



# Predicting the NBA MVP

Nick Narey, Luke Duthie

Datasci 112 Winter 2024

## Abstract

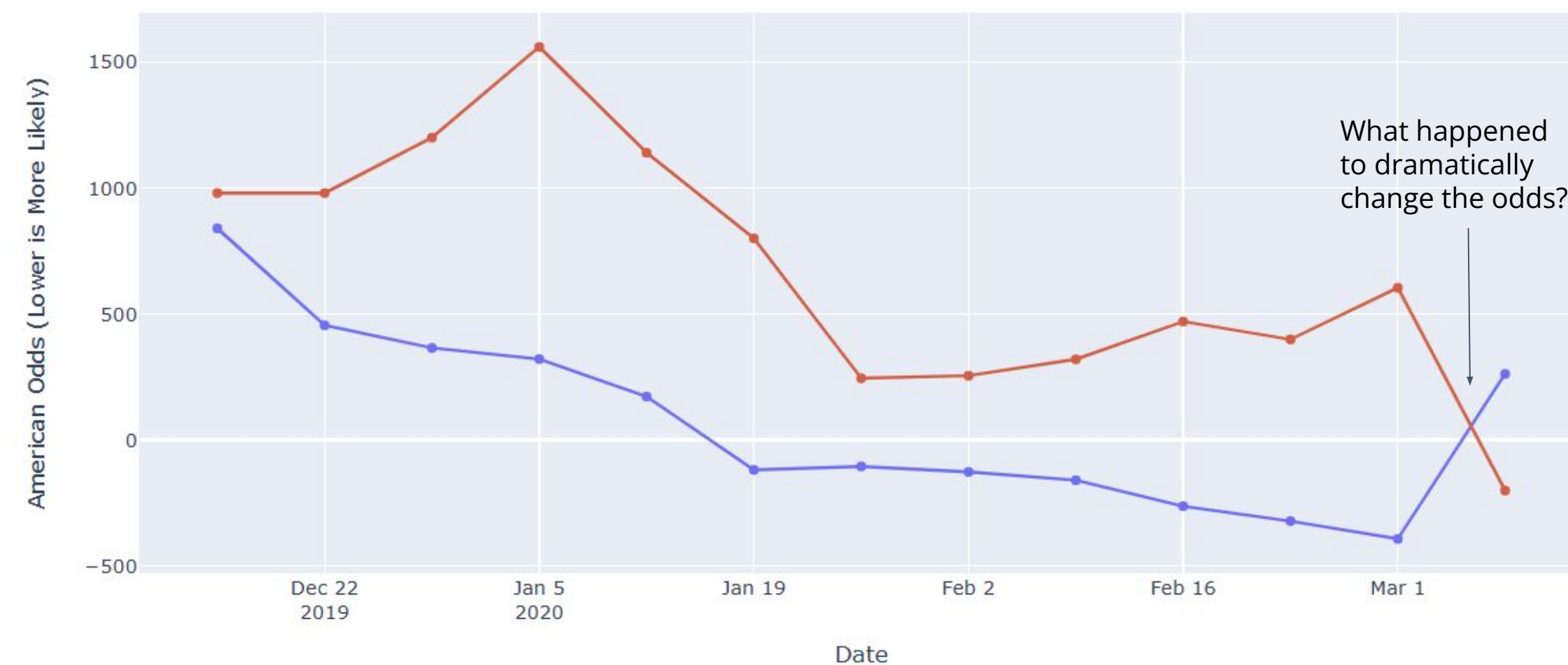
### The NBA's MVP discussion is out of control

The National Basketball Association (NBA) selects a Most Valuable Player (MVP) for every season based on individual regular season player statistics. Since the MVP is determined by media members voting based on personal opinion, this project explores ways that the NBA MVP can be determined without bias and with more consistency across each season. This project was inspired by the analytics and non-basketball debates of the 2023 NBA season, where Nikola Jokic and Joel Embiid got into a heated and controversial race for the award.



