Predicting NBA Player Performance

Christopher Bui
Project 2 - Luther

Tristan Thompson 2017-18 Game Log

Tristan Trevor James Thompson ▪ Twitter: RealTristan13
(Double T)
Position: Center and Power Forward ▪ **Shoots:** ~~Left~~ Right
6-9, 238lb (206cm, 107kg)
Team: Cleveland Cavaliers
Born: March 13, 1991 (Age: 27-129d) in Toronto, Canada

More bio, uniform, draft, salary info ▾

SUMMARY	G	PTS	TRB	AST	FG%	FG3%	FT%	eFG%	PER	WS
2017-18	53	5.8	6.6	0.6	56.2	-	54.4	56.2	13.6	2.5
Career	519	9.0	8.4	0.8	51.8	0.0	60.8	51.8	15.2	37.4

Contract: **$16 Million / year** for 5 years

$100 Million **spending limit**

Objective

- Construct a reliable model for predicting player performance

- Outcome:
  - Potential for team organizations invest smartly
  - Save ***tens - hundreds of millions*** of dollars

Define the variables

- **Response Variable**
  - Performance = Points Per Game for a season

- **Features**
  - Games played, PTS, FG%, TRB, TOV, AST

Methods

- selenium, BeautifulSoup, statsmodels, sklearn
- Start from 1979 due to creation of 3 point line
- Averaged first 3 seasons for each player
- 4th season PTS = future values

Per Game

Share & more ▼ Glossary

Season	Age	Tm	Lg	Pos	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%
1996-97	18	LAL	NBA	SG	71	6	15.5	2.5	5.9	.417	0.7	1.9	.375
1997-98 ★	19	LAL	NBA	SG	79	1	26.0	4.9	11.6	.428	0.9	2.8	.341
1998-99	20	LAL	NBA	SG	**50**	50	37.9	7.2	15.6	.465	0.5	2.0	.267
1999-00 ★	21	LAL	NBA	SG	66	62	38.2	8.4	17.9	.468	0.7	2.2	.319

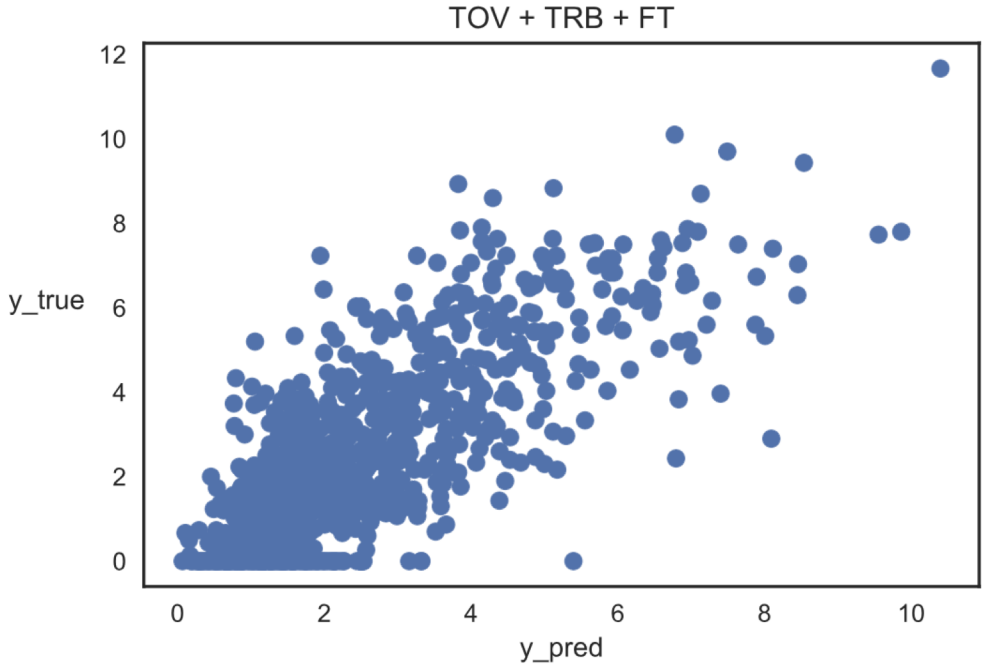
	NAME	G	GS	MP	FG	FGA	FG_pct	P3	PA3	pct_3P	...
0	Dalibor Bagaric	31.667	0.0	9.267	0.900	2.700	0.325	0.0	0.000	0.000	...
1	John Bagley	75.000	29.0	22.233	3.333	7.367	0.448	0.0	0.233	0.078	...
2	James Bailey	75.333	0.0	20.300	3.533	7.133	0.489	0.0	0.000	0.167	...

