NBA CONTRACT PREDICTION

Joonbum Yang (1005408186)
Mai Dang Khoi Nguyen (1005206199)
En-Chan Sing (1005512688)

yahoo!sports

Cristiano Ronaldo wants to leave Manchester United. But nobody wants Cristiano Ronaldo

Henry Bushnell

August 23, 2022 · 5 min read

In this article:

Manchester United Dec 21, 3:00 PM vs BRN	Cristiano Ronaldo | F | #7	Erik ten Hag Dutch association f...

Cristiano Ronaldo wants to leave Manchester United. He wanted to leave

How can we reduce inefficiency in NBA Contract Market?

Predicting Future Salary by Machine Learning Techniques

Contributions

To reduce market inefficiency in NBA market, we will adopt machine learning techniques to compare each players next salary contract between prediction. This will not only allow us to draw useful prediction in their salary based on their previous performances, it will also tell us whether if these players will be overpaid or underpaid.

The usage of machine learning models will be valuable in the research area, as the majority of current papers only use linear regression. Other papers in the research area also use missing/flawed datasets with lack of advanced metrics or players' characteristics in their model. Therefore, our paper will fill the missing gap in a research area.

Literature Review

The determinants of the salary in NBA and the overpayment in the year of signing a new contract (Li, Nuoya)

- Similarities: Performance statistics of the last 2 years before contract renewal, along with their personal characteristics.
- Differences: Only linear regression model is included. Advanced metrics are not included in the research. Data between 1995-2013

National Basketball Association Player Salary Prediction Using Supervised Machine Learning Methods (Emirhan Ozbalta, Mucahit Yavuz, Tolga Kaya)

- Similarities: Basic basketball performance statistics and the NBA 2K20 player ratings. Random forest and decision trees were used.
- Differences: Basketball statistics such as fouls or turnovers are ignored. Basketball players' in-game ratings as classifiers are flawed.

Estimating NBA players salary share according to their performance on court: A machine learning approach (Ioanna Papadaki, Michail Tsagris)

- Similarities: This paper proof that using linear regression is not ideal since there isn't a linear relationship. Lasso regression and random forest methods were used instead. The model used a lot of advanced metrics.
- Differences: Paper didn't exclude players already on contracts. Factors such as age or nationalities are also not included in their paper.

Data & Descriptive Statistics

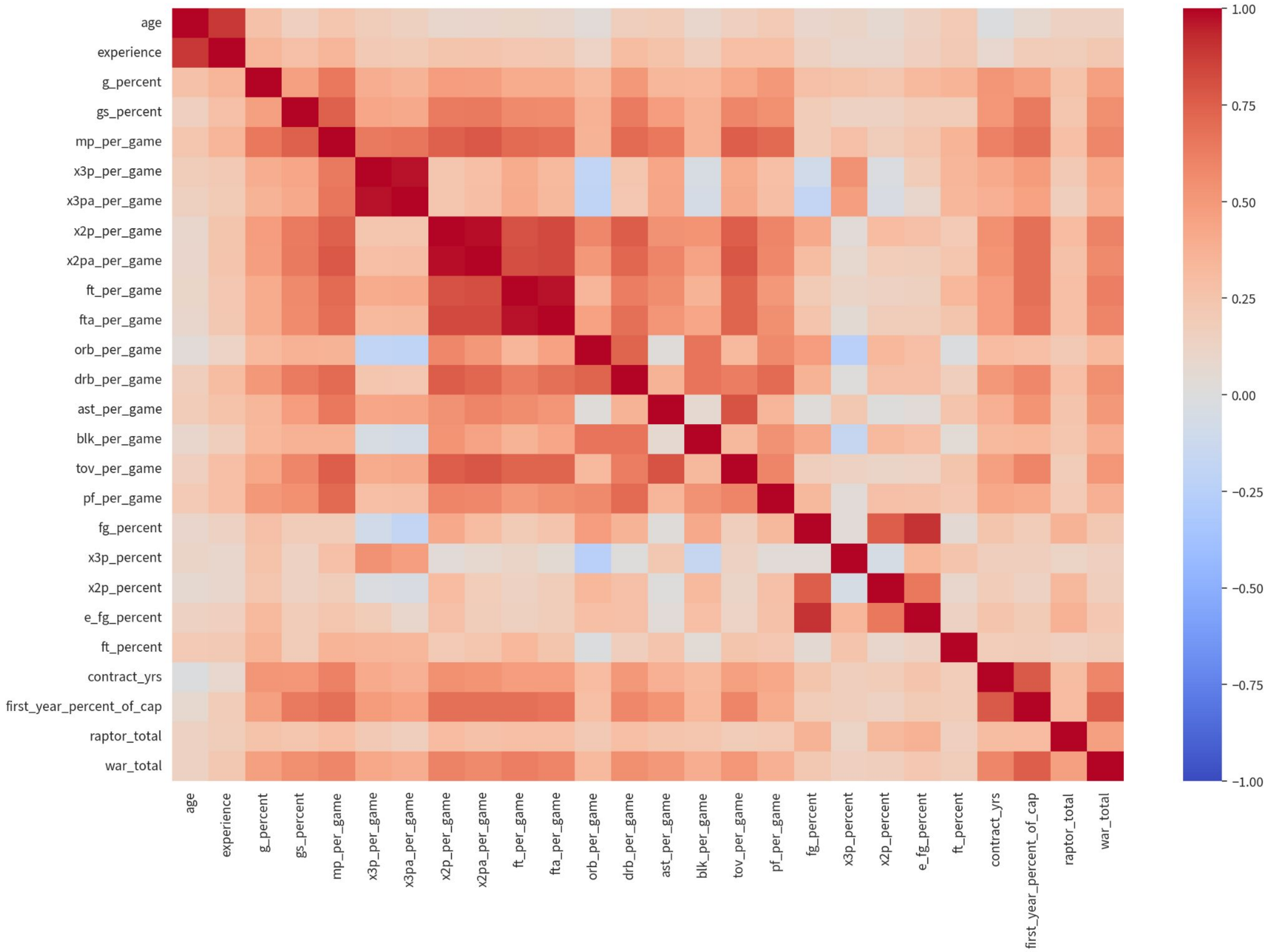
The data set consists of players who were elected free agency between 2016 and 2022, with traditional stats coming from Basketball-Reference and advanced metrics from FiveThirtyEight.