## Research Question: To what extent does the National Basketball Association have an unbiased and consistent MVP selection system?

### Inspiration: NBA MVP Odds 2022 - 2023



### Data Collection/Cleaning

- Scraped different stats tables for each year (2007-2023) from different websites
- Determined different minimum stats of former MVPs
- Applied minimums to training and test data for full roster
- Scraped Voting Shares proportion from another site and merged

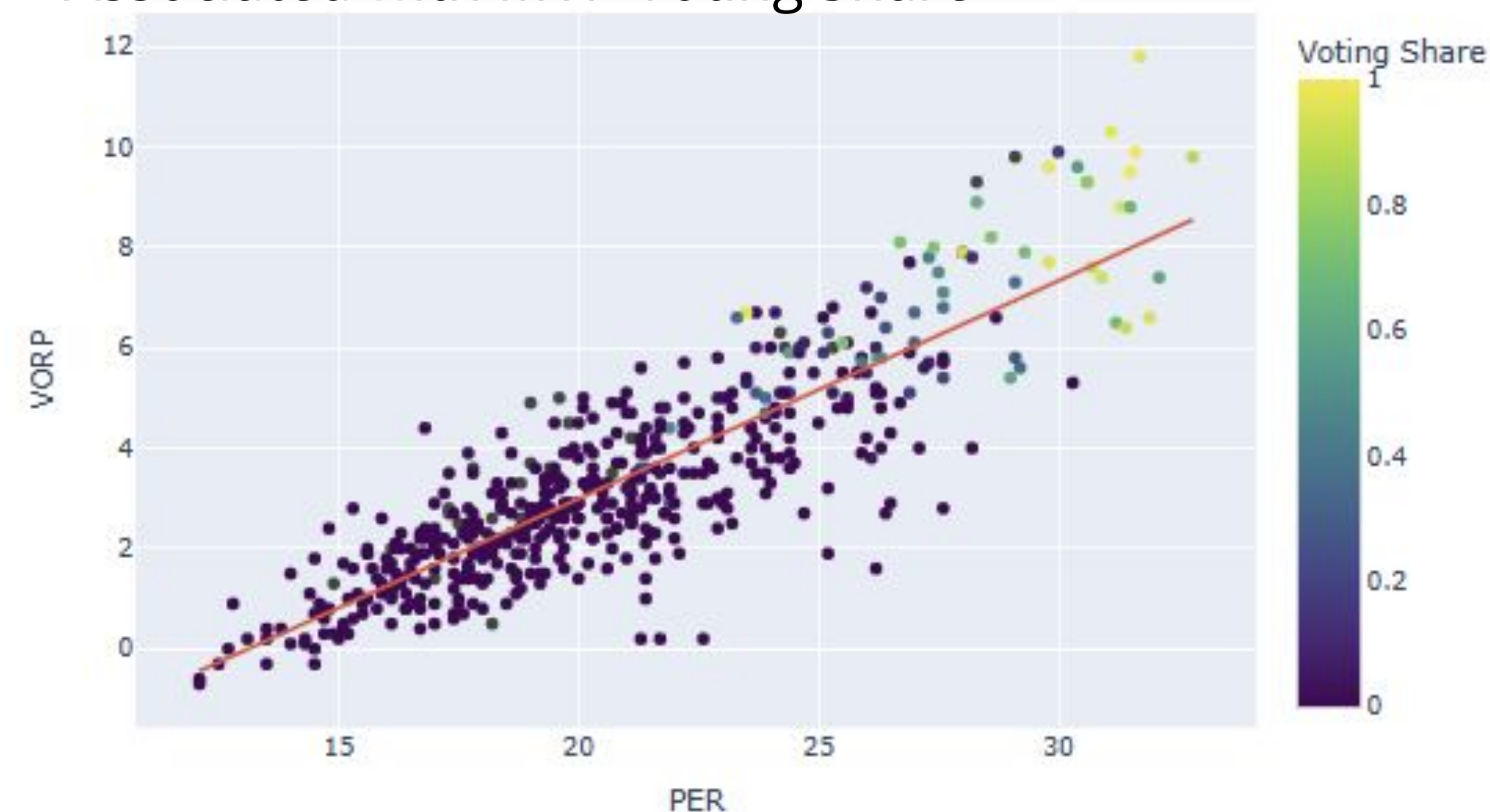
```
import requests
from bs4 import BeautifulSoup
import pandas as pd

url = "https://www.basketball-reference.com/leagues/NBA_2008_advanced."

def scrape_basketball_reference(url):
    response = requests.get(url)
    if response.status_code == 200:
        soup = BeautifulSoup(response.text, 'html.parser')
        table = soup.find('table')
        df = pd.read_html(str(table))[0]
        return df
    else:
        print('Error: Unable to retrieve data from the provided URL')
        return None

df_adv = scrape_basketball_reference(url)
df_adv = df_adv.drop_duplicates(subset=['Player'], keep='first')
df_adv
```

