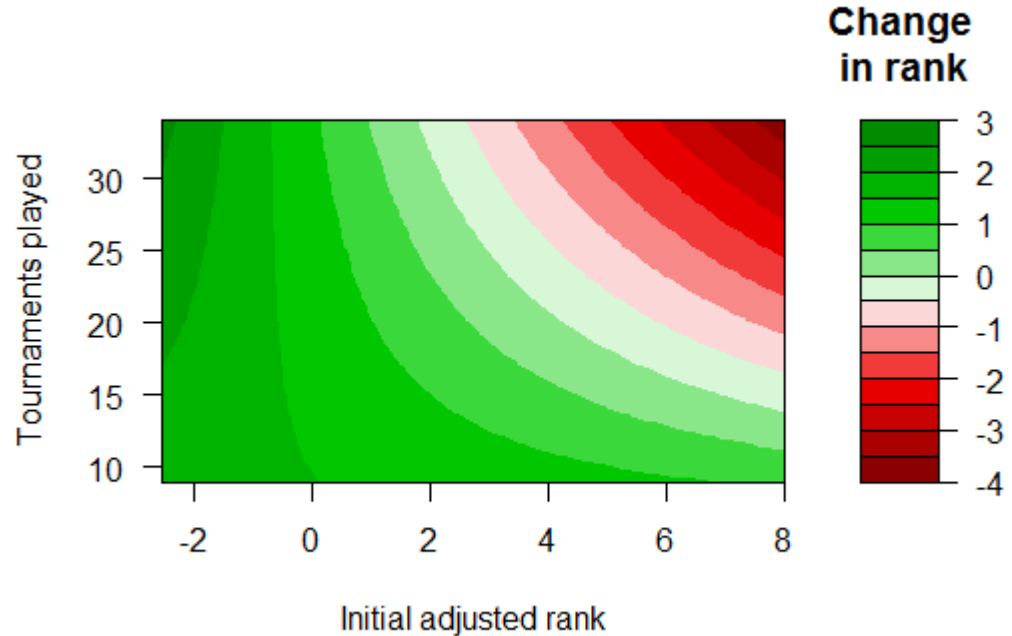
$$\Delta\text{rank} = \log_2\left(\frac{\text{initial rank}}{\text{final rank}}\right)$$

Change in ranks over time



Tournaments Played

- Removal of Davis Cup
- Ranges between 9 and 34
- Mean of 25

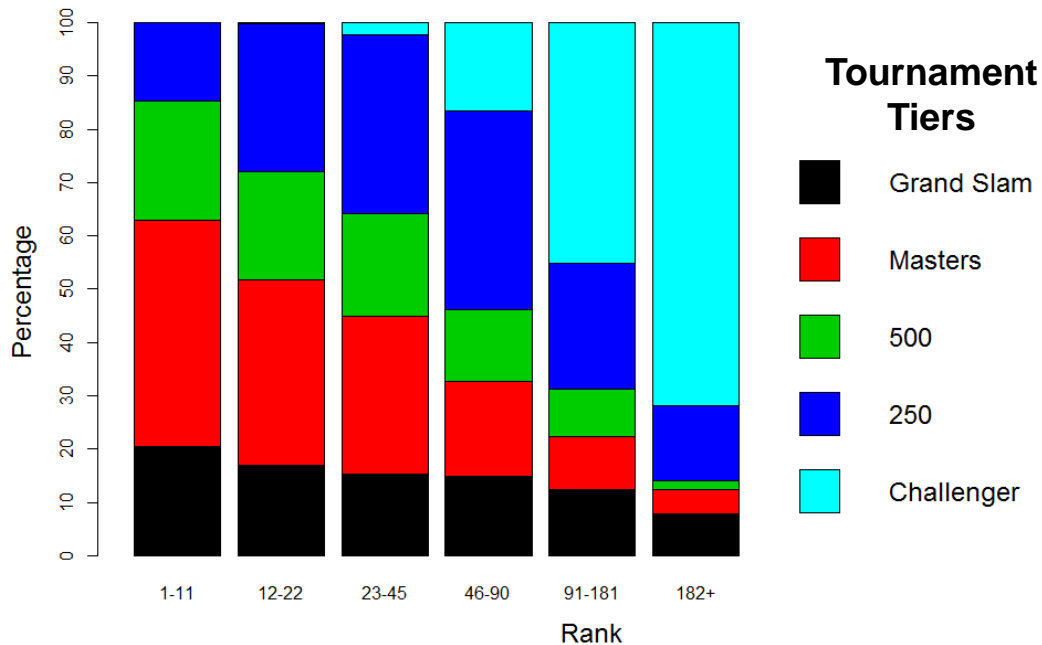


Tournament Tiers

Tournament tier name	Ranking points for winner	Number of tournaments run in 2016
Grand Slam	2000	4
Masters	1000	9
500	500	13
250	250	39
Challenger	Up to 125	167