Assumption: Injured 4th season = 0 PTS

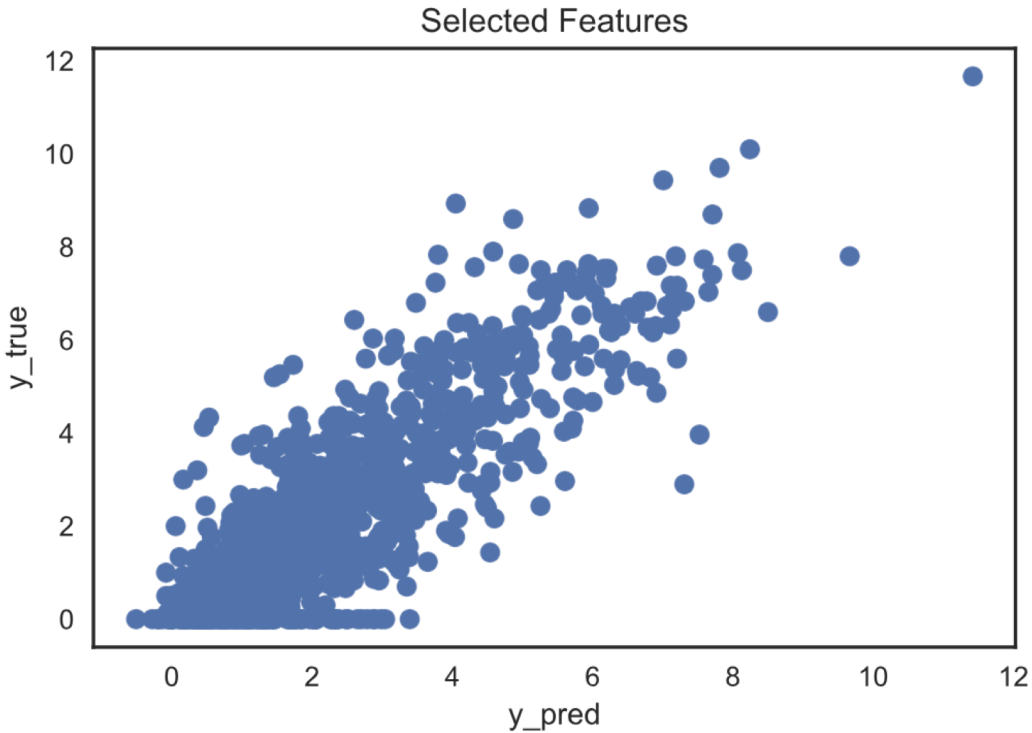
Selecting Key Features

- Initially include all features for Baseline
  - Select lowest P-values < 0.05
  - Reject extremely high P-values

- Selected Features:
  - 14/25
  - Took out steals, blocks, assists ...