### Feature Selection: NBA PER and VORP Associated with MVP Voting Share



- Used Regression instead of Classification, predicting voting points proportion
  - more points with y-values to train

### 2024 Prediction: Random Forest Regressor

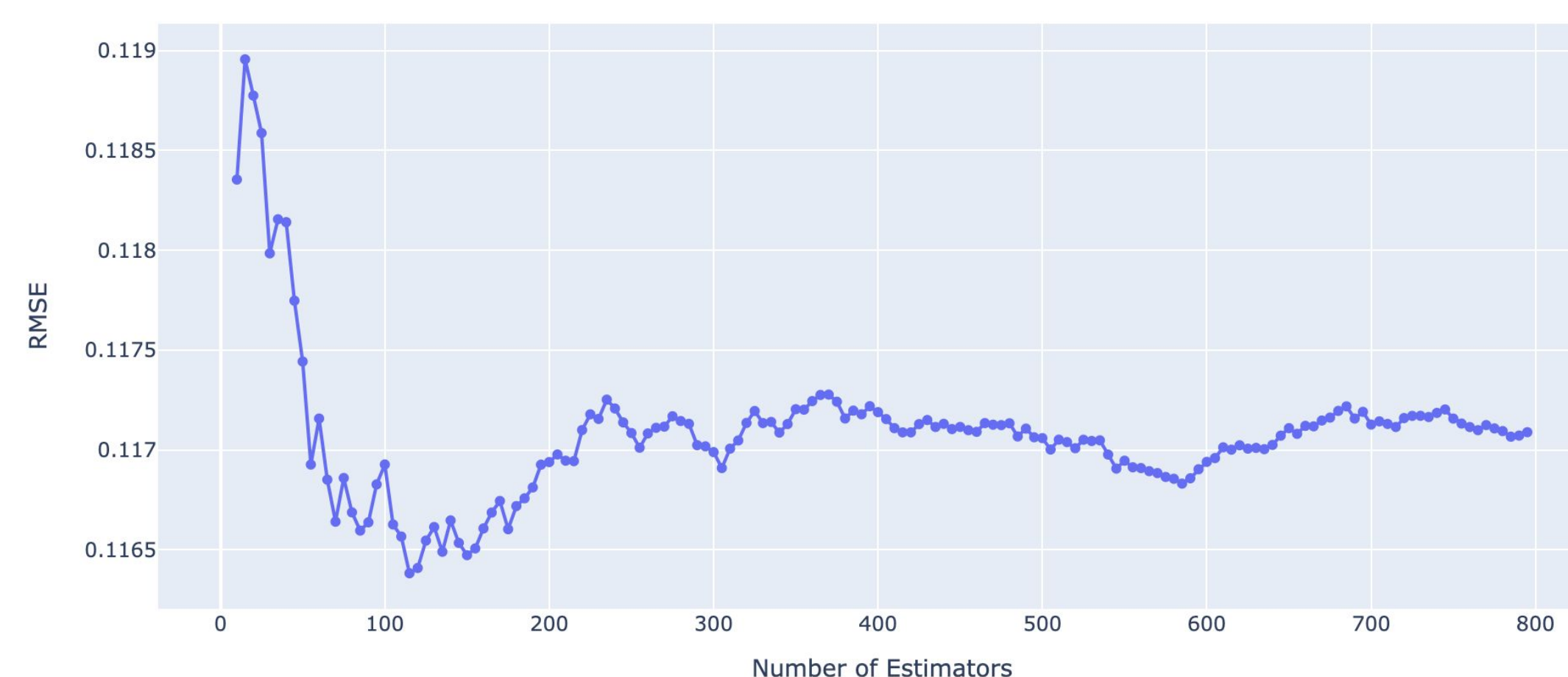
| Player        | Voting Share |
|---------------|--------------|
| Nikola Jokic  | .715         |
| Joel Embiid   | .664         |
| LeBron James  | .506         |
| Kevin Durant  | .445         |
| Stephen Curry | .053         |