Tournament Tiers



Congestion Score



0	1	2	3	4	5	6	7	8	9	10	11	12	13
3326	1117	344	116	41	21	6	5	5	2	2	0	1	1

Length of breaks between tournaments (weeks)

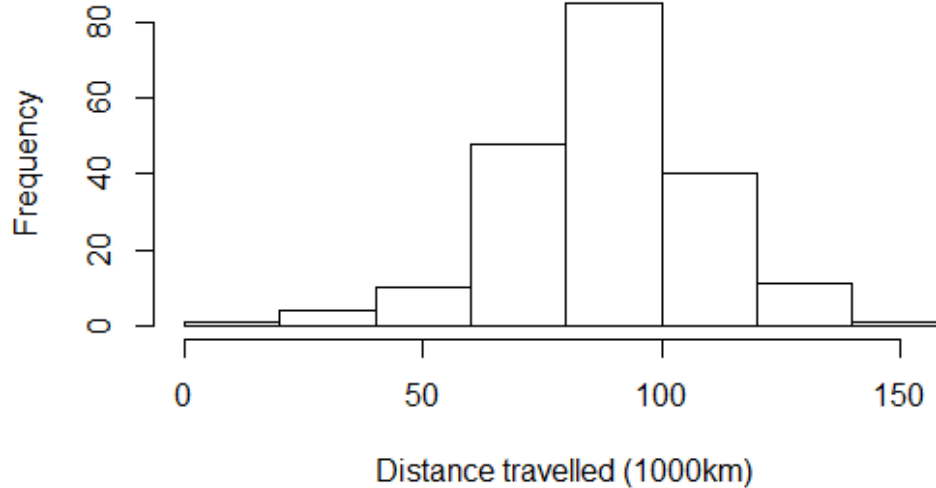


Curtin University

Distance Travelled

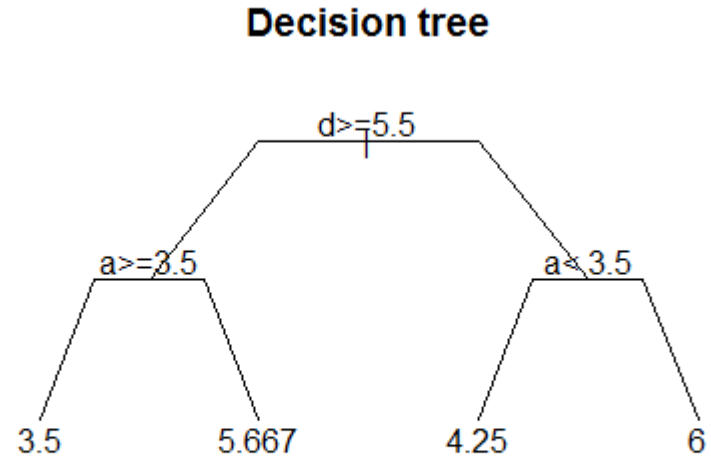


Distance Travelled



Random Forest Method

- Makes a 'forest' using many prediction models (trees) created from the data
- Creates an 'average' prediction model with lower variance than the single prediction models



Models

- Cross-validation
- Regression Model
- Random Forest Model
- Removal of tournaments played as variable

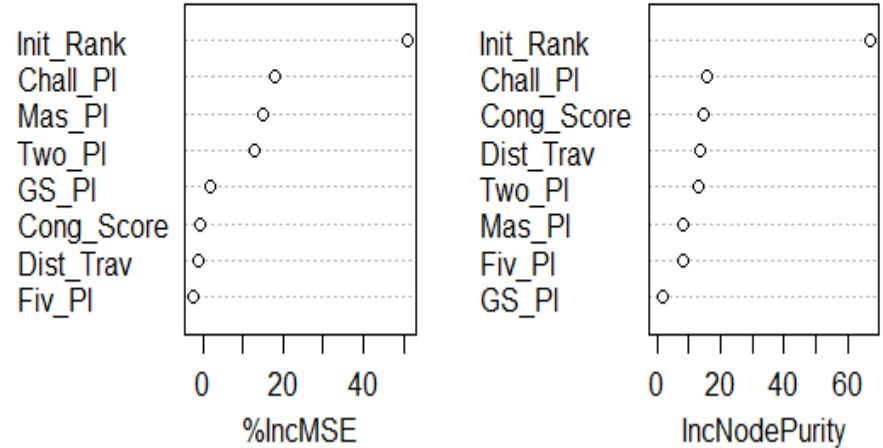
Coefficient Name	Coefficient Value
Masters	0.087
500s	0.145
250s	0.045
Challengers	-0.019
Initial rank	0.001
500:Initial rank	-0.062
250:Initial rank	-0.067
Chall:Initial rank	-0.034



