Introduction

Predicting NCAA “March Madness” outcomes has long attracted data scientists seeking to beat naive ranking-based strategies and “gut” picks. This paper describes my project, formulating a lightweight, interpretable scoring function based on normalized team metrics and optimizing its weights through simulation on historical games. I summarize related literature, pose my refined research questions, review ethical implications of sports data use, detail my data sources and exploration, outline the analysis plan, report on final results, discuss findings and limitations, and reflect on challenges faced.

Mini Literature Review

What convinced me to do this project firstly is my love of march madness, but also ideas that ways of picking the tournament could be improved. People with similar ideas have inspired me, including Kenpom. The idea that different metrics can change based on the round they are in comes from stats of when teams end up losing. Early work in tournament prediction highlighted the value of efficiency metrics: Jared Dean demonstrated that historical data patterns (e.g., 12th-seed upsets) improve predictions over intuition. Most comparable to my work is a March 2025 arXiv paper simplifying FiveThirtyEight’s framework to four key predictors with logistic regression.

Research Questions

Question 1: Can we predict the outcome of NCAA March Madness games using historical team metrics and optimized weighting parameters?

Question 2: Are offense or defensive team statistics most predictive of tournament success?

Question 3: Do the predictive power of certain factors change in later tournament rounds?

Ethics Review

Ethical considerations are very important when doing any project. My project follows the transparency principle by fully documenting all the data collected. All team metrics are from publicly available NCAA and ESPN websites.To uphold fairness, we avoid using demographic data or proprietary scouting reports that could introduce bias.

Privacy concerns are minimal, as all team-level metrics are publicly available, and no specific players are mentioned in the project. While specific data, such as player specific injury reports, could improve predictions, incorporating such sensitive details could raise concerns about privacy, and be difficult to implement the effect of differing injuries. This model also does not ensure any outcome to a specific game and is limited in its predicting power. It is not accountable for any decisions based on the results.

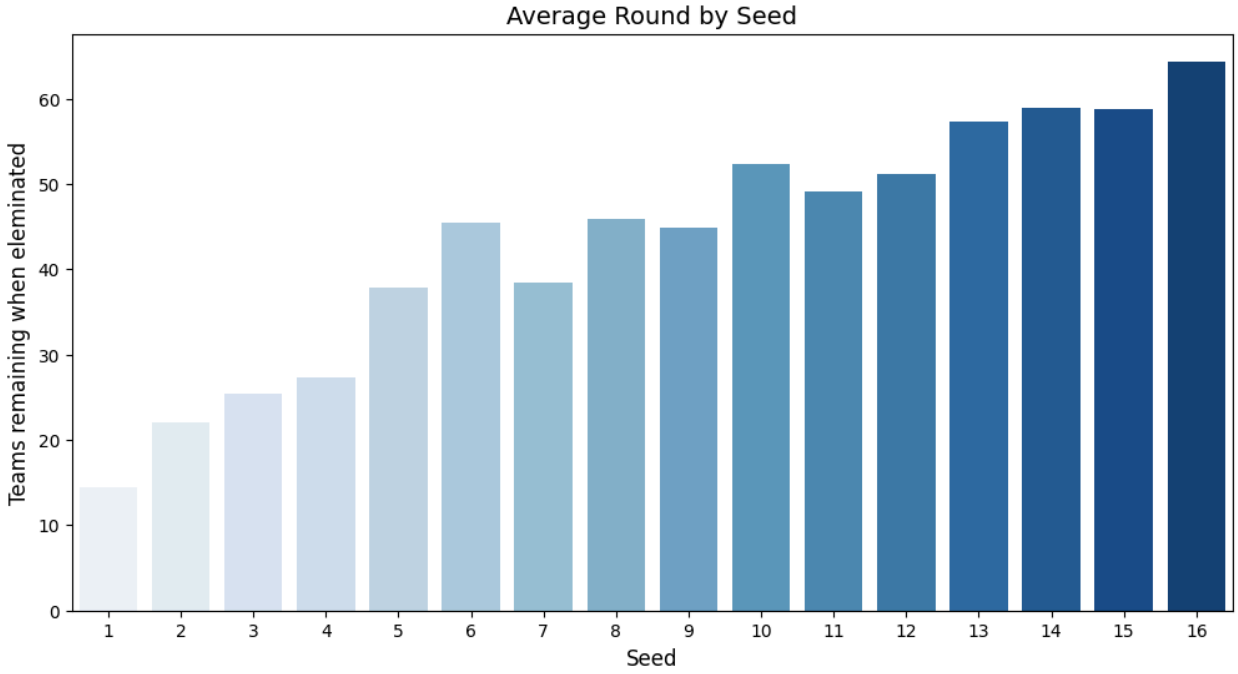
Data

I used two main CSVs: DEV _ March Madness.csv, containing historical team metrics (AdjOE, AdjDE, seed), and EvanMiya.csv, adding supplemental features (talent, height, three point percentage). Data was also used from NCAA and ESPN websites to get historical game results data. We chose these datasets for their completeness and their historical scope.

