

NBA Game Prediction and Season Simulation

MSc Thesis project for Teno González Dos Santos, supervisor Patrick Lambrix



Scoring:

- Free Throws → 1 point
- Inside 7.24m area → 2 points
- Outside 7.24m area → 3 points

Timing:

- 24 seconds / play
- 4 quarters
- **12** minutes / quarter

Background





- 30 teams
- 2 conferences
- 82 games / team
- 16 teams to POs

Background





- 8 teams / conference
- Best out of 7 games
- Format: 2-2-1-1-1





Two questions:

- 1. Are the features from **pre-season** as **significant** as the ones from **in-season** when predicting the outcome of a single game?
- 2. Do **probabilities** (for the championship or reaching different play-offs stages) assigned **by simulation** reflect what has occurred in **reality** in a sufficiently accurate manner?

Framework

Phases

- 1. Preprocessing: constructing features.
- 2. Model Selection: obtaining the optimal model.
- 3. Simulation: constructing final probabilities.



Framework: Preprocessing



Framework: Model Selection



Framework: Simulation



Raw Data: Web scraping and Pre-processing

- Extracting information contained online
- Unifying team + player names
- Merging different sources of information
- Removing unnecessary information



Feature Extraction: Box Scores



Golden State Warriors

121

103

17-38

Dallas Mavericks (17-38) Share & more V Glossary

		Basic Box Score Stats																		
Starters	МР	FG	FGA	FG%	ЗР	ЗРА	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	тоу	PF	PTS	+/-
Maxi Kleber	20:03	1	6	.167	0	3	.000	0	0		2	3	5	0	0	1	1	2	2	-29
Yogi Ferrell	30:42	3	9	.333	2	5	.400	1	2	.500	0	2	2	1	1	0	3	1	9	-25
<u>Salah Mejri</u>	16:37	5	10	.500	0	1	.000	0	0		3	2	5	1	0	2	1	5	10	-24
Dennis Smith	30:19	8	17	.471	4	9	.444	2	2	1.000	1	1	2	3	2	1	2	1	22	-19
Kyle Collinsworth	19:08	1	4	.250	0	0		0	0		0	0	0	1	1	0	1	3	2	-14
Wesley Matthews	28:52	7	17	.412	2	9	.222	1	2	.500	0	2	2	2	2	0	1	2	17	-4
Jalen Jones	6:48	0	1	.000	0	1	.000	1	2	.500	0	1	1	0	1	0	0	0	1	-1
Dwight Powell	31:23	7	9	.778	0	0		4	4	1.000	3	3	6	2	2	0	0	1	18	+6
J.J. Barea	28:11	2	9	.222	0	5	.000	2	2	1.000	0	0	0	8	1	1	1	0	6	+9
<u>Dirk Nowitzki</u>	27:57	6	10	.600	1	4	.250	3	3	1.000	0	11	11	1	5	2	0	1	16	+11
Johnathan Motley	Did Not Play																			
Josh McRoberts		Did Not Play																		
Team Totals	240	40	92	.435	9	37	.243	14	17	.824	9	25	34	19	15	7	10	16	103	

www.basketball-reference.com

Feature Extraction: Box Scores

- Home Team
- Season Stage
- League Record
- Games Played
- Back to back

- Total Wins in the last 3, 8 and 15 games
- Home Wins in the last 3, 8 and 15 games
- Away Wins in the last 3, 8 and 15 games

Feature Extraction: Previous Year's Stats

Team Misc

								Adva	anced	Offense Four Factors			Defense Four Factors									
	w	L	PW	PL	ΜΟΥ	SOS	SRS	ORtg	DRtg	Pace	FTr	3PAr	eFG%	τον%	ORB%	FT/FGA	eFG%	τον%	DRB%	FT/FGA	Arena	Attendance
Team	47	35	49	33	2.89	0.00	2.89	107.9	104.8	95.0	.245	.282	.507	12.2	23.7	.189	.507	13.2	78.6	.170	AT&T Center	754,562
Lg Rank	12	18	8	8	8	17	8	17	3	28	17	27	26	5	6	18	8	12	8	3		12

www.basketball-reference.com

Feature Extraction: Previous Year's Stats

- Previous year's furthest stage
- Regular Season record
- Offensive Rating
- Defensive Rating