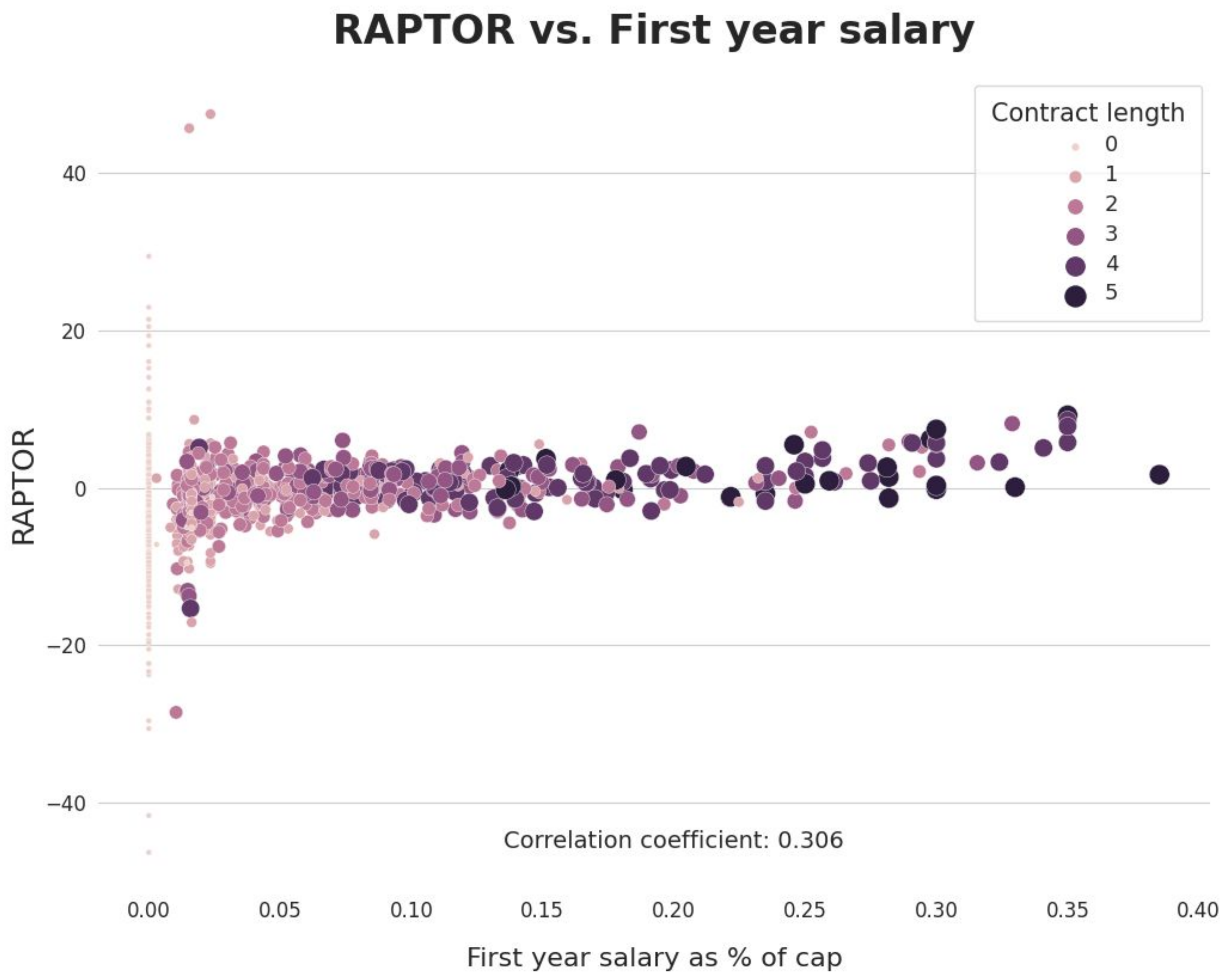
Dependent variable

- First year salary as a percentage of the salary cap

Independent variables

- Traditional box scores (ex. field goal percentage)
- Player's characteristics (ex: age, experience)
- Advanced metrics such as RAPTOR (takes advantage of play-by-play and player tracking data) and WAR (benchmarking against a theoretical replacement player)


age
experience
g_percent
gs_percent
mp_per_game
x3p_per_game
x3pa_per_game
x2p_per_game
x2pa_per_game
ft_per_game
fta_per_game
orb_per_game
drb_per_game
ast_per_game
blk_per_game
tov_per_game
pf_per_game
fg_percent
x3p_percent
x2p_percent
e_fg_percent
ft_percent
contract_yrs
first_year_percent_of_cap
raptor_total
war_total
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00



	First year salary	RAPTOR
count	1389.000000	1389.000000
mean	0.039650	-1.818240
std	0.064813	5.296957
min	0.000000	-46.307943
25%	0.000000	-3.845888
50%	0.015900	-1.472998
75%	0.044500	0.665477
max	0.384999	47.473611

Methods & Model Comparison

Decision Tree: By recursively splitting on predictors that will produce the most information gain, regression tree sets series of splits with each predictors that will lead to prediction of variable of interest.

Random Forest: Instead of using a single tree, random forest takes advantage of bootstrapping within training data and produce aggregated prediction of variable of interest.

Gradient Boosting: Technique which involves building a stronger model upon weaker-performing previous models to minimize its loss function and increase the predictive power.

