

NCAA Basketball Analysis

Maclean Sherren & Kayla Tansiongco

Brigham Young University

December 13, 2023

Outline

- 1 Introduction
- 2 Data Collection
- 3 Exploratory Data Analysis
- 4 Methodology
- 5 Results
- 6 Conclusion
- 7 Q & A
- 8 References

Introduction

- Due to Covid, 2020 NCAA tournament did not happen.
- Question: Are there certain regular season variables that are better predictors for postseason outcomes?
- Goal: Use existing data/trends from other years and regular season data from 2020 to come up with potential methods of prediction and postseason outcomes.

- College Basketball Dataset folder found on Kaggle. Dataset "cbb.csv" has regular and postseason statistics from years 2013-2021. Excludes 2020 season due to lack of postseason statistics.
- Folder included datasets for each individual year, with 2020 ("cbb20.csv") having regular season stats only.
- Variables of final dataset: YEAR, TEAM, CONF, G, W, BARTHAG, POSTSEASON, SEASON_FINAL.

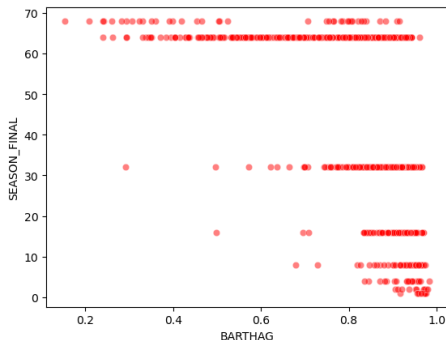
- Created numerically coded column, SEASON_FINAL, derived from POSTSEASON (string data type) for cbb.csv dataframe. (Ex: "Champions" = 1, "2ND" = 2, "F4" = 4)
- Initialized variables POSTSEASON and SEASON_FINAL as NaN for cbb20.csv dataframe. This is what we will fill in through our prediction methods.

Exploratory Data Analysis

- Determine what variables are most correlated with POSTSEASON outcomes.
- Used Spearman correlation since it is more appropriate for measurements taken from ordinal scales.

Exploratory Data Analysis

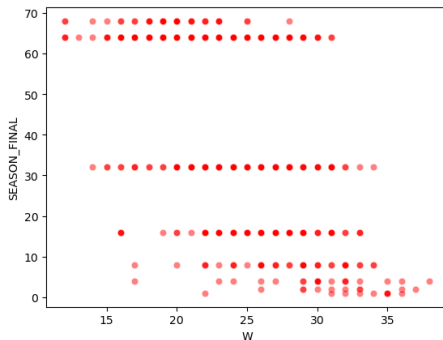
- BARTHAG vs SEASON_FINAL
- BARTHAG = Power Rating (Chance of beating an average Division 1 team)



Corr: -0.6461273899341722

Exploratory Data Analysis

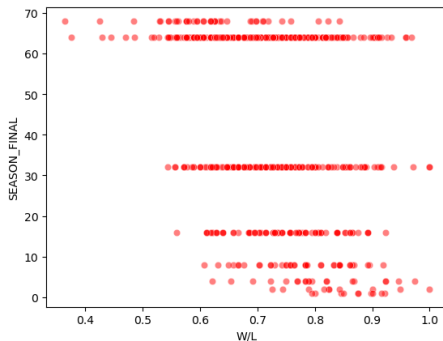
- W vs SEASON_FINAL
- W = Number of games won



Corr: -0.5125436890870254

Exploratory Data Analysis

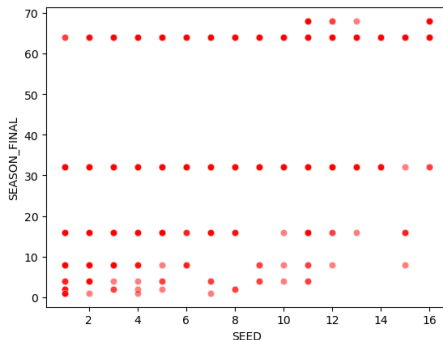
- W/L vs SEASON_FINAL
- W/L = Win-Loss Ratio



Corr: -0.3532346489556209

Exploratory Data Analysis

- SEED vs SEASON_FINAL
- SEED = Seed in the NCAA March Madness Tournament



Corr: 0.6065146756151623

- Package contains a function that takes a season and method of prediction, then produces simulated results of an NCAA tournament.
- Compare predicted results to actual results.
- Create a prediction for 2020 NCAA postseason tournament.
- Since seeding was not determined for 2020 postseason, variable RK will be used as a predictor.