- Four factors:
 - eFG% (effective field goal %)
 - TOV% (turnover %)
 - ORB% (offensive rebound %)
 - DRB% (defensive rebound %)
 - FT/FGA ratio

(free throws / field goal attempts)

Feature Extraction: Players' Performance in the Previous Year

Rk		Age	G	MP	PER	TS%	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	ows	DWS	ws	WS/48	ОВРМ	DBPM	ВРМ	VORP
1	Klay Thompson	27	73	2506	16.0	.598	.442	.083	1.5	10.3	6.2	11.2	1.1	1.1	9.5	23.7	3.1	1.8	4.9	.094	1.8	-2.4	-0.6	0.9
2	Kevin Durant	29	<u>68</u>	2325	26.0	.640	.338	.331	1.6	19.5	11.2	25.5	1.0	4.0	12.9	30.4	7.5	2.9	10.4	.215	5.0	0.7	5.6	4.5
3	Draymond Green	27	ZQ	2287	16.1	.556	.414	.279	3.9	21.1	13.1	29.0	2.0	3.1	22.5	17.2	2.6	3.4	6.1	.127	0.4	2.7	3.2	3.0
4	Stephen Curry	29	5	1631	28.2	.675	.580	.350	2.7	14.4	9.0	30.3	2.4	0.4	13.3	31.0	7.2	1.9	9.1	.267	9.9	-1.3	8.6	4.4
5	Andre Iguodala	34	64	1622	11.2	.536	.366	.272	3.7	12.6	8.5	16.2	1.6	1.9	15.8	11.5	1.6	1.6	3.2	.094	-0.7	0.9	0.2	0.9
6	Nick Young	32	80	1393	10.0	.562	.668	.133	1.7	7.9	5.0	3.5	1.3	0.4	7.0	17.5	0.8	0.9	1.7	.059	-0.2	-3.4	-3.7	-0.6
7	Shaun Livingston	32	Z.	1129	11.6	.533	.015	.178	3.4	9.2	6.5	16.8	1.5	1.5	13.6	16.6	0.9	0.9	1.8	.077	-2.6	-0.9	-3.5	-0.4
8	David West	37	73	999	20.9	.600	.021	.209	7.7	18.2	13.3	20.5	2.3	5.9	16.2	21.7	1.8	1.7	3.5	.168	-0.2	3.2	3.0	1.3
9	Zaza Pachulia	33	<u>69</u>	972	17.0	.612	.004	.352	11.0	25.0	18.5	15.2	1.9	1.4	19.1	17.0	1.8	1.4	3.3	.161	-0.5	1.3	0.8	0.7
10	Patrick McCaw	22	57	961	7.4	.477	.356	.151	2.0	7.2	4.8	10.6	2.2	0.8	14.6	12.8	-0.3	0.8	0.6	.029	-2.9	-0.5	-3.4	-0.3
11	Kevon Looney	21	<u>66</u>	910	14.8	.588	.026	.399	11.0	15.2	13.2	6.0	1.8	4.9	12.0	12.4	1.5	1.3	2.8	.145	-0.9	2.1	1.2	0.7
12	Jordan Bell	23	<u>57</u>	809	18.0	.641	.022	.238	8.9	19.0	14.3	16.9	2.1	5.5	20.0	13.9	1.6	1.4	2.9	.174	0.0	3.7	3.7	1.2
13	Quinn Cook	24	33	740	12.8	.593	.413	.099	1.6	10.2	6.2	16.6	0.8	0.1	10.8	17.5	1.1	0.4	1.6	.101	0.6	-2.8	-2.2	0.0
14	Omri Casspi	29	<u>5</u> 3	740	15.5	.632	.106	.333	4.9	24.0	15.1	9.5	1.2	2.0	13.5	16.3	1.2	1.0	2.2	.141	-1.5	0.0	-1.4	0.1
15	JaVale McGee	30	<u>65</u>	615	22.3	.641	.027	.237	11.5	18.7	15.4	8.0	1.6	7.3	10.0	19.2	1.7	1.1	2.7	.212	0.4	2.1	2.6	0.7
16	Damian Jones	22	15	89	6.8	.517	.000	.227	6.8	10.6	8.8	2.9	0.5	2.7	14.2	13.9	0.0	0.1	0.1	.031	-4.9	-1.7	-6.6	-0.1
17	Chris Boucher	25	1	1	-28.5	.000	1.000	.000	0.0	100.0	55.9	0.0	0.0	0.0	0.0	43.9	0.0	0.0	0.0	-0.899	-38.9	-20.1	-59.0	0.0

