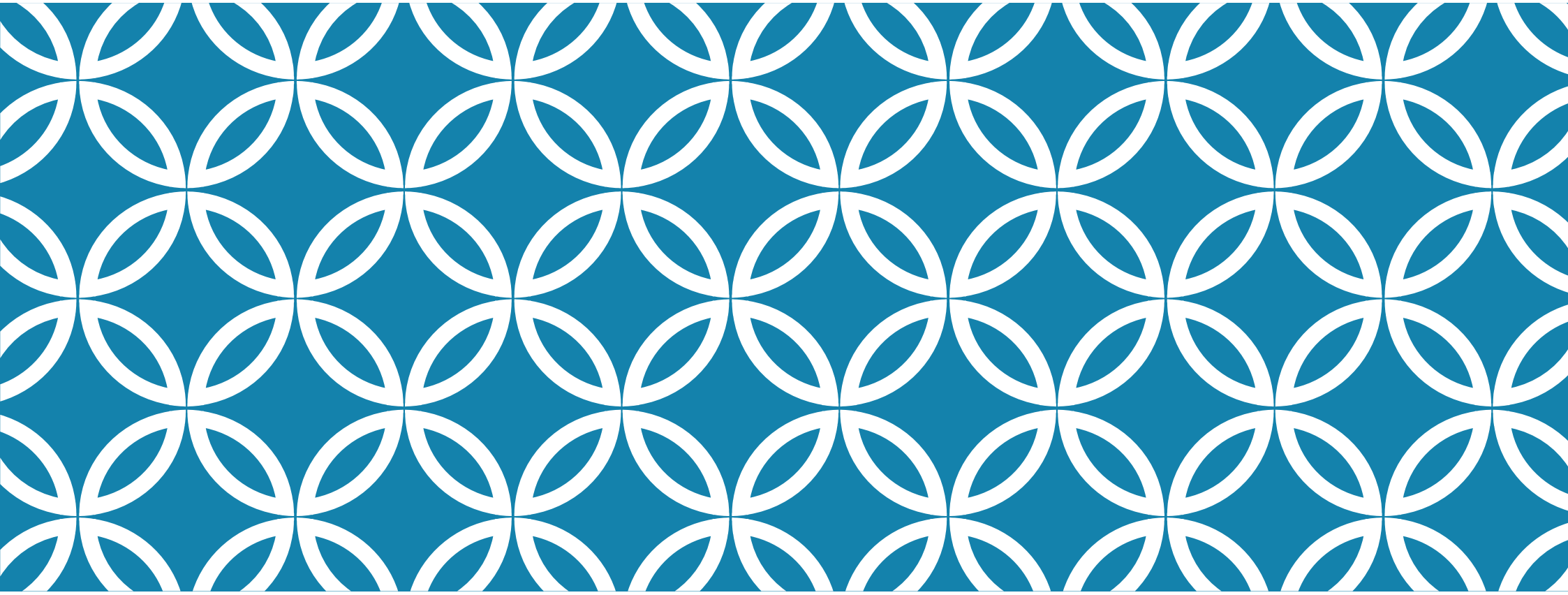


ACE.AI SCORE PREDICTOR MID-TERM REVIEW

Carson Chen

May 15, 2019

HIGHLIGHTS

- I turned the full match history file into the format into a prediction dataset made up of prediction labels and variables.

- Current accuracy (cross-validated, looking back) is at 65%

- Tennis match results seem more random than expected, there are many factors that are difficult to be captured.

REVIEW PROGRESS

1. Exploratory data analysis and data cleansing

- Data overview, descriptive statistics, and cleaning
- Dataset transformation to make it a standard format for predictive modelling

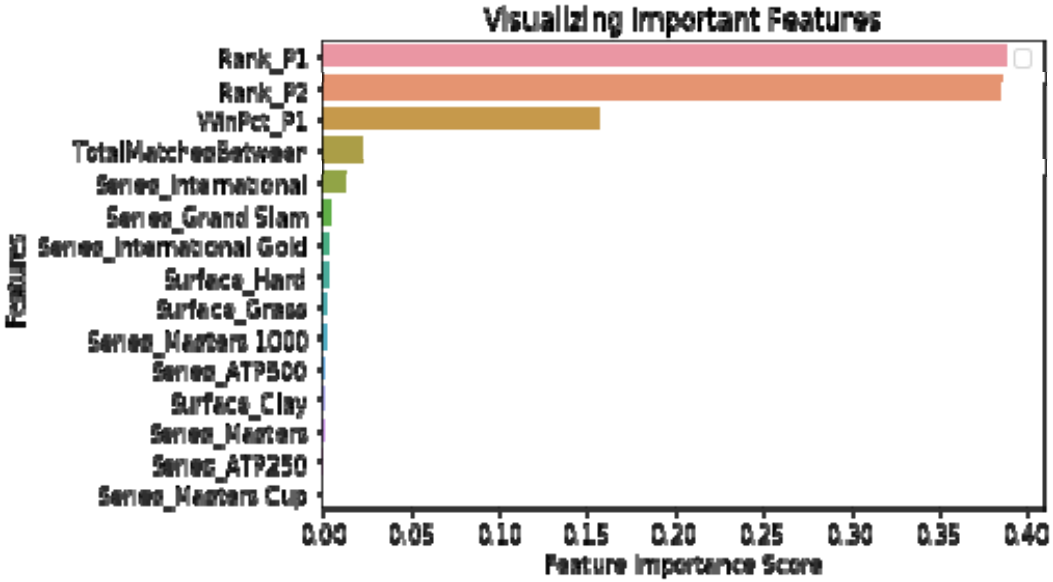
2. Model building and validation

- Engineer feature set as predictor variables (ongoing)
- Build an initial benchmark model for reference

3. Product development

- Build data pipeline for the project
- Initialize database in RDS

DEMO/ANALYSIS



Variable Importance

Prediction data set

	Rank_P1	Rank_P2	TotalMatchesBetween	WinPct_P1	matchresult	Series_ATP250	Series_ATP500	Series_Grand Slam	Series_International
0	63	77	0	0.500000	1	0	0	0	1
1	6	59	0	0.500000	1	0	0	0	1
2	174	73	0	0.500000	0	0	0	0	1
3	78	33	0	0.500000	0	0	0	0	1
4	35	206	0	0.500000	0	0	0	0	1
5	219	211	0	0.500000	0	0	0	0	1
6	2000	60	0	0.500000	0	0	0	0	1
7	36	85	0	0.500000	0	0	0	0	1
8	128	14	0	0.500000	1	0	0	0	1
9	104	38	0	0.500000	0	0	0	0	1
10	14	121	0	0.500000	1	0	0	0	1
11	163	128	0	0.500000	0	0	0	0	1

LESSONS LEARNED

Specifics of the data:

- Player rank is the single most important factor to determine a match outcome, which is indicating the current ranking mechanism is quite accurate

General deployment strategies

- It's more effective to present result in the angle of "my player" versus "the opponent player", rather than of a detached stance 'A vs B'

- Offline prediction is preferred for a better user experience, but it requires setting up a larger database

RECOMMENDATIONS/NEXT STEPS

2. Model building and validation

- Iteratively develop a set of models with engineered features, optimize parameters to find the best model

- Validate the model using primary and potential alternative metrics such as F1-score

3. Product development

- Build Flask app for deployment

- Build user interface prototype for the project