

Prediction of Tennis Matches

IEOR 242 - UC Berkeley





Introduction

US OPEN (TENNIS)



US Open Tennis 2019: Updated Men's Bracket and Final Predictions

RORY MARSDEN 
SEPTEMBER 7, 2019

0 



NEWS

World Cup: AI predicts that Germany will win

Machine learning predicts World Cup winner

Researchers have predicted the outcome after simulating the entire soccer tournament 100,000 times.

by **Emerging Technology from the arXiv**

June 12, 2018

2020 Australian Open women's odds, picks, predictions: Tennis expert says Serena Williams poised for upset

Gavin Mair is up over \$22,000 on his women's tennis tournament picks over the past three years.

by **CBS Sports Staff** Jan 19, 2020 at 11:27 am ET • 2 min read

A MACHINE LEARNING ANALYSIS OF THE NFL: PREDICTING NEW PLAYOFF CONTENDERS

 SEPTEMBER 5, 2018  BY HARVARDSPORTS  4 MIN READ  1 COMMENT

By Matty Cheng

US OPEN (TENNIS)

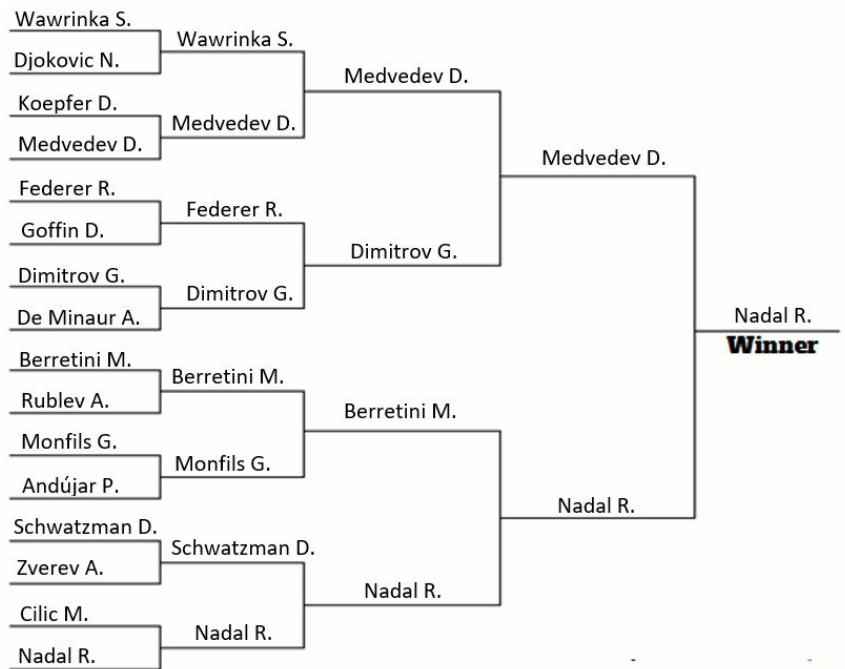


US Open Tennis 2019: Bracket Predictions, Odds for Top Players, Prize-Money Info

GILL CLARK 
AUGUST 25, 2019

0 

Our Goal: Predict 2019 US Open Winner





Approach

Data Mining

Dataset from tennis-data.co.uk/ - Data manipulation with Python

R Models

Baseline - Logistic Regression - CART - Random Forest

Python Models

Principal Component Analysis - Neural Networks

Final Model

Model with highest accuracy selected - Simulation



Data Mining

1. Raw Data Pulled From tennis-data.co.uk/

Tournament	Date	Surface	Round	Best of	Winner	Loser	WRank	LRank	W1	L1	W2	L2	W3	L3
US Open	8/29/05	Hard	1st Round	5	Massu N.	Gambill J.M.	58	151	7	6	6	2	6	3
US Open	8/29/05	Hard	1st Round	5	Baker B.	Gaudio G.	195	9	7	6	6	2	6	4
US Open	8/29/05	Hard	1st Round	5	Ljubicic I.	Summerer T	19	198	6	3	7	6	6	3
US Open	8/29/05	Hard	1st Round	5	Malisse X.	Hernych J.	54	81	6	2	7	5	6	2
US Open	8/29/05	Hard	1st Round	5	Berdych T.	Kohlischreiber	34	79	7	6	7	6	6	4
