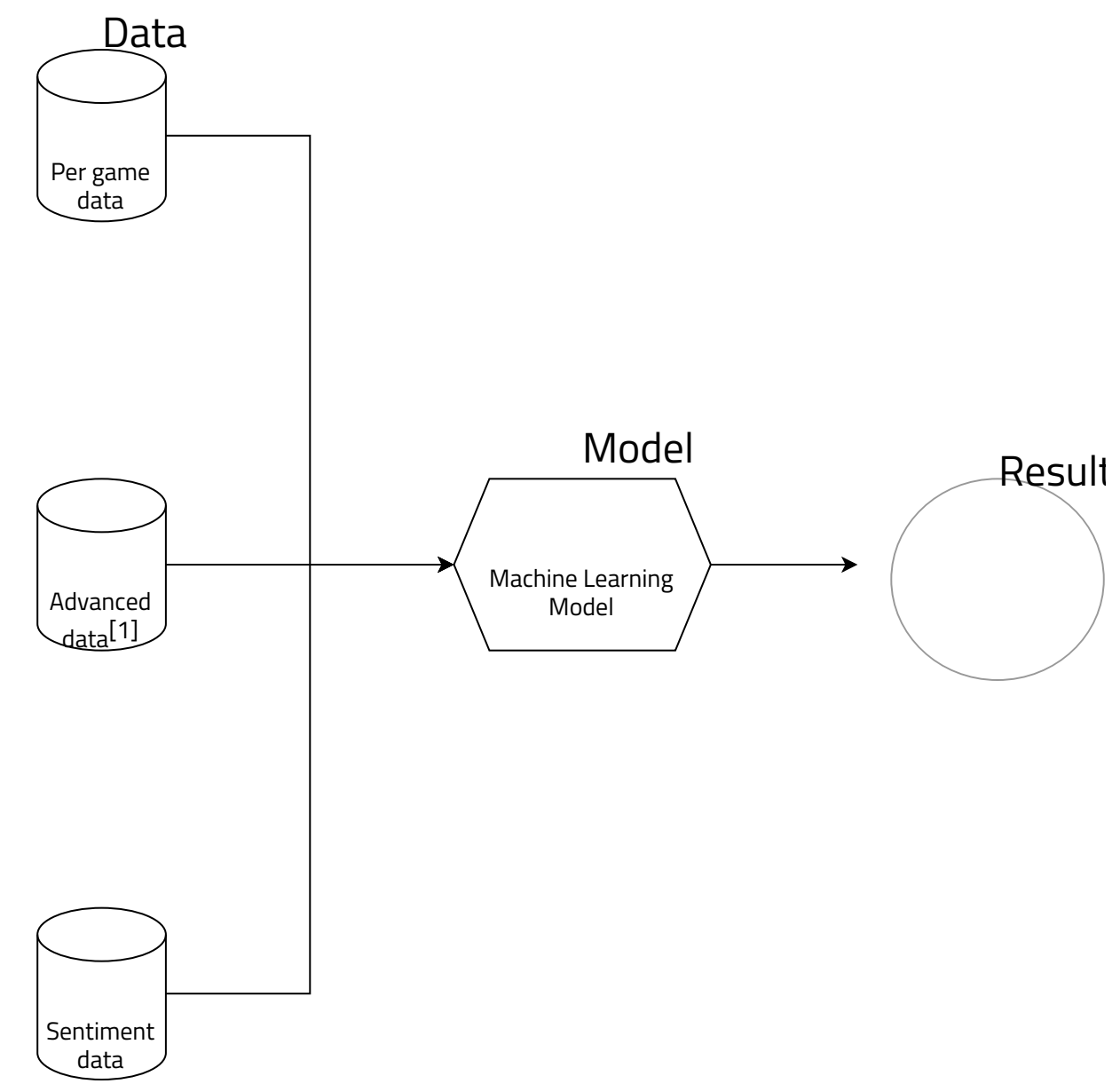


Machine Learning: Predicting the NBA MVP

Kobe Sarausad
6/4/2022

Mission



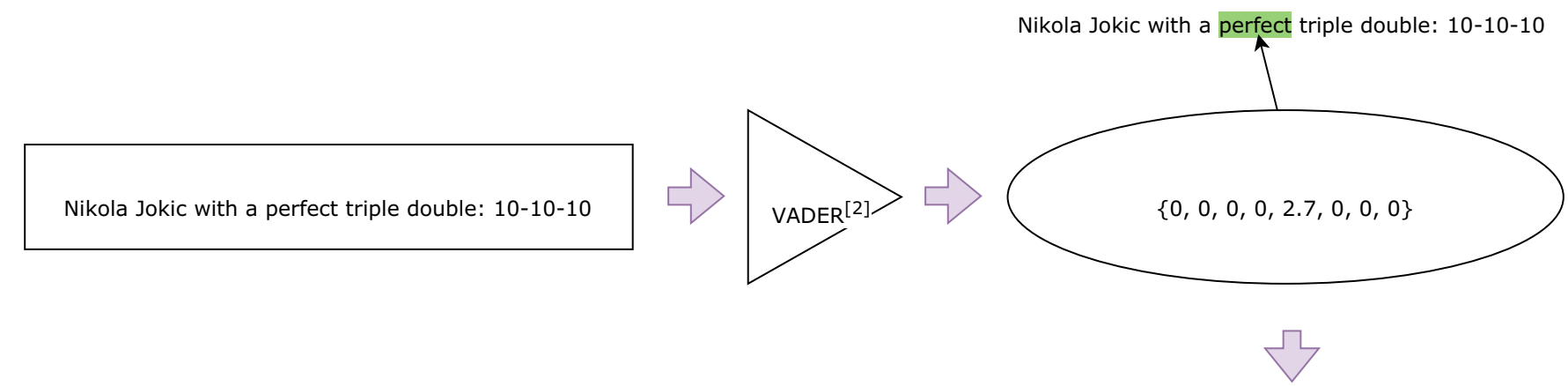
Mission

- 172 MVP candidates from 2010-2022
- With 40 total variables

```
##  mvppts ptsmax share games  mpg  ppg rbpq astpg
## 102   9 1310 0.007  82 38.1 29.0 6.1  7.5
## 51   765 1210 0.632  81 38.5 28.1 7.9  4.6
## 163  595 1000 0.595  67 32.9 29.9 11.6  5.8
## 158   1 1010 0.001  44 36.6 24.6 7.9 10.8
## 103   6 1310 0.005  77 37.0 21.2 4.7  6.4
```

- And 34 more variables...

Sentiment Analysis



- Repeated for all ~3700 Reddit post titles
- vader^[3] package in R made it very straightforward