### Testing on Different Years

| MVP                        | Random Forest         | K-Nearest Neighbors   | Gradient Boosting     |
|----------------------------|-----------------------|-----------------------|-----------------------|
| 2008 Kobe Bryant           | LeBron James          | LeBron James          | LeBron James          |
| 2009 LeBron James          | LeBron James          | LeBron James          | LeBron James          |
| 2010 LeBron James          | LeBron James          | LeBron James          | LeBron James          |
| 2011 Derrick Rose          | LeBron James          | LeBron James          | LeBron James          |
| 2012 LeBron James          | LeBron James          | LeBron James          | LeBron James          |
| 2013 LeBron James          | LeBron James          | LeBron James          | LeBron James          |
| 2014 Kevin Durant          | Kevin Durant          | Kevin Durant          | Kevin Durant          |
| 2015 Steph Curry           | James Harden          | James Harden          | James Harden          |
| 2016 Steph Curry           | Steph Curry           | Steph Curry           | Steph Curry           |
| 2017 Russell Westbrook     | Russell Westbrook     | James Harden          | Russell Westbrook     |
| 2018 James Harden          | James Harden          | James Harden          | James Harden          |
| 2019 Giannis Antetokounmpo | Giannis Antetokounmpo | Giannis Antetokounmpo | Giannis Antetokounmpo |
| 2020 Giannis Antetokounmpo | Giannis Antetokounmpo | Giannis Antetokounmpo | Nikola Jokic          |
| 2021 Nikola Jokic          | Nikola Jokic          | Nikola Jokic          | Nikola Jokic          |
| 2022 Nikola Jokic          | Nikola Jokic          | Nikola Jokic          | Nikola Jokic          |
| 2023 Joel Embiid           | Nikola Jokic          | Nikola Jokic          | Joel Embiid           |