www.basketball-reference.com

Feature Extraction: Players' Performance in the Previous Year

$$eWS_{48} = WS_{48} \cdot \frac{\frac{Minutes Played}{\# games per season}}{\# \min s per game}$$

Win Share: estimate of the number of wins contributed by a player per 48 minutes



Feature Extraction: Players' Performance in the Previous Year

- Staying players weighted mean of eWS₄₈
- Leaving players weighted mean of eWS₄₈
- Signed players weighted mean of eWS₄₈

Feature Extraction: Previous Draft Picks

				Round 1
Rk	Pk	Tm	Player	College
1	1	PHO	Deandre Ayton	University of Arizona
2	2	SAC	Marvin Bagley	Duke University
3	3	ATL	Luka Doncic	
4	4	MEM	Jaren Jackson	Michigan State University
5	5	DAL	Trae Young	University of Oklahoma
6	6	ORL	Mohamed Bamba	University of Texas at Austin
7	Z	CHI	Wendell Carter	Duke University
8	8	CLE	Collin Sexton	University of Alabama
9	2	NYK	Kevin Knox	University of Kentucky
10	10	PHI	Mikal Bridges	Villanova University
11	11	<u>CHO</u>	Shai Gilgeous-Alexander	University of Kentucky
12	12	LAC	Miles Bridges	Michigan State University
13	<u>13</u>	LAC	Jerome Robinson	Boston College
14	<u>14</u>	DEN	Michael Porter	University of Missouri
15	15	WAS	Troy Brown	University of Oregon
16	<u>16</u>	PHO	Zhaire Smith	Texas Tech University
17	17	MIL	Donte DiVincenzo	Villanova University
18	18	SAS	Lonnie Walker	University of Miami
19	<u>19</u>	ATL	Kevin Huerter	University of Maryland
20	20	MIN	Josh Okogie	Georgia Institute of Technology



Presented by **& State Farm**[®]

www.basketball-reference.com

Feature Extraction: Previous Draft Picks

- # Picks in positions 1-3
- # Picks in positions 4-10
- # Picks in positions 11-20
- # Picks in positions 20- end of 1st Round
- # Picks in 2nd Round



Presented by 🉈 State Farm[®]

Feature Extraction: Players' Salaries and Salary Cap

Salaries

Rk Salary Magic Johnson \$2,500,000 1 Kareem Abdul-Jabbar \$2,030,000 2 Mitch Kupchak \$1,150,000 3 James Worthy \$550,000 4 Maurice Lucas \$475,000 5 **Kurt Rambis** \$450,000 6 7 **Byron Scott** \$420,000 Michael Cooper \$366,000 8 9 Mike McGee \$275,000 Larry Spriggs \$135,000 10 11 A.C. Green \$128,750 12 Ronnie Lester \$100,000