[2] <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109/8122>

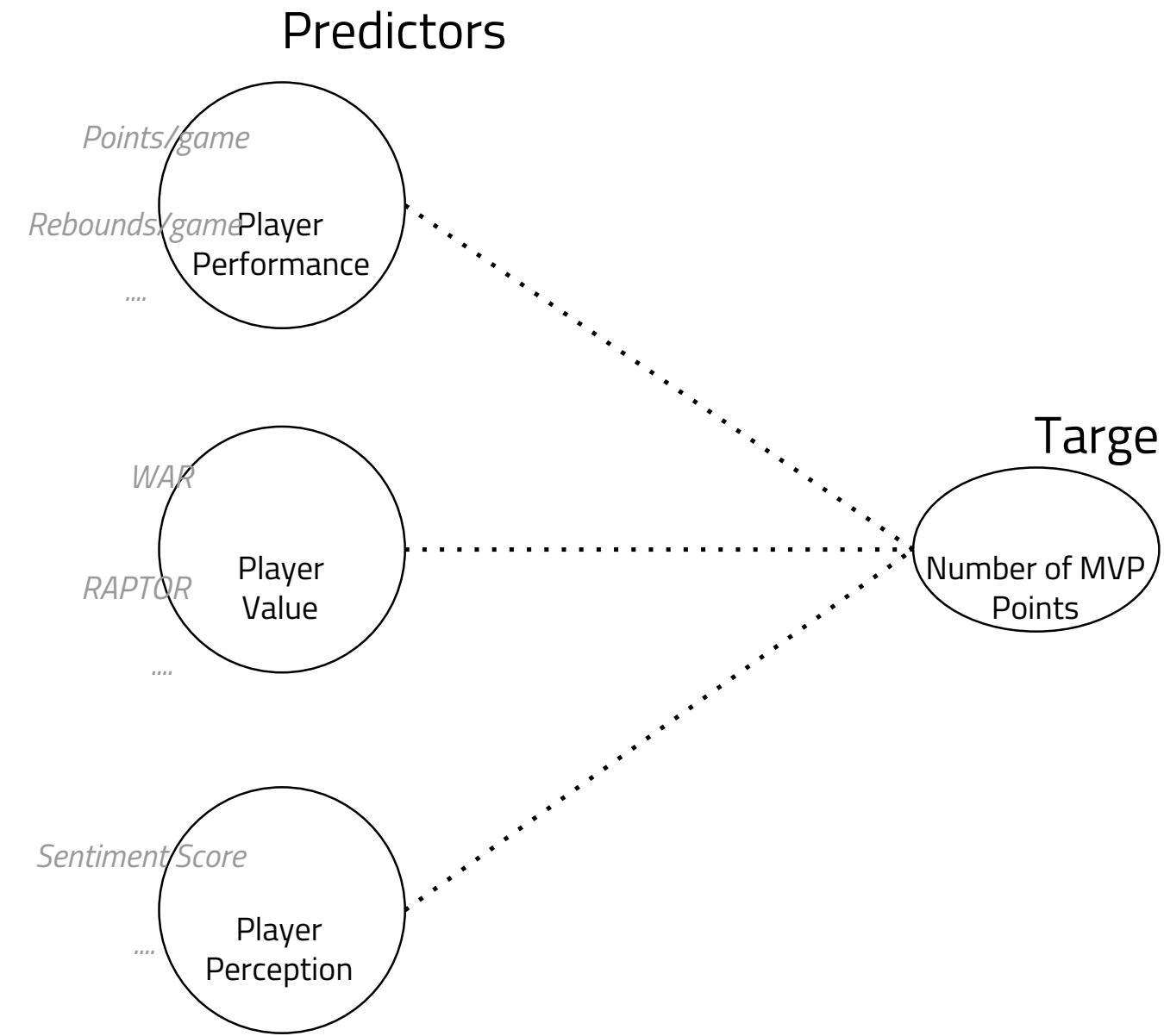
[3] <https://cran.r-project.org/web/packages/vader/index.html>

Sentiment Analysis

Example Text	Sentiment Score (VADER ^[2])
According to Dirk Nowitzki, reporters should pay no attention to Jason Terry after games	-0.382
[Draymond Green in 2010] “the world is flat fool”	-0.440
Stephen Curry is 5th All-Time in career True Shooting Percentage at 62.0%, and 1st All-Time among guards	0.421
[Murphy] Serge Ibaka interrupts DeRozan’s scrum to ask what he got the team for Christmas. Told nothing, Ibaka says he’s going to go play with John Wall.	0.026