Adj. R-squared: 0.529

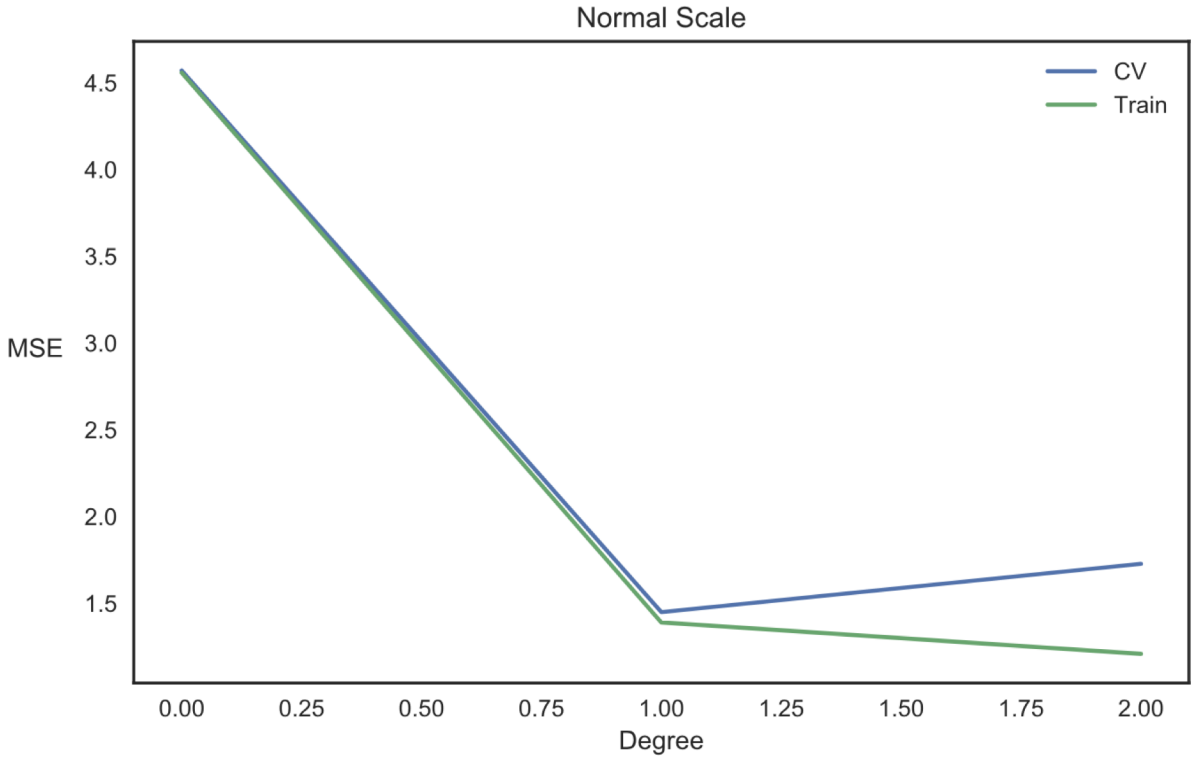


Adj. R-squared:	0.690

Determining Type of Fit

- Was there a better fit than linear?
- If so, see at what degree we would be underfitting or overfitting
- Started: Degree 3

```
The degree 1 had an MSE of: 1.44977046381
```



K-Fold Cross Validation

- 952 Rows
- Test Size = 30 %
- 10 Folds
- **Mean R^2** : 0.670 +/- =.06
- **Mean Square Error**: 1.449 +/- 0.212

Improvements to the model

- Get more relevant data not appearing on the stat sheet
  - Minutes of rest
  - Diet intake
  - Daily routine, etc.

