PREDICTING NBA TEAM PERFORMANCE USING MACHINE LEARNING

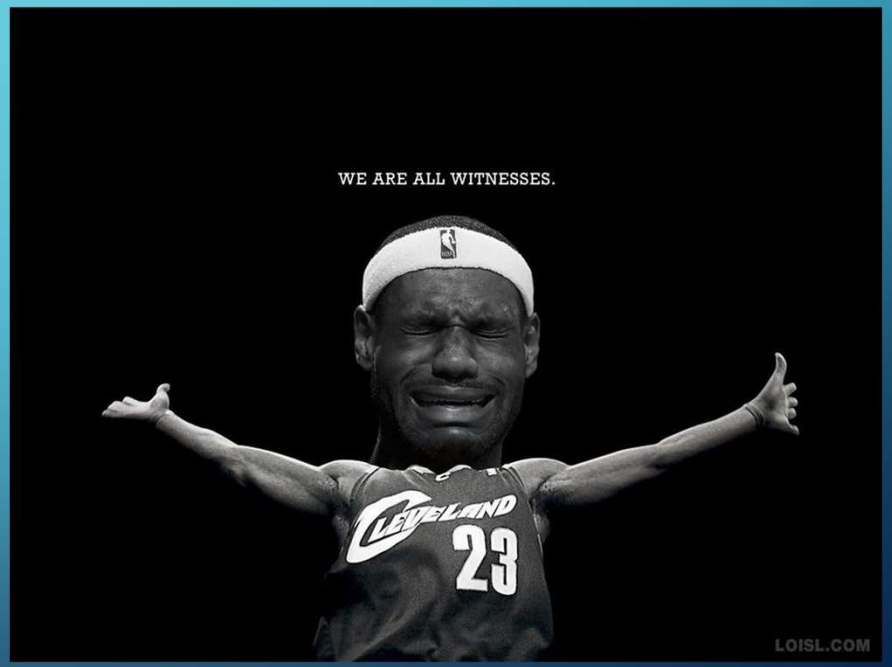
ANDREW ELKIN

GOAL

- Take a small training dataset (n=20 games) of limited team performance metric features
- Use several machine learning models to predict the final record of each team (Wins/Losses)
- Discover which feature and model combinations work best

WHY?

- Tons of applications:
  - NBA front office data analysis
  - Growing sports gambling market
  - It's fun ☺



DATA GATHERING AND FEATURE SELECTION

- Pulled from the open source NBA API
- Used data from 2018-2024
- Features Used:
  - Effective Field Goal Percentage, Opponent Effective Field Goal Percentage
  - Offensive Rebounding Rate, Opponent Offensive Rebounding Rate
  - Turnover Rate, Opponent Turnover Rate
  - Free Throw Attempt Rate, Opponent Free Throw Attempt Rate
- Target: Total Regular Season Wins

MODELS USED

- Ordinary Least Squares Linear Regression
- Ridge Regression
- Random Forest
- Boosted Trees

THE WINNER????

Ridge Regression!!!! R^2 of 0.77

TREE MODELS DID NOT FAIR WELL

Random Forest: R^2 of 0.55

Boosted Tree: R^2 of 0.47

TAKEAWAYS

- Chosen features are very strong predictors of overall team success
- Relationship is close to linear
- Dark Forest and Boosted Trees need a lot more data

WHAT'S NEXT?

- Adding custom weights to each feature to optimize predictions
- Adding more seasons of data

REFERENCES

- Images:
  - https://www.espn.com/video/clip/_/id/28529738
  - https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSg69swJU6Lvcy1YUmg_Y2JICTXWM1ulT883g&s
  - https://live.staticflickr.com/2620/5848748388_28e5e43da9_b.jpg