Can Machine Learning Predict the Outcome of NBA Games?

Vaibhav Malik

Machine Learning Techniques

1) Regression will be used to create a model for the amount of points a team scores based on the team's average stats prior to a game
2) Classification and regression will then be used to predict whether a team will win or lose their next game based on different factors such as home court advantage

Some Things to Know Before Getting Started

Conventional NBA wisdom: more big players= more wins

Big players=get closer to the basket=higher chance of scoring

Then…

Golden State Warriors win championship 4 years ago:

Shoot high volume of 3s + rarely keep big players on the court

Choosing a Dataset

Regular Season

Turnovers do not include team turnovers   Share & more ▾   Glossary

Scroll Right For More Stats · *Switch to Widescreen View* ▶

								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
1	1	2015-10-27		DET	L	94	106	37	82	.451	8	27	.296	12	15	.800	7	40	22	9	4	15	25	37	96	.385	12	29	.414	20	26	.769	23	59	23
2	2	2015-10-29	@	NYK	W	112	101	42	83	.506	10	24	.417	18	26	.692	7	39	26	11	4	15	18	38	93	.409	6	29	.207	19	21	.905	16	48	21
3	3	2015-10-30		CHO	W	97	94	36	83	.434	8	23	.348	17	22	.773	8	45	23	9	4	15	17	36	88	.409	12	30	.400	10	13	.769	13	54	23
4	4	2015-11-01	@	CHO	W	94	92	37	88	.420	7	29	.241	13	14	.929	9	48	22	9	6	11	16	32	86	.372	14	37	.378	14	15	.933	9	45	27
5	5	2015-11-03	@	MIA	W	98	92	37	90	.411	7	21	.333	17	22	.773	16	51	25	10	1	15	14	38	86	.442	5	28	.179	11	16	.688	13	49	13
6	6	2015-11-04		BRK	W	101	87	37	76	.487	8	22	.364	19	24	.792	6	34	27	15	10	13	17	36	83	.434	5	21	.238	10	13	.769	14	43	24
7	7	2015-11-06	@	NOP	W	121	115	41	77	.532	9	21	.429	30	35	.857	6	41	30	10	6	15	21	42	84	.500	10	30	.333	21	24	.875	6	32	30
8	8	2015-11-07		WAS	W	114	99	43	92	.467	13	33	.394	15	15	1.000	10	39	37	12	4	14	15	38	78	.487	8	24	.333	15	17	.882	8	41	27
9	9	2015-11-09		MIN	L	107	117	44	86	.512	13	31	.419	6	7	.857	4	27	30	13	6	17	19	46	80	.575	7	15	.467	18	22	.818	8	40	27
10	10	2015-11-11		NOP	W	106	98	42	86	.488	10	26	.385	12	16	.750	8	44	26	10	7	12	16	37	83	.446	14	29	.483	10	15	.667	8	40	21
11	11	2015-11-13	@	BOS	L	93	106	36	76	.474	6	20	.300	15	20	.750	5	35	26	3	2	16	14	42	103	.408	11	33	.333	11	12	.917	17	50	27
12	12	2015-11-15		UTA	L	96	97	37	76	.487	8	22	.364	14	17	.824	8	33	26	9	5	15	16	39	76	.513	7	21	.333	12	17	.706	10	40	21
13	13	2015-11-17	@	BRK	L	88	90	33	76	.434	12	29	.414	10	14	.714	8	44	22	7	7	20	17	37	84	.440	9	20	.450	7	13	.538	9	44	22
14	14	2015-11-18		SAC	W	103	97	38	89	.427	9	30	.300	18	24	.750	12	45	23	9	6	11	15	38	86	.442	5	16	.313	16	18	.889	13	52	26
15	15	2015-11-21	@	CLE	L	97	109	32	86	.372	11	28	.393	22	25	.880	12	38	22	10	4	12	20	41	85	.482	11	29	.379	16	19	.842	11	51	27
16	16	2015-11-24		BOS	W	121	97	45	80	.563	12	24	.500	19	22	.864	6	43	33	8	7	20	25	34	82	.415	9	21	.429	20	26	.769	10	33	23
17	17	2015-11-25	@	MIN	L	95	99	31	82	.378	8	31	.258	25	35	.714	15	53	14	10	7	14	20	34	82	.415	9	21	.429	22	29	.759	7	45	22
18	18	2015-11-27	@	MEM	W	116	101	35	75	.467	14	39	.359	32	36	.889	6	40	25	4	3	10	21	35	80	.438	12	30	.400	19	21	.905	5	37	20
19	19	2015-11-28	@	SAS	L	88	108	37	87	.425	6	26	.231	8	13	.615	12	43	28	9	5	18	17	40	85	.471	9	17	.529	19	20	.950	12	49	25
20	20	2015-11-30		OKC	W	106	100	39	90	.433	10	28	.357	18	20	.900	14	51	23	9	8	12	15	37	93	.398	7	19	.368	19	22	.864	13	48	19