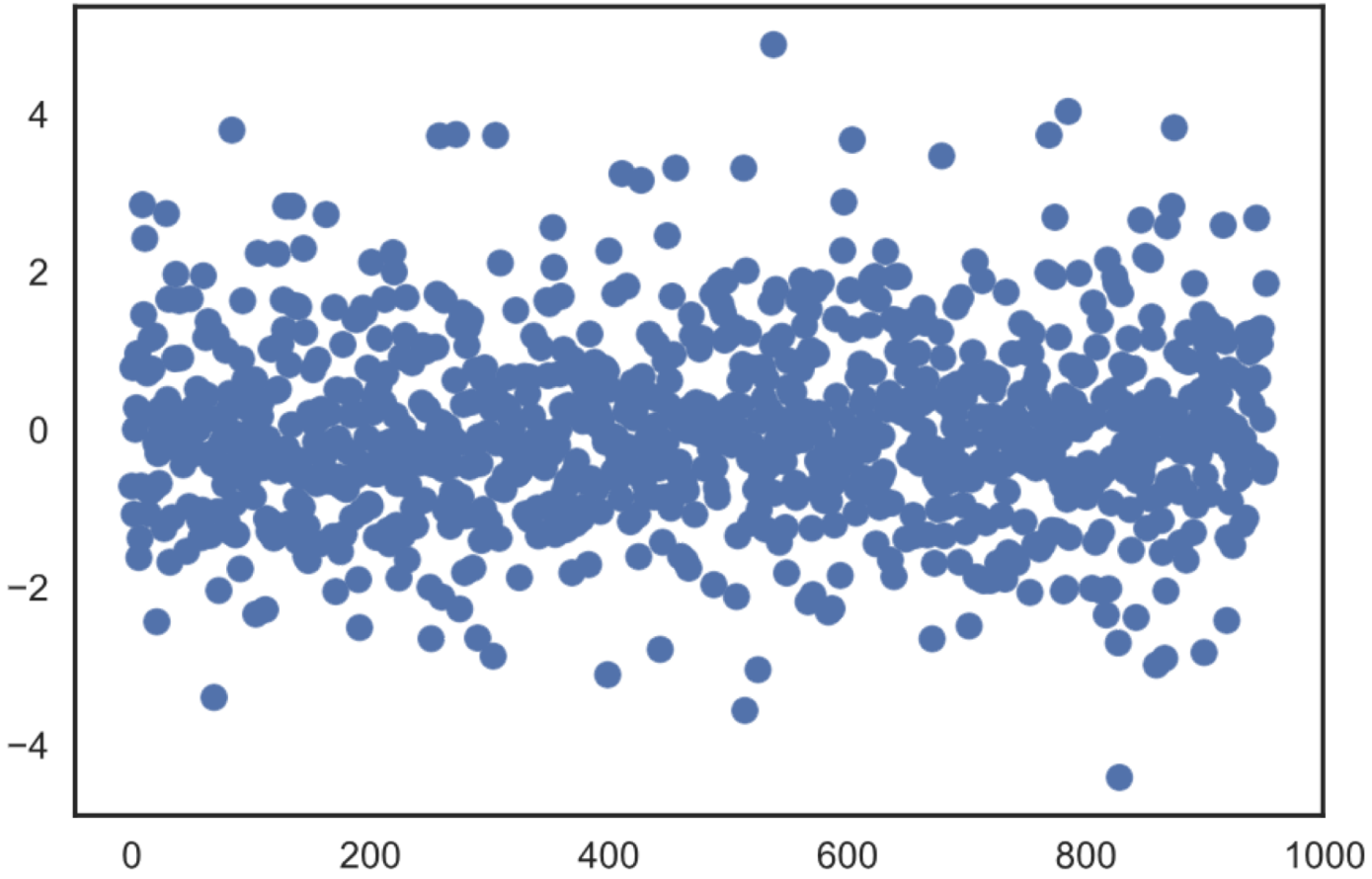
- Perhaps reduce the number of features given new data