- Season: 2021
- Method: BARTHAG (Power Rating)

Predicted Final Four

- Baylor
- Michigan
- Gonzaga
- Houston

Predicted Championship

- Gonzaga vs Baylor
- Winner: Baylor

Actual Final Four

- Baylor
- Gonzaga
- Houston
- UCLA

Actual Championship

- Gonzaga vs Baylor
- Winner: Baylor

- Season: 2021
- Method: W (Wins)

Predicted Final Four

- Winthrop
- Alabama
- Gonzaga
- Loyola Chicago

Predicted Championship

- Alabama vs Gonzaga
- Winner: Alabama

Actual Final Four

- Baylor
- Gonzaga
- Houston
- UCLA

Actual Championship

- Gonzaga vs Baylor
- Winner: Baylor

- Season: 2021
- Method: SEED

Predicted Final Four

- Baylor
- Gonzaga
- Michigan
- Illinois

Predicted Championship

- Michigan vs Gonzaga
- Winner: Gonzaga

Actual Final Four

- Baylor
- Gonzaga
- Houston
- UCLA

Actual Championship

- Gonzaga vs Baylor
- Winner: Baylor

Results

- Season: 2020
- Method: BARTHAG (Power Rating)

```
Round 1:
  Baylor vs Texas Southern -> Winner: Baylor
  Cincinnati vs VCU -> Winner: Cincinnati
  Illinois vs East Tennessee St. -> Winner: Illinois
  Auburn vs College of Charleston -> Winner: Auburn
  Iowa vs Mississippi St. -> Winner: Iowa
  Seton Hall vs Bowling Green -> Winner: Seton Hall
  Ohio St. vs Northern Iowa -> Winner: Ohio St.
  Louisville vs Murray St. -> Winner: Louisville

Round 2:
  Baylor vs Cincinnati -> Winner: Baylor
  Illinois vs Auburn -> Winner: Illinois
  Iowa vs Seton Hall -> Winner: Seton Hall
  Ohio St. vs Louisville -> Winner: Ohio St.

Round 3:
  Baylor vs Illinois -> Winner: Baylor
  Seton Hall vs Ohio St. -> Winner: Ohio St.

Round 4:
  Baylor vs Ohio St. -> Winner: Baylor

Baylor advances to the Final Four!
```

Final Four: Baylor, Duke, Gonzaga, Kansas

Playoff Finals: Baylor vs Kansas

NCAA D1 Division Champs: Baylor

Results

- Season: 2020
- Method: W (Wins)