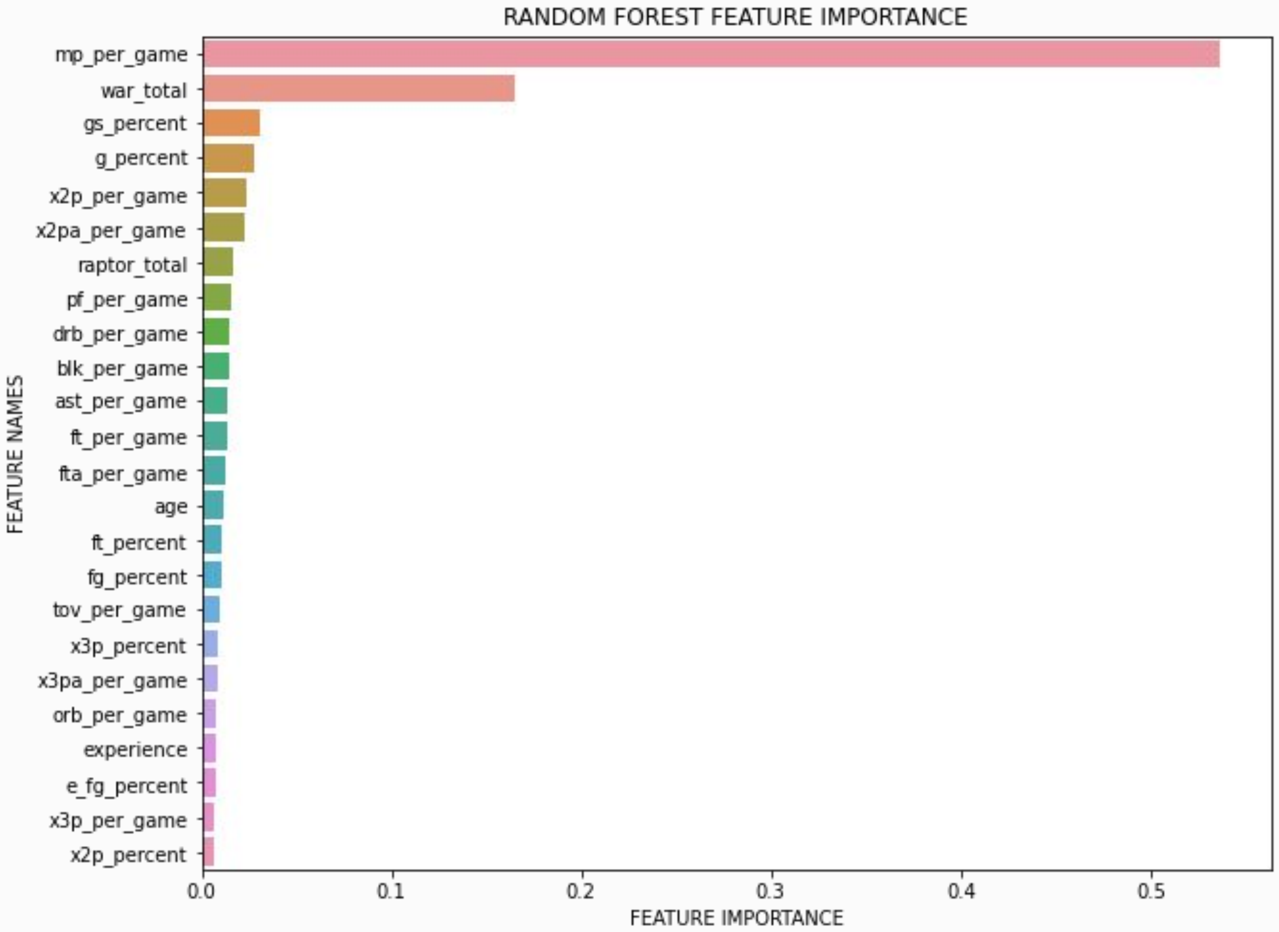
XGBoost: Model follows the principle of gradient boosting, but used more regularized model formalization to control over-fitting