Salary Cap History Share & more v

Year	Salary Cap	Salary Cap (2015)
<u>1984-85</u>	\$3,600,000	\$7,934,034
<u>1985-86</u>	\$4,233,000	\$9,153,509
<u>1986-87</u>	\$4,945,000	\$10,317,292
<u>1987-88</u>	\$6,164,000	\$12,354,015
<u>1988-89</u>	\$7,232,000	\$13,829,137
<u>1989-90</u>	\$9,802,000	\$17,783,913
<u>1990-91</u>	\$11,871,000	\$20,658,962
<u>1991-92</u>	\$12,500,000	\$21,113,943
<u>1992-93</u>	\$14,000,000	\$22,972,126
<u>1993-94</u>	\$15,175,000	\$24,268,416
<u>1994-95</u>	\$15,964,000	\$24,833,276
<u>1995-96</u>	\$23,000,000	\$34,761,086
<u>1996-97</u>	\$24,363,000	\$35,973,807
<u>1997-98</u>	\$26,900,000	\$39,119,783
<u>1998-99</u>	\$30,000,000	\$42,686,775
<u>1999-00</u>	\$34,000,000	\$46,798,454
2000-01	\$35,500,000	\$47,522,364
2001-02	\$42,500,000	\$56,008,788
2002-03	\$40,271,000	\$51,890,436
2003-04	\$43,840,000	\$55,016,264

www.basketball-reference.com

2017-18: Salary cap: 99 M\$ Luxury tax: 119 M\$

Feature Extraction: Players' Salaries and Salary Cap

- Total team salary
- Total Salary / Salary Cap ratio
- Top-1, Top-2, Top-3 and Top-5 of players' salary ratio



Logistic regression

- C=0.1
- 23 features selected



SVM models

0.70 Accuracies of Support Vector Machines Support Vector Machines 0.68 C=10.66 Accuracy 25 features selected 0.64 0.62 0.60 100 1000 2 \sim 2 C value

Random Forest

- Number of trees = 100
- Minimum number of samples in a leaf = 4
- Minimum number of samples in a Split = 5



Multilayer Perceptron

- Hidden Layer Sizes = [100]
- Alpha = 0.05
- Learning rate = 'constant'
- Solver = 'ADAM'
- Activation function = 'tahn'



Model selection: Cross-Validation of Best Models

Mean accuracies:

- Logistic Regression: 68.3%
- SVMs: 68.12%
- Random Forest: 69.68%
- Multilayer Perceptron: 68.96%





Top 50 important features





50 less important features





Simulation Probabilities

 Table 4.11: Comparison of simulation probabilities and true positions for season 2009/10.

	1 st	Conf.	Conf.	NBA	NBA	Doolity			
Team	Round	Semifinals	Finals	Finals	Champion	(fourth eat ato go)			
	(Model)	(Model)	(Model)	(Model)	(Model)	(furthest stage)			
LAL	86.4	73.1	43.9	27.9	16.6	NBA Champion			
CLE	86.5	58.0	39.8	24.3	16.2	Conf. Semifinals			
BOS	86.5	51.6	38.1	24.2	16.2	NBA Finals			
ORL	85.5	56.1	36.9	21.4	9.4	Conf. Finals			
ATL	86.3	64.5	35.7	18.4	9.0	Conf. Semifinals			
SAS	81.6	50.5	29.8	15.2	6.3	Conf. Semifinals			
DEN	78.8	57.0	29.8	13.6	5.1	1 st Round			
POR	67.5	46.1	19.8	9.3	5.5	1 st Round			
DAL	65.9	32.8	15.6	6.0	3.2	1 st Round			
MIA	58.9	30.6	13.1	9.6	3.1	1 st Round			
HOU	79.2	33.1	14.0	7.7	2.5	9 th West conf.			
UTA	69.0	26.1	12.4	5.8	1.9	Conf. Semifinals			
NOP	66.6	37.5	13.5	3.4	1.7	11 th West conf.			
РНО	54.0	21.7	8.0	3.1	1.1	Conf. Finals			
	Team LAL CLE BOS ORL ATL SAS DEN POR DAL MIA HOU UTA NOP PHO	1** Round (Model) LAL 86.4 CLE 86.5 BOS 86.3 BORL 85.5 ATL 86.3 DEN 78.8 POR 67.5 DAL 65.9 MIA 58.9 HOU 79.2 UTA 69.0 NOP 66.6 PHO 54.0	1st Conf. Round Semifinals (Model) (Model) LAL 86.4 73.1 CLE 86.5 58.0 BOS 86.5 51.6 ORL 85.5 56.1 ATL 86.3 64.5 SAS 81.6 50.5 DEN 78.8 57.0 POR 67.5 46.1 DAL 65.9 32.8 MIA 58.9 30.6 HOU 79.2 33.1 UTA 69.0 26.1 NOP 66.6 37.5 PHO 54.0 21.7	1st Conf. Conf. Round Semifinals Finals (Model) (Model) (Model) LAL 86.4 73.1 43.9 CLE 86.5 58.0 39.8 BOS 86.5 51.6 38.1 ORL 85.5 56.1 36.9 ATL 86.3 64.5 35.7 SAS 81.6 50.5 29.8 DEN 78.8 57.0 29.8 POR 67.5 46.1 19.8 DAL 65.9 32.8 15.6 MIA 58.9 30.6 13.1 HOU 79.2 33.1 14.0 UTA 69.0 26.1 12.4 NOP 66.6 37.5 13.5 PHO 54.0 21.7 8.0	1 st Conf. Conf. NBA Round Semifinals Finals Finals (Model) (Model) (Model) (Model) LAL 86.4 73.1 43.9 27.9 CLE 86.5 58.0 39.8 24.3 BOS 86.5 51.6 38.1 24.2 ORL 85.5 56.1 36.9 21.4 ATL 86.3 64.5 35.7 18.4 SAS 81.6 50.5 29.8 15.2 DEN 78.8 57.0 29.8 13.6 POR 67.5 46.1 19.8 9.3 DAL 65.9 32.8 15.6 6.0 MIA 58.9 30.6 13.1 9.6 HOU 79.2 33.1 14.0 7.7 UTA 69.0 26.1 12.4 5.8 NOP 66.6 37.5 13.5 3.4 PHO <t< td=""><td>Image: sec: sec: sec: sec: sec: sec: sec: se</td></t<>	Image: sec: sec: sec: sec: sec: sec: sec: se			