US Open	8/29/05	Hard	1st Round	5	Nadal R.	Reynolds B.	2	132	6	3	6	3	6	4
US Open	8/29/05	Hard	1st Round	5	Novak J.	Norman D.	25	112	6	4	6	4	6	2

2. Feature Generation Performed (21 Variables) to Predict Output Variable (Win/Loss)

'diff_rank', 'diff_match_win_percent', 'diff_games_win_percent', 'diff_5_set_match_win_percent',
'diff_close_sets_percent', 'diff_match_win_percent_hard', 'diff_games_win_percent_hard',
'diff_5_set_match_win_percent_hard', 'diff_close_sets_percent_hard', 'diff_match_win_percent_52',
'diff_games_win_percent_52', 'diff_5_set_match_win_percent_52', 'diff_close_sets_percent_52',
'diff_match_win_percent_hard_60', 'diff_games_win_percent_hard_60', 'diff_5_set_match_win_percent_hard_60',
'diff_close_sets_percent_hard_60', 'diff_match_win_percent_hh', 'diff_games_win_percent_hh',
'diff_match_win_percent_hard_hh', 'diff_games_win_percent_hard_hh']



R Models

Naive Baseline

Training

0	1
596	231

Test

0	1
255	99

Selecting all 0's:

72.03% of accuracy

Logistic Regression

Features Used:

Percent Difference in Matches Won

Percent Difference Between "Close" Sets

Percent Difference in Matches Won in 52 wks

	FALSE	TRUE
--	-------	------

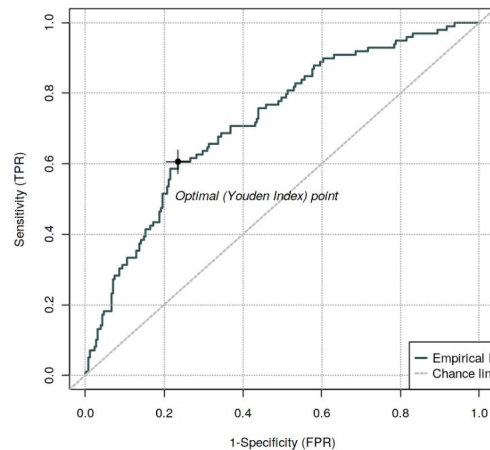
0	146	109
---	-----	-----

1	29	70
---	----	----

Accuracy: 66.3%

FPR: 32.3%

TPR: 56.5%



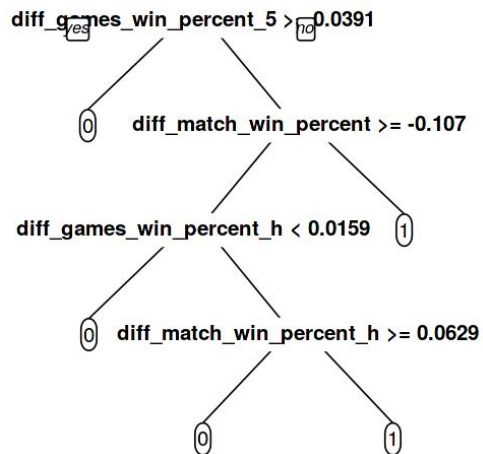
AUC: .691



R Models

CART

Random Forest

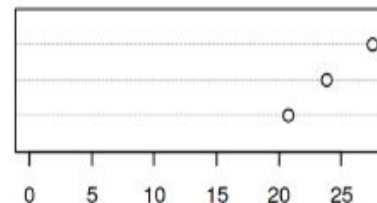


Accuracy in the test set: 70%

Variable Importance:

diff_games_win_percent_52
diff_match_win_percent
diff_match_win_percent_52

modelRF



Accuracy in the test set: 74%



Python Models

Neural Network

- Built Using Keras Library in Python
 - Layer 1: 64 Units (Relu Activation)
 - Layer 2: 32 Units (Relu Activation)
 - Layer 3: 1 Unit (Sigmoid Optimizer)
- Train Accuracy: 0.758
- Test Accuracy: **0.776**