<p>Round 1:</p> <p>Gonzaga vs Stanford → Winner: Gonzaga Florida vs Stanford → Winner: Stanford Penn St. vs Stanford → Winner: Stanford Oregon vs New Mexico State → Winner: New Mexico State Gonzaga vs DU → Winner: DU Oklahoma vs Colorado → Winner: Colorado Michigan at Arkansas → Winner: Arkansas West Virginia at North Dakota St. → Winner: North Dakota St.</p> <p>Round 2:</p> <p>Gonzaga vs Stanford → Winner: Gonzaga Stephen F. Austin vs New Mexico State → Winner: New Mexico State Ohio vs Colorado → Winner: Colorado Arkansas vs North Dakota St. → Winner: North Dakota St.</p> <p>Round 3:</p> <p>Gonzaga vs New Mexico State → Winner: New Mexico State Colorado vs North Dakota St. → Winner: North Dakota St.</p> <p>Round 4:</p> <p>Gonzaga vs New Mexico State → Winner: New Mexico State Colorado vs North Dakota St. → Winner: North Dakota St.</p> <p>North Dakota St. advances to the Final Four</p>	<p>Round 1:</p> <p>Bayler vs Texas Southern → Winner: Bayler Cincinnati vs WCU → Winner: Cincinnati Illinois vs East Tennessee St. → Winner: East Tennessee St. Auburn vs College of Charleston → Winner: Auburn Iowa vs Mississippi St. → Winner: Mississippi St. Seton Hall vs Bowling Green → Winner: Bowling Green Ohio St. vs Northern Iowa → Winner: Northern Iowa Louisville vs Murray St. → Winner: Louisville</p> <p>Round 2:</p> <p>Bayler vs Cincinnati → Winner: Bayler East Tennessee St. vs Auburn → Winner: East Tennessee St. Mississippi St. vs Bowling Green → Winner: Bowling Green Northern Iowa vs Louisville → Winner: Northern Iowa</p> <p>Round 3:</p> <p>Bayler vs East Tennessee St. → Winner: East Tennessee St. Mississippi St. vs Bowling Green → Winner: Bowling Green Northern Iowa vs Louisville → Winner: East Tennessee St.</p> <p>Round 4:</p> <p>East Tennessee St. vs Northern Iowa → Winner: East Tennessee St.</p> <p>East Tennessee St. advances to the Final Four</p>
<p>Round 1:</p> <p>San Diego St. vs Memphis → Winner: Memphis Saint Mary's vs Texas Tech → Winner: Saint Mary's Kentucky vs Arizona St. → Winner: Kentucky Butler vs Vermont → Winner: Vermont Creighton vs Virginia → Winner: Creighton Maryland vs UC Irvine → Winner: Maryland Rutgers vs Houston → Winner: Houston Duke vs Wright St. → Winner: Wright St.</p> <p>Round 2:</p> <p>Memphis vs Saint Mary's → Winner: Saint Mary's Kentucky vs Vermont → Winner: Vermont Creighton vs Maryland → Winner: Maryland Houston vs Wright St. → Winner: Wright St.</p> <p>Round 3:</p> <p>Saint Mary's vs Vermont → Winner: Vermont Maryland vs Wright St. → Winner: Wright St.</p> <p>Round 4:</p> <p>Saint Mary's vs Vermont → Winner: Vermont Maryland vs Wright St. → Winner: Wright St.</p> <p>Vermont advances to the Final Four</p>	<p>Round 1:</p> <p>Kansas vs Princeton → Winner: Kansas Wisconsin vs Rhode Island → Winner: Rhode Island Arizona vs Liberty → Winner: Liberty Michigan St. vs North Texas → Winner: Michigan St. LSU vs Purdue → Winner: LSU Florida St. vs Little Rock → Winner: Florida St. Marquette vs Indiana → Winner: Indiana Dayton vs Winthrop → Winner: Dayton</p> <p>Round 2:</p> <p>Kansas vs Rhode Island → Winner: Kansas Liberty vs Michigan St. → Winner: Liberty LSU vs Florida St. → Winner: Florida St. Indiana vs Dayton → Winner: Dayton</p> <p>Round 3:</p> <p>Kansas vs Liberty → Winner: Liberty Florida St. vs Dayton → Winner: Dayton</p> <p>Round 4:</p> <p>Kansas vs Liberty → Winner: Liberty Florida St. vs Dayton → Winner: Dayton</p> <p>Liberty advances to the Final Four</p>

Final Four: East Tennessee St., Vermont, Gonzaga, Liberty

Playoff Finals: East Tennessee St. vs Gonzaga

NCAA D1 Division Champs: East Tennessee St.

Results

- Season: 2020
- Method: RK (Rank/Equivalent to SEED)