- Data obtained from basketball-reference website
- Only data from past 3 seasons will be used
- Discrepancy in tactics pre and post GSW championship

Stats Used in Points Scored Regression Model

3PA= Number of 3 point shots a team attempts

3P%= Percentage of 3 point shots made

OFFENSIVE RATING= Advanced metric measuring how efficiently a team scores points

OPPONENT DEFENSIVE RATING=Advanced metric measuring how efficiently the opponent prevents other teams from scoring points

Accuracy of Points Scored Predictive Models

```
Residuals:
    Min      1Q  Median      3Q     Max 
-32.962  -8.042   0.269   8.227  37.387 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  92.6893    49.0396   1.890   0.0600 .
V8            0.2123     0.4194   0.506   0.6132  
V9           -3.9415    41.0413  -0.096   0.9236  
V14          -0.4120     0.4519  -0.912   0.3629  
V15           0.4633     0.1995   2.322   0.0211 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.09 on 238 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.03101, Adjusted R-squared:  0.01472 
F-statistic: 1.904 on 4 and 238 DF,  p-value: 0.1105
```

- R-squared ↓↓↓
- Form=meaningless
- Too many unquantifiable factors

Significant Factors

3PA: Significant in 33.3% of models

3P%: Significant in 20% of models

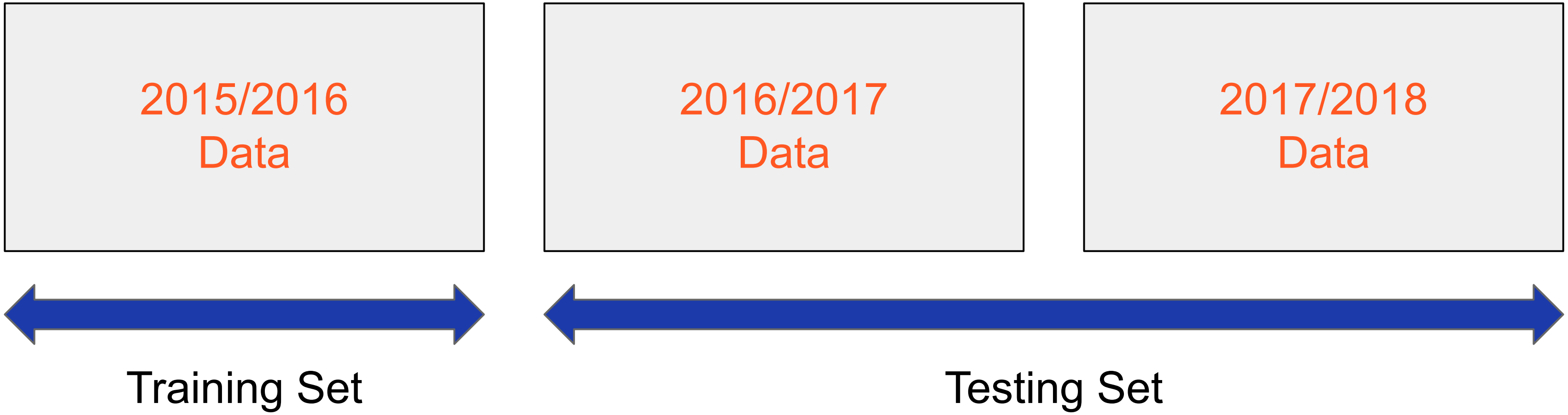
OFFENSIVE RATING:

Significant in 6.7% of models

OPPONENT DEFENSIVE RATING:

Significant in 90% of models

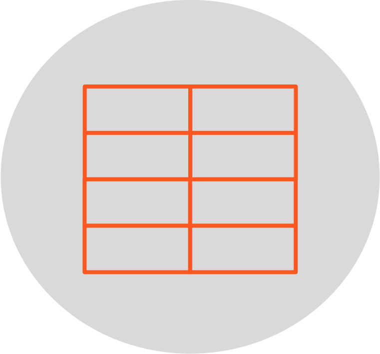
Predicting Wins and Losses Classification



Optimal model in first iteration correctly predicts win/loss with 61.8% accuracy

Improving the Model

Download More Stats



Restructure Data

Test Different Combinations

Loop through groups of stats

↓

Loop through every combination of stats in that group

↓

Cross validate by randomly choosing training and testing sets and then finding average accuracy

↓

Maximum true accuracy: 63%

Predicting Wins and Losses Regression

Team's significant stats = WIN 1

Team's significant stats = LOSS 0

Regression equation with stats as variables

Plug in team’s stats into regression equation to output “win probability”

↓

Team with higher win probability is predicted to win the game

↓

Compare predictions to actual results

↓

Prediction accuracy: 66%