Simulation Probabilities

Table 4.12: Comparison of simulation probabilities and true positions reality for season 2010/11.

		1 st	Conf.	Conf.	NBA	NBA	Deality
	Team	Round	Semifinals	Finals	Finals	Champion	(fourth oat atogo)
		(Model)	(Model)	(Model)	(Model)	(Model)	(Iurtnest stage)
1	LAL	88.4	63.1	47.9	24.9	14.5	Conf. Semifinals
2	BOS	86.5	59.3	39.8	23.6	14.1	Conf. Semifinals
3	ORL	83.5	53.6	34.1	19.9	13.5	1 st Round
4	SAS	85.5	57.1	36.9	21.8	12.9	1 st Round
5	DAL	86.3	64.8	35.7	22.4	12.6	NBA Champion
6	MIA	81.6	51.3	29.8	18.1	11.9	NBA Finals
7	ATL	71.8	53.0	29.8	15.6	5.1	Conf. Semifinals
8	РНО	67.5	41.1	19.8	7.3	3.3	10 th West conf.
9	DEN	65.9	30.8	15.6	7.0	1.9	1 st Round
10	CLE	58.9	30.6	13.1	6.6	1.9	15 th East conf.
11	POR	69.2	33.1	14.0	5.7	1.8	1 st Round
12	OKC	58.3	26.1	12.4	5.8	1.3	Conf. Finals
13	UTA	68.6	37.5	13.5	7.4	1.3	11 th West conf.
14	NOP	54.0	21.7	8.0	3.1	1.0	1 st Round
15	СНІ	69.6	28.9	6.8	5.3	1.0	Conf. Finals

Conclusions

- Pre-season and in-season variables "work" together.
- Good performance of the model and improvements over some of previous studies.
- Game prediction + Simulation shows as a good approximation of reality.
- Struggles to capture **drastic changes** in team performances in consecutive years.

Future Work

- Using the output probabilities of this thesis (or the betting companies' probabilities) as the starting point.
- Develop a **model** that could **update** these **probabilities** while the season is being played.
- **Result:** a table of **probabilities** that reflect the progress of the teams in a more accurate manner during the season.