<p>Round 1:</p> <p>Gonzaga vs Montana → Winner: Gonzaga Florida vs Stanford → Winner: Florida Purdue vs Saint Mary's → Winner: Saint Mary's Oregon vs New Mexico State → Winner: New Mexico State Colorado vs BYU → Winner: BYU Villanova vs Colgate → Winner: Villanova Michigan vs Arkansas → Winner: Michigan West Virginia vs North Dakota St. → Winner: West Virginia</p> <p>Round 2:</p> <p>Gonzaga vs Florida → Winner: Gonzaga Saint Mary's vs New Mexico State → Winner: New Mexico State BYU vs Villanova → Winner: Villanova Michigan vs West Virginia → Winner: Michigan</p> <p>Round 3:</p> <p>Gonzaga vs New Mexico State → Winner: New Mexico State Villanova vs Michigan → Winner: Villanova</p> <p>Round 4:</p> <p>Villanova advances to the Final Four! → Winner: Villanova</p>	<p>Round 1:</p> <p>Baylor vs Texas Southern → Winner: Baylor Cincinnati vs VCU → Winner: Cincinnati Illinois vs East Tennessee St. → Winner: Illinois Auburn vs College of Charleston → Winner: Auburn Iowa vs Mississippi St. → Winner: Iowa Seton Hall vs Bowling Green → Winner: Seton Hall Ohio St. vs Northern Iowa → Winner: Ohio St. Louisville vs Murray St. → Winner: Louisville</p> <p>Round 2:</p> <p>Baylor vs Cincinnati → Winner: Baylor Illinois vs Auburn → Winner: Illinois Iowa vs Seton Hall → Winner: Seton Hall Ohio St. vs Louisville → Winner: Ohio St.</p> <p>Round 3:</p> <p>Baylor vs Illinois → Winner: Baylor Seton Hall vs Ohio St. → Winner: Ohio St.</p> <p>Round 4:</p> <p>Baylor advances to the Final Four! → Winner: Baylor</p>
<p>Round 1:</p> <p>San Diego St. vs Memphis → Winner: Memphis Saint Mary's vs Texas Tech → Winner: Texas Tech Kentucky vs Arizona St. → Winner: Kentucky Butler vs Vermont → Winner: Butler Creighton vs Virginia → Winner: Creighton Maryland vs UC Irvine → Winner: Maryland Rutgers vs Houston → Winner: Houston Duke vs Wright St. → Winner: Duke</p> <p>Round 2:</p> <p>Memphis vs Texas Tech → Winner: Texas Tech Kentucky vs Butler → Winner: Butler Creighton vs Maryland → Winner: Creighton Houston vs Duke → Winner: Duke</p> <p>Round 3:</p> <p>Texas Tech vs Butler → Winner: Texas Tech Creighton vs Duke → Winner: Duke</p> <p>Round 4:</p> <p>Duke advances to the Final Four! → Winner: Duke</p>	<p>Round 1:</p> <p>Kansas vs Princeton → Winner: Kansas Wisconsin vs Rhode Island → Winner: Wisconsin Arizona vs Liberty → Winner: Arizona Michigan St. vs North Texas → Winner: Michigan St. LSU vs Purdue → Winner: Purdue Florida St. vs Little Rock → Winner: Florida St. Marquette vs Indiana → Winner: Marquette Dayton vs Winthrop → Winner: Dayton</p> <p>Round 2:</p> <p>Kansas vs Wisconsin → Winner: Kansas Arizona vs Michigan St. → Winner: Michigan St. Purdue vs Florida St. → Winner: Florida St. Marquette vs Dayton → Winner: Dayton</p> <p>Round 3:</p> <p>Kansas vs Michigan St. → Winner: Kansas Florida St. vs Dayton → Winner: Dayton</p> <p>Round 4:</p> <p>Kansas advances to the Final Four! → Winner: Kansas</p>

Final Four: Baylor, Duke, Gonzaga, Kansas

Playoff Finals: Baylor vs Kansas

NCAA D1 Division Champs: Kansas

Conclusion

- While there is a lot of randomness and uncertainty to account for in the NCAA tournament, there are variables that can accurately predict the outcomes.
- Among the variables observed, power rating was the strongest method of prediction.
- Potential to apply package to other variables as well.
- Aside from our own observations, we could apply further analyses to evaluate the accuracy of our predictors.
- Potential to use same methods to make predictions for future seasons.

Questions or comments?

Thank you! Happy Holidays!

References

- Kaggle Dataset
<https://www.kaggle.com/datasets/andrewsundberg/college-basketball-dataset?select=cbb.csv>
- 2021 NCAA Tournament Results
<https://www.ncaa.com/news/basketball-men/article/2022-07-20/2021-ncaa-bracket-scores-stats-records-march-madness-mens-tournament>
- 2020 Predicted Bracket
<https://www.ncaa.com/news/basketball-men/article/2020-02-28/ncaa-predictions-projections-2020-bracket-andy-katz>
- BracketMaker (preliminary guide for package structure)
<https://github.com/klauscurde/BracketMaker>