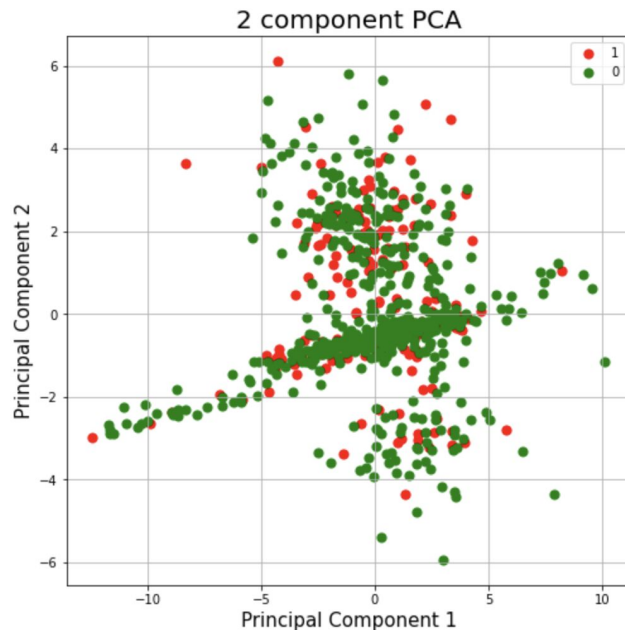


Python Models

Principal Component Analysis

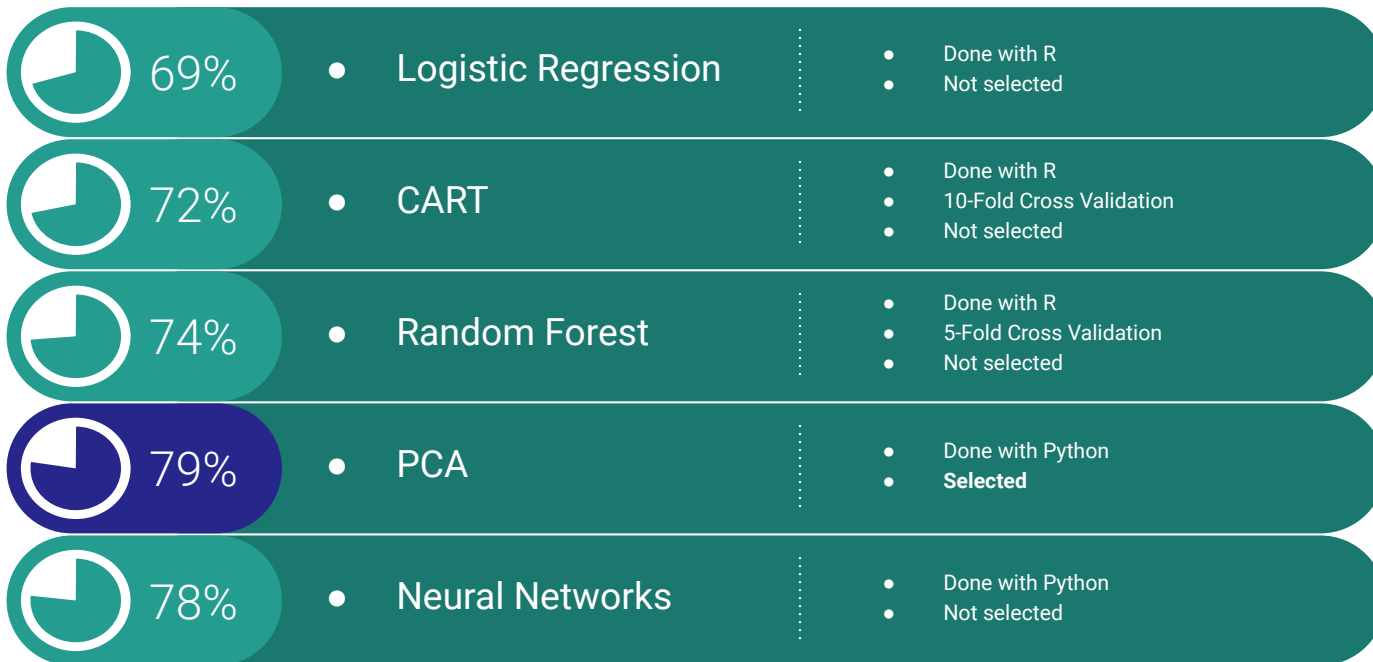
- **Standardizing Features**
- **We keep 2 principal components**
- **Test Accuracy: 0.79**

