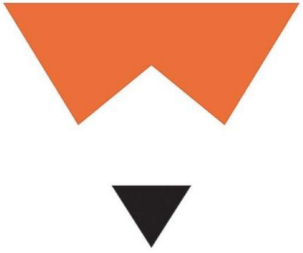
Ripton Rosen

NBA Prediction Model

Basketball Reference and the NBA

Goal

Create a classification model that can accurately predict NBA wins and losses



FiveThirtyEight

Forbes

Data Process

Web Scrape

From Basketball-Reference.com

Data Cleaning

Deal with missing values and rename columns

Data Exploration

Sort through and understand the data

Modeling

Create models based on the data to predict game outcomes

Interpretation

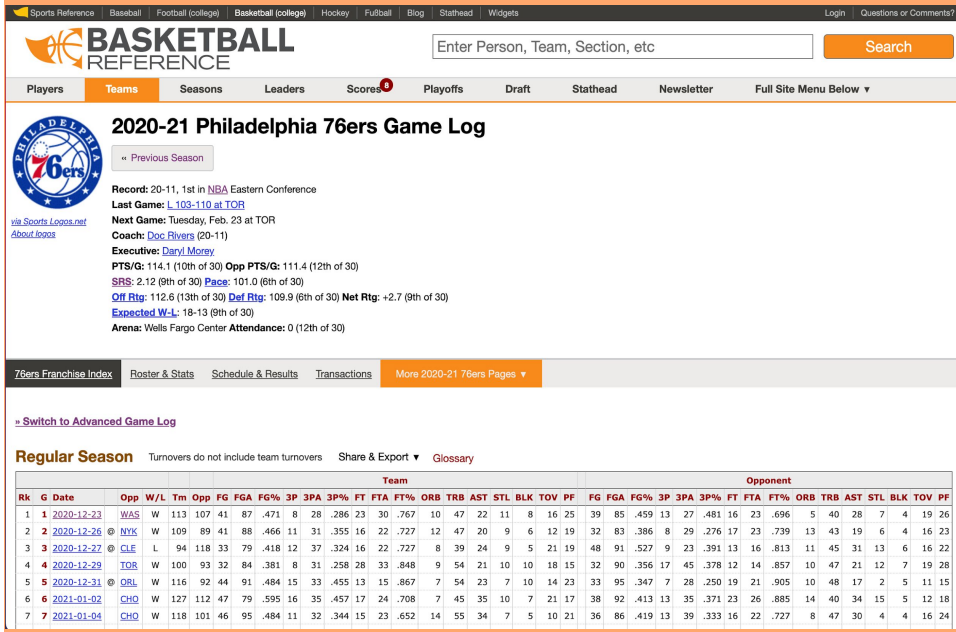
Learn from the models

Further Work

Come up with plans for the next steps in the project

Game Logs

- All 30 NBA teams
- 2009-2010 season up until 2019-2020 season
- Basic game statistics and advanced game statistics



Sports Reference | Baseball | Football (college) | Basketball (college) | Hockey | Fußball | Blog | Stathead | Widgets — Login | Questions or Comments?

BASKETBALL REFERENCE

Enter Person, Team, Section, etc — Search

Players | Teams | Seasons | Leaders | Scores | Playoffs | Draft | Stathead | Newsletter | Full Site Menu Below ▾

2020-21 Philadelphia 76ers Game Log

« Previous Season

via Sports Logos.net
About logos

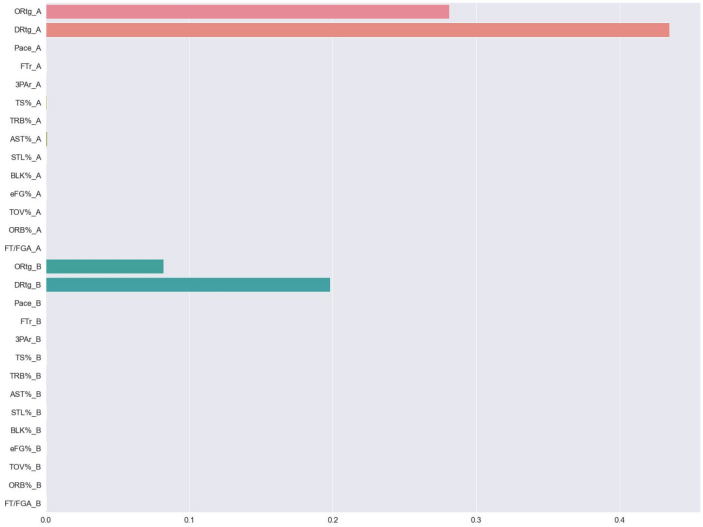
Record: 20-11, 1st in NBA Eastern Conference
Last Game: L 103-110 at TOR
Next Game: Tuesday, Feb. 23 at TOR
Coach: Doc Rivers (20-11)
Executive: Daryl Morey
PTS/G: 114.1 (10th of 30) **Opp PTS/G:** 111.4 (12th of 30)
SRS: 2.12 (9th of 30) **Pace:** 101.0 (6th of 30)
Off Rtg: 112.6 (13th of 30) **Def Rtg:** 109.9 (6th of 30) **Net Rtg:** +2.7 (9th of 30)
Expected W-L: 18-13 (9th of 30)
Arena: Wells Fargo Center **Attendance:** 0 (12th of 30)