Models

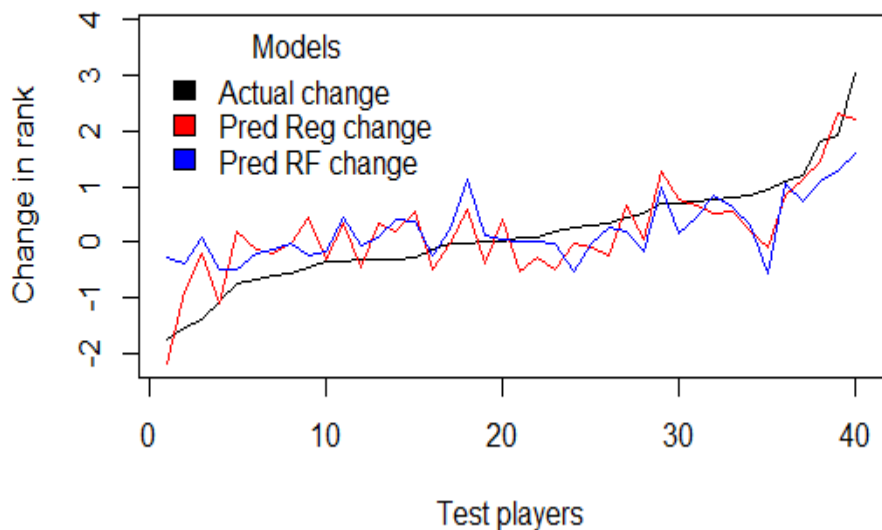
- Cross-validation
- Regression Model
- Random Forest Model
- Removal of tournaments played as variable

Variable Importance Plot



Model Comparison

Predicted change in rank for test players



Characteristics of 'difference' vectors

	Regression Model	Random Forest Model
Mean	-0.039	-0.065
Variance	0.303	0.446
RMSE	0.545	0.663

Model Application

	S1	S2	S3	S4	S5	S6
Grand Slams	4	4	4	4	3	4
Masters	6	8	3	6	2	2
500s	3	5	3	4	2	1
250s	4	7	8	12	5	1
Challengers	1	0	6	4	15	13
Congestion Score	0.187	0.155	0.326	0.270	0.095	0.026
Distance Travelled	73.38	91.47	68.93	99.76	80.50	107.63

Model Application

Rank 5	Random Forest Prediction	Regression Prediction
Schedule 1	7	8
Schedule 2	6	17

Rank 32	Random Forest Prediction	Regression Prediction
Schedule 1	30	21
Schedule 2	29	25
Schedule 3	58	59
Schedule 4	46	66

Model Application

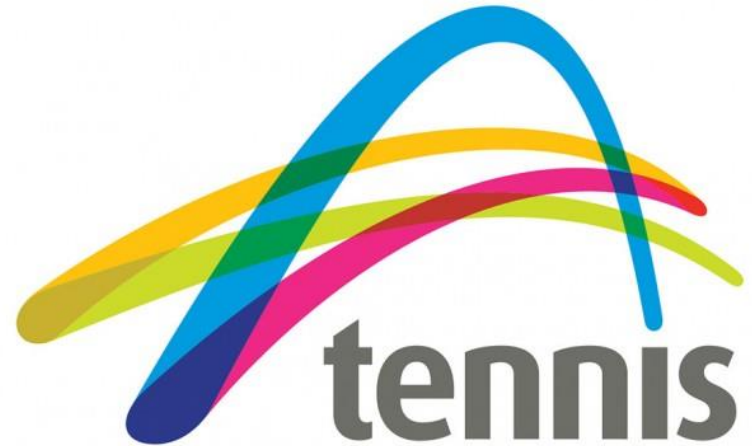
Rank 72	Random Forest Prediction	Regression Prediction
Schedule 1	44	32
Schedule 2	33	30
Schedule 3	69	63
Schedule 4	57	57
Schedule 6	79	98

Rank 100	Random Forest Prediction	Regression Prediction
Schedule 1	50	38
Schedule 3	71	64
Schedule 4	56	53
Schedule 5	99	97
Schedule 6	103	99



Further Research

- First step for Tennis Australia
- More data for wider use
- Additional schedule variables
- Additional player variables
- Use optimisation techniques to find the optimal schedule



Summary

- Seven variables found that characterise a schedule
- Regression and random forest models created to predict changes in ranks for top male players

