

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

College Basketball Win Prediction

Joshua Majano



What and How?

Question

- Can we predict the number of wins using only the team stats?

Which features will we use?

1. “Four Factors of Basketball Success” from Dean Oliver
 - a. There are four most important strategies to win a basketball game
 - i. Score on every possession (effective fields goals)
 - ii. Pick up all rebounds (rebounding percentage)
 - iii. Get to the foul line (free throw rate)
 - iv. Protect the ball (turnover percentage)
2. Adjusted Efficiency Stats
3. All of them



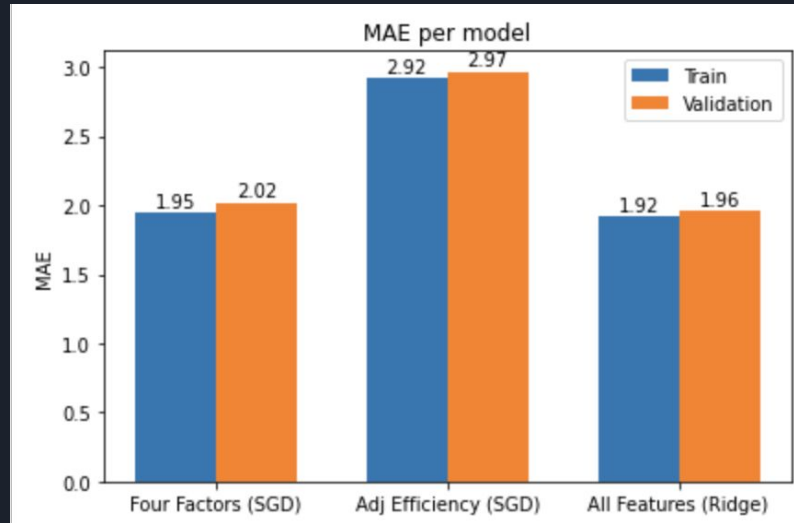
The Data

- Contains data from the 2015-2019 Division I college basketball seasons
- The only missing values come from the “POSTSEASON” and “SEED” columns
- 1757 rows with ~353 teams and 24 columns of team stats and features
 - 351 teams from 2015-2018 seasons
 - 2 new teams were added in 2019 season
- Only used data from 2015-2018 seasons
 - Trained/Validated on 2015-2017 seasons
 - Tested on 2018 season



Model Picks and Metrics

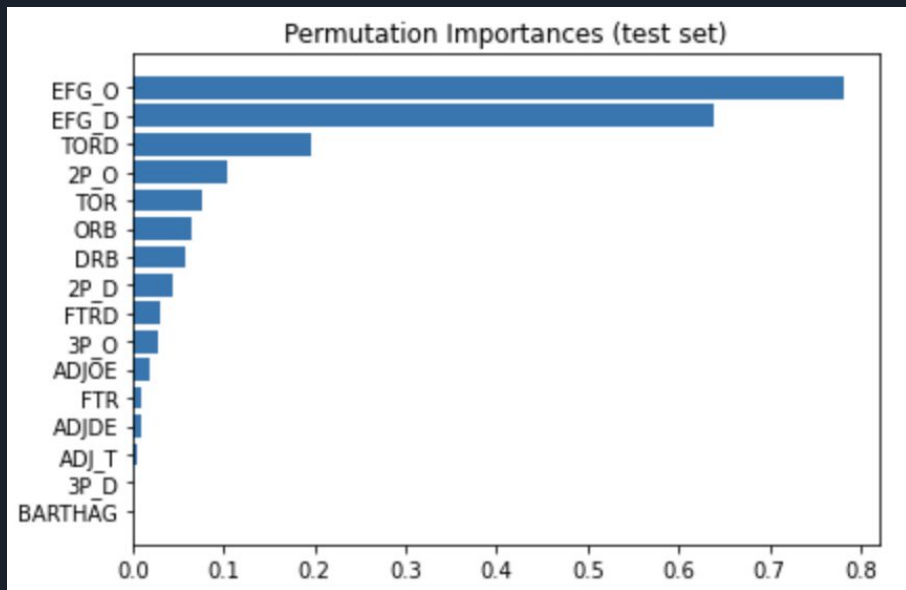
- Four Factors Model (8 features)
 - EFG_O, EFG_D, TOR, TORD, ORB, DRB, FTR, FTRD
- Adjusted Efficiency Stats Model (3 features)
 - ADJOE, ADJDE, ADJ_T
- “All” Features Model (16 features)



Winner: Ridge model using all of the team stats features

Notable Results

- This final model had an MAE = 1.93 on the test data
 - On average predicted within “1.93” wins of the actual team wins in 2018 season
- What did this model consider important?





Conclusion

- “That’s why we play the game...”
 - Our best model utilized as many features possible about each team to make most accurate predictions
- Yet, our prediction accuracy was pretty respectable
- There is a lot to uncover from the numbers
- Future models:
 - Consider including school name, conference played in, etc...
 - Can we expand predictions to success in the postseason?
- Beyond this dataset:
 - Are there other stats not in this dataset that can improve our predictions?
Could there be a different or new stat created that can be just as predictive as all of these stats together?