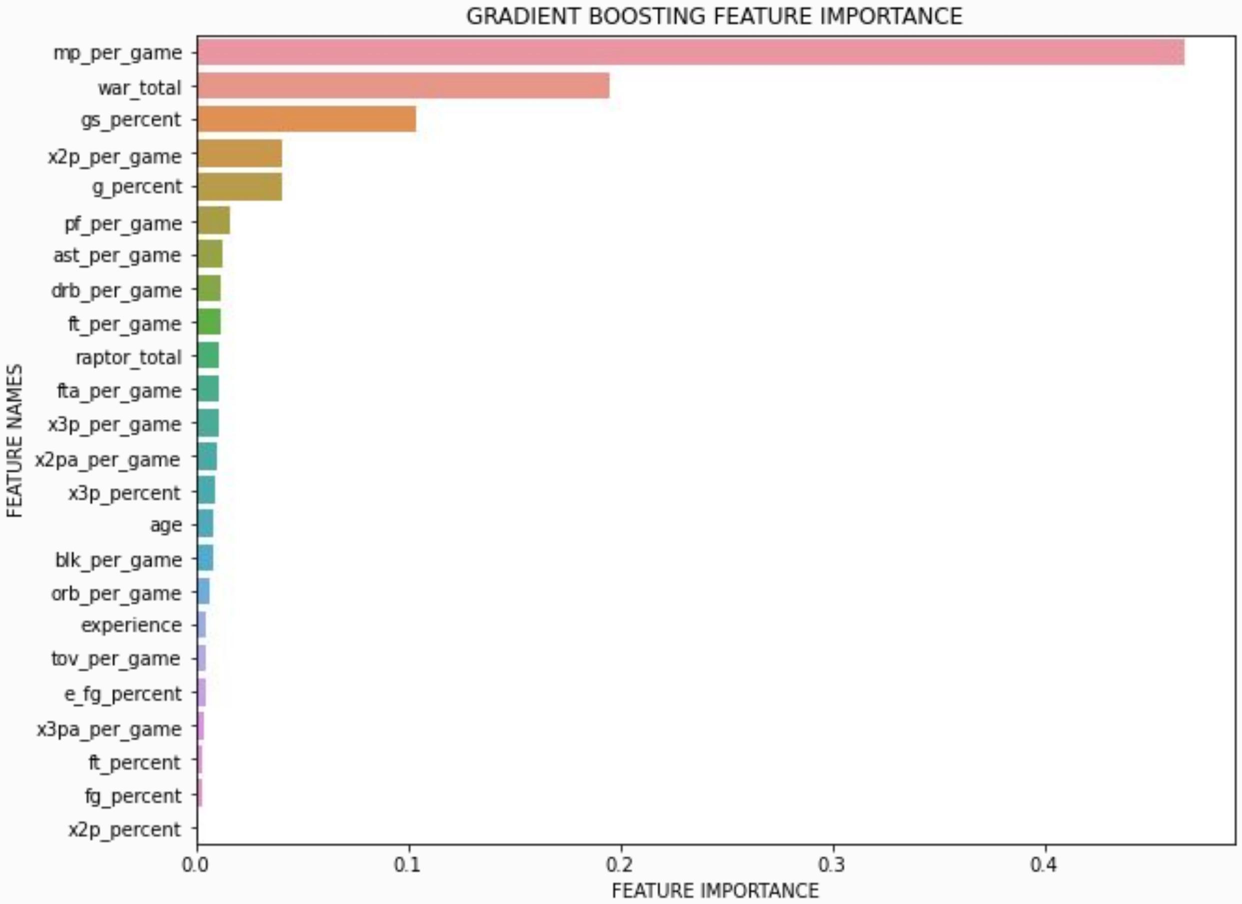
Methods & Model Comparison

	Predictor	Score	Mean Absolute Error	Mean Squared Error	Cross Validation Score (CVS)	CVS Mean	CVS Standard Deviation
0	Decision Tree	0.525186	0.025836	0.001873	[0.40764236967909473, 0.5229026201761826, 0.39...	0.440661	0.047654
1	Random Forest	0.776947	0.018468	0.000880	[0.7123533366840384, 0.7468880255816666, 0.653...	0.709921	0.030756
2	Gradient Boosting Classifier	0.776976	0.018648	0.000880	[0.6954943788319898, 0.7127173041276716, 0.660...	0.691659	0.017134
3	XGBoost Classifier	0.765872	0.018739	0.000923	[0.7340363767161733, 0.7206057083106109, 0.681...	0.713958	0.017972

Results

Feature Importance





Approximately 1.5 Billion USD (1390% of a team salary cap) are mismatched in the last 6 years

Top 5 Overvalued Players

	player	season	age	experience	first_year_percent_of_cap	predicted	diff
340	Timofey Mozgov	2016	29	6	0.170000	0.034011	0.135989
530	Otto Porter Jr.	2017	23	4	0.250000	0.126501	0.123499
87	Dirk Nowitzki	2016	37	18	0.265600	0.155813	0.109787
913	Klay Thompson	2019	28	8	0.300000	0.194008	0.105992
1243	Bradley Beal	2022	28	10	0.384999	0.292444	0.092555

Top 5 Undervalued Players

	player	season	age	experience	first_year_percent_of_cap	predicted	diff