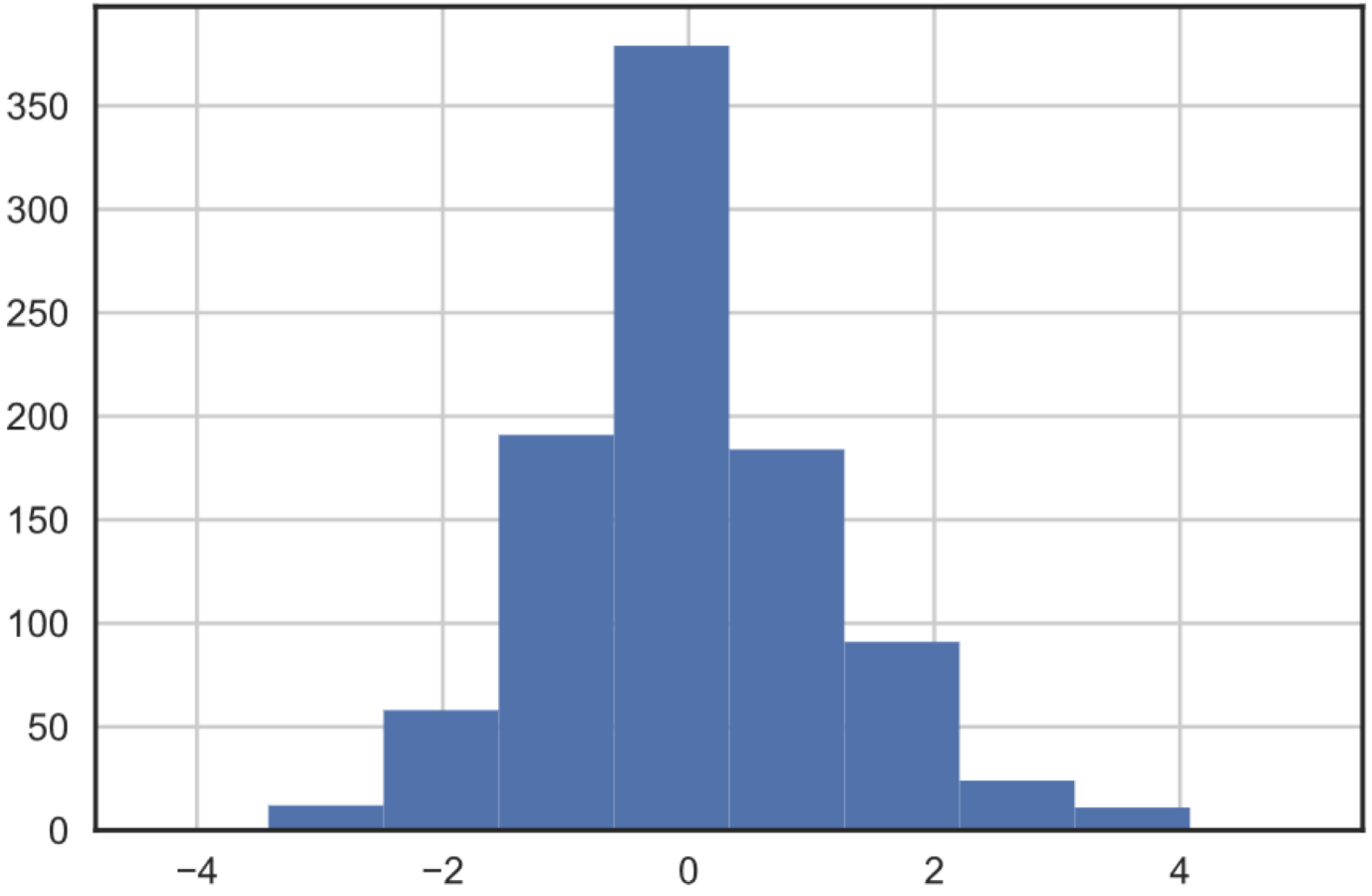
Future

- Gives a quantitative platform to reduce risk

- Teams can make sound business decisions in their interests
  - Attracts more business/sponsors/fans

Appendix


4
2
0
−2
−4
0
200
400
600
800
1000


350
300
250
200
150
100
50
0
−4
−2
0
2
4

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0227	0.253	-0.090	0.928	-0.520	0.474
PTS	0.5610	0.076	7.422	0.000	0.413	0.709
G	0.0170	0.003	5.658	0.000	0.011	0.023
FG_pct	2.8665	2.020	1.419	0.156	-1.098	6.831
FG	-0.3844	0.187	-2.056	0.040	-0.751	-0.018
MP	-0.0564	0.018	-3.140	0.002	-0.092	-0.021
ORB	-2.4015	1.378	-1.742	0.082	-5.106	0.303
DRB	-2.3329	1.377	-1.695	0.090	-5.034	0.369
TOV	-0.2902	0.130	-2.227	0.026	-0.546	-0.034
eFG_pct	-2.7790	1.988	-1.398	0.162	-6.680	1.122
TRB	2.4233	1.374	1.764	0.078	-0.273	5.120
STL	0.4896	0.158	3.092	0.002	0.179	0.800
PF	-0.2305	0.095	-2.417	0.016	-0.418	-0.043