	Explained Variance percentage
First PC	64 %
Second PC	23 %

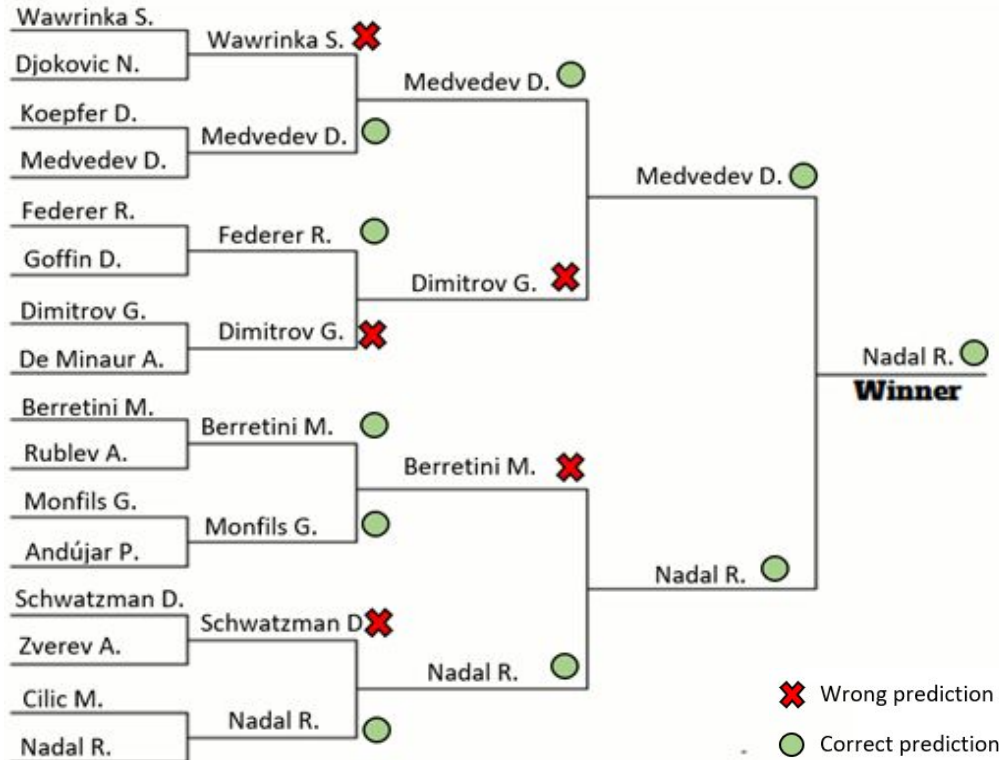




Comparison & Selection



Results (2019 US Open Model Prediction)



Next step:

US Open 2020



Thank You

Group:

Agustin Toll Villagra - MEng IEOR 2020

David Scanlan - MEng IEOR 2020

Johan Gerfaux - MEng IEOR 2020

Joshua Edwards - MEng IEOR 2020