76ers Franchise Index | Roster & Stats | Schedule & Results | Transactions | More 2020-21 76ers Pages ▾

» Switch to Advanced Game Log

Regular Season Turnovers do not include team turnovers Share & Export ▾ Glossary

						Team														Opponent											
Rk	G	Date		Opp	W/L	Tm	Opp	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	TRB	AST	STL	BLK	TOV	PF	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	TRB	AST	STL	BLK	TOV	PF
1	1	2020-12-23		WAS	W	113	107	41	87	.471	8	28	.286	23	30	.767	10	47	22	11	8	16	25	39	85	.459	13	27	.481	16	23	.696	5	40	28	7	4	19	26
2	2	2020-12-26	@	NYK	W	109	89	41	88	.466	11	31	.355	16	22	.727	12	47	20	9	6	12	19	32	83	.386	8	29	.276	17	23	.739	13	43	19	6	4	16	23
3	3	2020-12-27	@	CLE	L	94	118	33	79	.418	12	37	.324	16	22	.727	8	39	24	9	5	21	19	48	91	.527	9	23	.391	13	16	.813	11	45	31	13	6	16	22
4	4	2020-12-29		TOR	W	100	93	32	84	.381	8	31	.258	28	33	.848	9	54	21	10	10	18	15	32	90	.356	17	45	.378	12	14	.857	10	47	21	12	7	19	28
5	5	2020-12-31	@	ORL	W	116	92	44	91	.484	15	33	.455	13	15	.867	7	54	23	7	10	14	23	33	95	.347	7	28	.250	19	21	.905	10	48	17	2	5	11	15
6	6	2021-01-02		CHO	W	127	112	47	79	.595	16	35	.457	17	24	.708	7	45	35	10	7	21	17	38	92	.413	13	35	.371	23	26	.885	14	40	34	15	5	12	18
7	7	2021-01-04		CHO	W	118	101	46	95	.484	11	32	.344	15	23	.652	14	55	34	7	5	10	21	36	86	.419	13	39	.333	16	22	.727	8	47	30	4	4	16	24

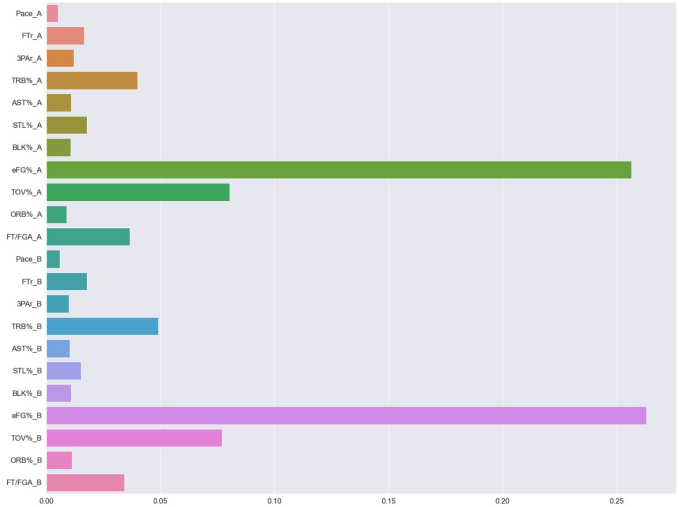


	Train_accuracy	Test_accuracy	F1_score	False_Negative	True_Positive
Decision Tree	100.00	99.62	1.00	9.0	3290.0
XGBoost	100.00	99.89	1.00	2.0	3297.0
Logistic Regression	100.00	100.00	1.00	0.0	3299.0
Support Vector Machine	99.99	100.00	1.00	0.0	3299.0
Random Forest	100.00	99.07	0.99	25.0	3274.0
K-Nearest Neighbors	98.53	97.27	0.97	77.0	3222.0

Findings

- Offensive Rating and Defensive Rating have complete priority

- Models' accuracy perfect/nearly perfect



	Train_accuracy	Test_accuracy	F1_score	False_Negative	True_Positive
XGBoost	99.95	94.54	0.95	196.0	3103.0
RS XGBoost	99.96	94.82	0.95	175.0	3124.0
Random Forest	100.00	90.85	0.91	307.0	2992.0
Random Search Random Forest	99.55	91.05	0.91	290.0	3009.0
Logistic Regression	87.55	87.83	0.88	402.0	2897.0
Support Vector Machine	83.53	84.10	0.84	542.0	2757.0
Decision Tree	81.96	80.36	0.81	579.0	2720.0
K-Nearest Neighbors	81.89	70.97	0.71	963.0	2336.0

New Findings

(Without Offensive and Defensive ratings)

- Other metrics find new life within the models

- Models no longer perfect… simply nearly

Further Work

Last 10 games statistics

Aggregate statistics on a last 10 games level to avoid running into data issues

Betting lines

Use point spreads as a next level for the model to predict

Artificial Neural Network

Implement a neural network model into the project to help with predictions

Thank You!

Does anyone have any questions?

Ripton.Rosen@gmail.com