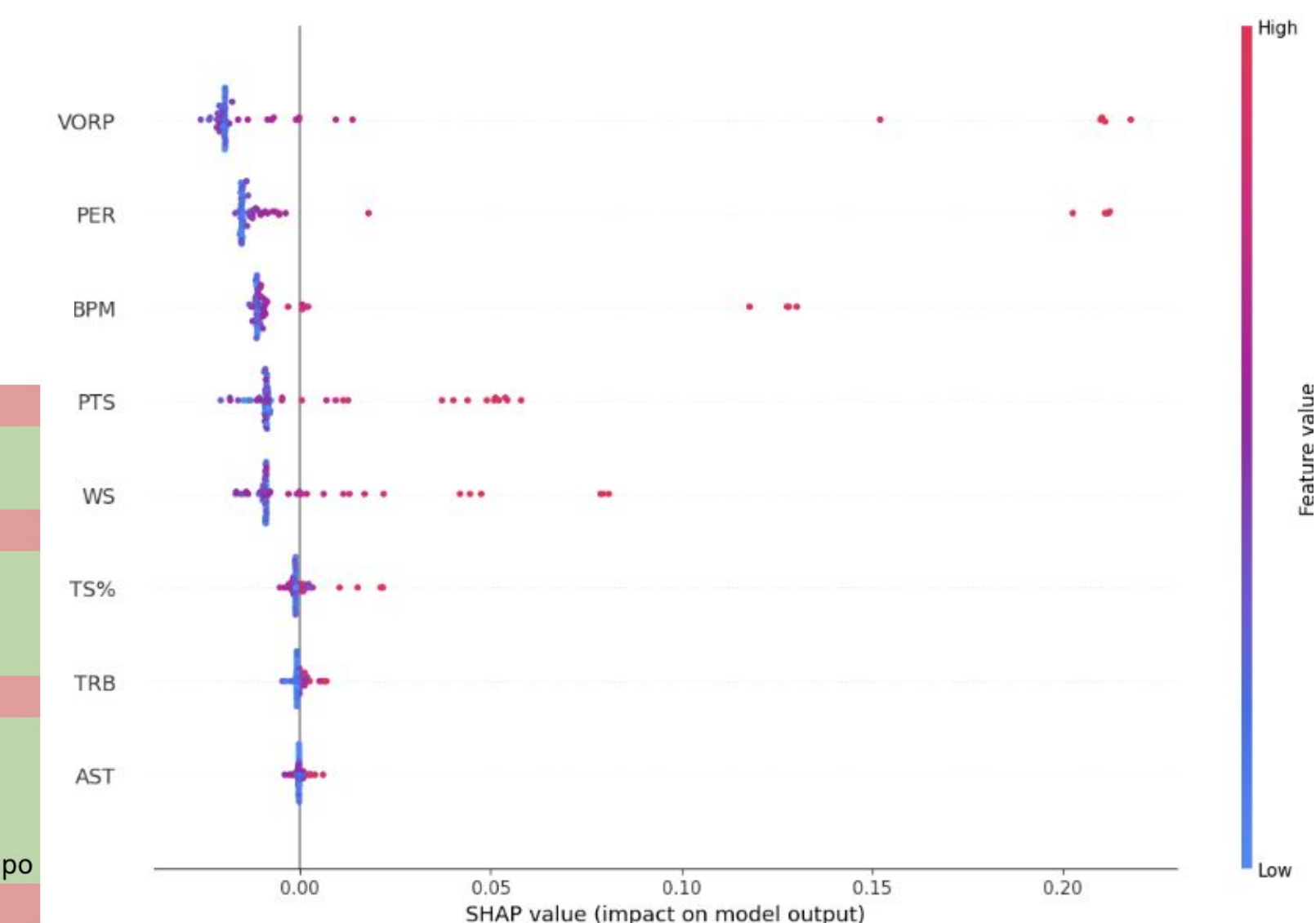
- Used Same CV method to predict Voting Share on given years
- Assigned MVP to player with highest voting share predicted

### Model Selection: Random Forest n\_estimators Tuning



- Implemented our own Cross validation grid search method
  - Entire season set aside to predict
  - Trained on remaining data
  - Average RMSE across all seasons
- Used to find best params for 3 models (one example above)

### Most Impactful Features



### Analysis

- 2008 - Kobe Bryant
  - Bryant gets "long overdue" first MVP
  - Lakers do well without Shaquille O'Neal
- 2011 - Derrick Rose
  - LeBron already had two MVPs
  - Miami Heat were expected to do better
  - Rose would be the youngest MVP ever
- 2023 - Joel Embiid
  - Stat padding
  - Racist allegations

### Conclusions

- Our results suggest that storylines often impact the way that voters decide on the MVP award.
- To fully be certain that media research has a massive influence on voting, we would like to see future research analyzing the relationship between the positivity of textual media articles and the changing of betting odds.
- We suggest that the NBA moves away from systems vulnerable to inconsistencies and toward a quantitative system to choose MVPs.