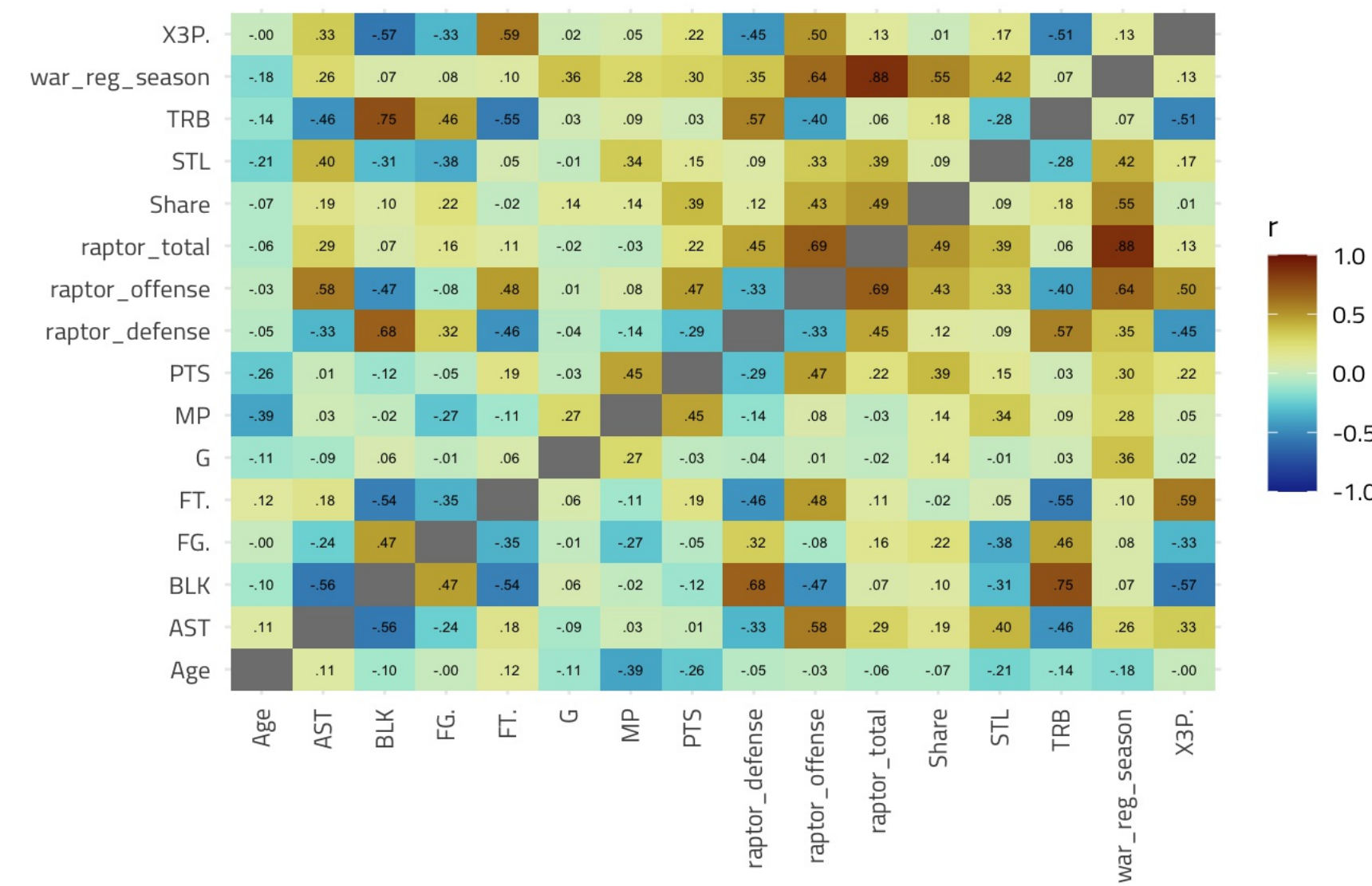
<!-- --> -->

Modeling

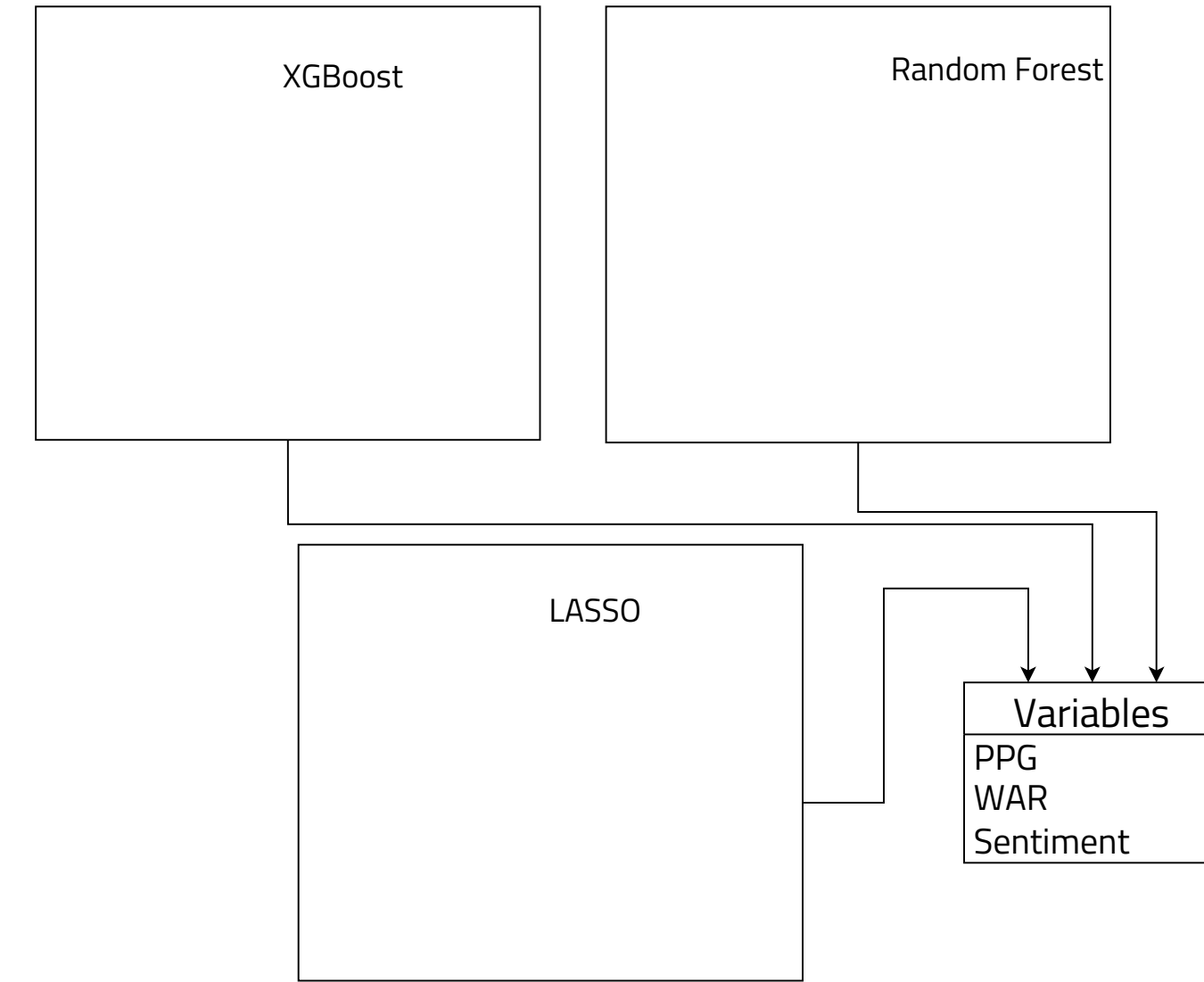


Modeling

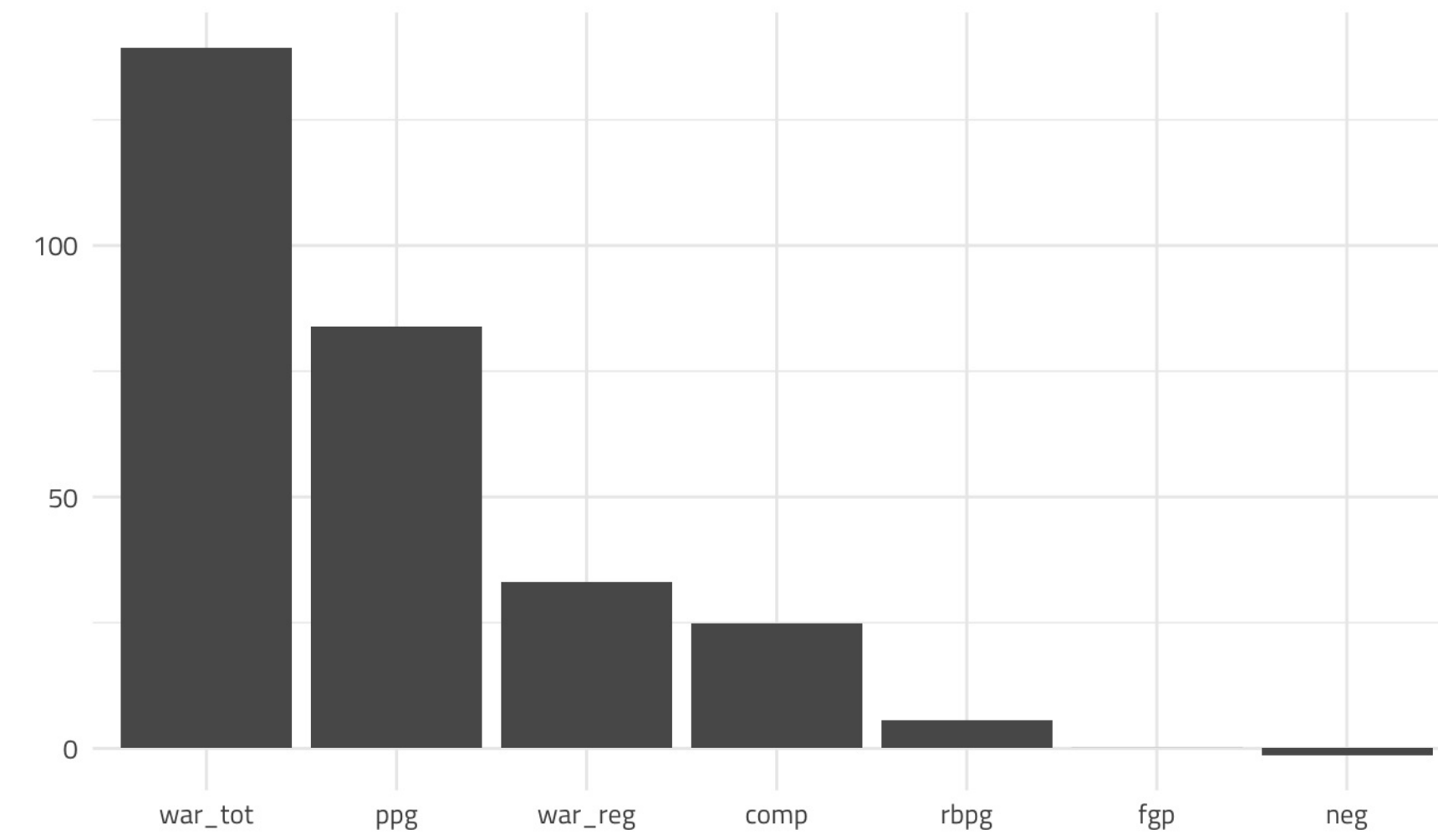
Correlation Matrix



Important Features



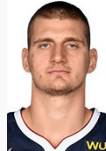


Important Features











Results

- Model was able to predict the MVP winner with *77% accuracy* (2010-2022)
- With average deviation from actual rank of *0.41*

Predictions for 2022 Season

Player	Rank	Predicted Rank	Sentiment Score
 Nikola Jokic	1	1	0.94
 Joel Embiid	2	2	-0.22
 Giannis Antetokounmpo	3	3	1.00

Full Predictions

SEASON	IMG	PLAYER	RANK	PREDICTED RANK	SENTIMENT
2010		LeBron James	1	1	-0.1
2010		Kobe Bryant	3	2	1.3
2010		Kevin Durant	2	3	0.14
2011		Derrick Rose	1	1	1.51
2011		Dwight Howard	2	2	-0.79
2011		LeBron James	3	3	-0.1
2012		LeBron James	1	1	-1.8
2012		Kevin Durant	2	2	-0.1

1-10 OF 39 ROWS

PREVIOUS 1 2 3 4 NEXT

Future

- Test out model midseason
- Add text sentiment analysis of famous critics like Stephen A Smith
- Use comments instead of headlines
- Comedy scores
- Account for jargon

Reflection

- Learned the arduous task of gathering data
- How we different models can be used

Resources Used For Presentaiton

- R package hexagons from <https://www.tidymodels.org/>
- Player headshots from <https://www.basketball-reference.com/>
- Diagrams created on <https://app.diagrams.net>