1226	Victor Oladipo	2021	28	8	0.0213	0.136921	-0.115621
13	Andre Drummond	2021	27	9	0.0214	0.133593	-0.112193
622	Carmelo Anthony	2018	33	15	0.0235	0.129099	-0.105599
654	DeMarcus Cousins	2019	28	9	0.0321	0.122636	-0.090536
1339	Miles Bridges	2022	23	4	0.0000	0.088592	-0.088592

Interested Players: Overpaying Big Names?

player	season	age	experience	first_year_percent_of_cap	predicted	diff
Kawhi Leonard	2019	27	8	0.3000	0.299933	0.000067
Kawhi Leonard	2021	29	10	0.3500	0.335217	0.014783
Kevin Durant	2016	27	9	0.2819	0.289722	-0.007822
Kevin Durant	2017	28	10	0.2523	0.268168	-0.015868
Kevin Durant	2018	29	11	0.2945	0.296681	-0.002181
Kevin Durant	2019	30	12	0.3408	0.295192	0.045608
LeBron James	2016	31	13	0.3289	0.329576	-0.000676
LeBron James	2018	33	15	0.3500	0.328436	0.021564
Nikola Jokić	2018	22	3	0.2458	0.249462	-0.003662
Stephen Curry	2017	28	8	0.3500	0.330931	0.019069

Conclusion

Conclusion & Drawbacks

Conclusion

- The existence of market inefficiency is evidently clear, the biggest gap being 13% to 11% salary cap.
- Using machine learning to predict salary cap can reduce asymmetrical information for both parties

Drawbacks

- Previous Salary?
- Overpowering Variable
- Contract length

Thank you!