Data Exploration

	Adjusted Offensive Efficiency	Adjusted Defensive Efficiency
count	8314.000000	8314.000000
mean	103.693613	103.693649
std	7.432575	6.458849
min	71.500000	84.100000
25%	98.700000	99.300000
50%	103.600000	103.900000
75%	108.700000	108.400000
max	130.400000	125.000000

This output provides summary statistics such as mean, median, and standard deviation for AdjOE and AdjDE. It shows that both have a similar median and mean which makes sense because for every negative play on defense there is a positive play for the offense and vice versa. What is interesting is that the spread for offense seems to be much larger, which makes it seem like offense should be more impactful because the difference between a good offense and a bad offense is greater than that of defense.



This bar chart visualizes how far each seed got into the tournament. Higher seeds (lower numbers) tend to get further (lower teams left when eliminated), indicating stronger teams which makes sense. What is interesting is the bigger drop off once you get to the top 4 seeds showing that seeding may be more predictive for the higher seeds making it a better measure for picking champions than early round predictions.

Analysis Plan

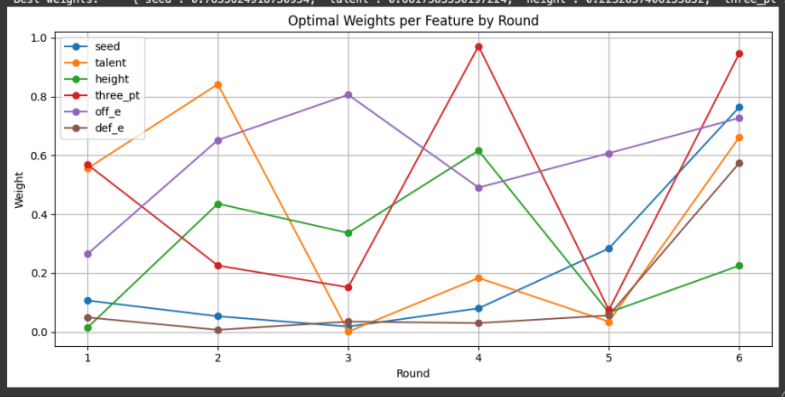
Q1 Can we predict the outcome of NCAA March Madness games using historical team metrics and optimized weighting parameters? The plan here is to train the model by taking the team statistics and weighing them by random amounts and plotting those weights by the amount of games it predicts correctly. Q2: Are offense or defensive team statistics most predictive of tournament success? Take the plots and look to see

at the highest accuracy if the weighted number is higher on offense or defense. Q3: Do the predictive power of certain factors change in later tournament rounds? The plan for this would be when we are putting in the equation for the model we can multiply or divide by the round number so that the effects either increase or decrease based on what round they are in. To find if each metric should increase or decrease I treat it like a regression problem. First run simulations round-by-round; for each round, optimize weights; track which features are most useful per round and use this to create a round-dependent weight schedule per feature.

Results and Findings & Limitations

So first we should answer the question about which stats are important for each round. This was done by randomly applying weights 10,000 times for each round and calculating the accuracy. The following graph shows what the weight was for each metric that created the highest accuracy simulated in each round.

```
--- Round 1 ---
Games considered: 160/160
Best accuracy:    0.787
Best weights:     {'seed': 0.10692272877297493, 'talent': 0.5553715391117405, 'height': 0.015005208958239535, 'three_pt': 0.5691413155244582, 'off_e': 0.2648183616613592, 'def_e': 0.04950147745040279}

--- Round 2 ---
Games considered: 80/80
Best accuracy:    0.762
Best weights:     {'seed': 0.05343226041594218, 'talent': 0.8417709368112914, 'height': 0.43577153930867085, 'three_pt': 0.22569369554397756, 'off_e': 0.6520602715205884, 'def_e': 0.007318834270946262}

--- Round 3 ---
Games considered: 40/40
Best accuracy:    0.700
Best weights:     {'seed': 0.018310535546602447, 'talent': 0.0010609447106811887, 'height': 0.3365189378749779, 'three_pt': 0.15186091702253945, 'off_e': 0.8060252391634497, 'def_e': 0.03532816309152664}

--- Round 4 ---
Games considered: 20/20
Best accuracy:    0.850
Best weights:     {'seed': 0.08033129538426809, 'talent': 0.18359378670759918, 'height': 0.6162259463881226, 'three_pt': 0.9709867975368319, 'off_e': 0.4906321951082053, 'def_e': 0.030355303109954046}

--- Round 5 ---
Games considered: 10/10
Best accuracy:    0.800
Best weights:     {'seed': 0.28376747039272165, 'talent': 0.034887242532204765, 'height': 0.06629860130049725, 'three_pt': 0.07493747455672228, 'off_e': 0.6074053184618162, 'def_e': 0.056222642097169095}

--- Round 6 ---
Games considered: 5/5
Best accuracy:    0.800
Best weights:     {'seed': 0.7635624918730934, 'talent': 0.6617583550197224, 'height': 0.2252837406155832, 'three_pt': 0.9455408554891561, 'off_e': 0.7269933573427976, 'def_e': 0.5739693572175714}
```



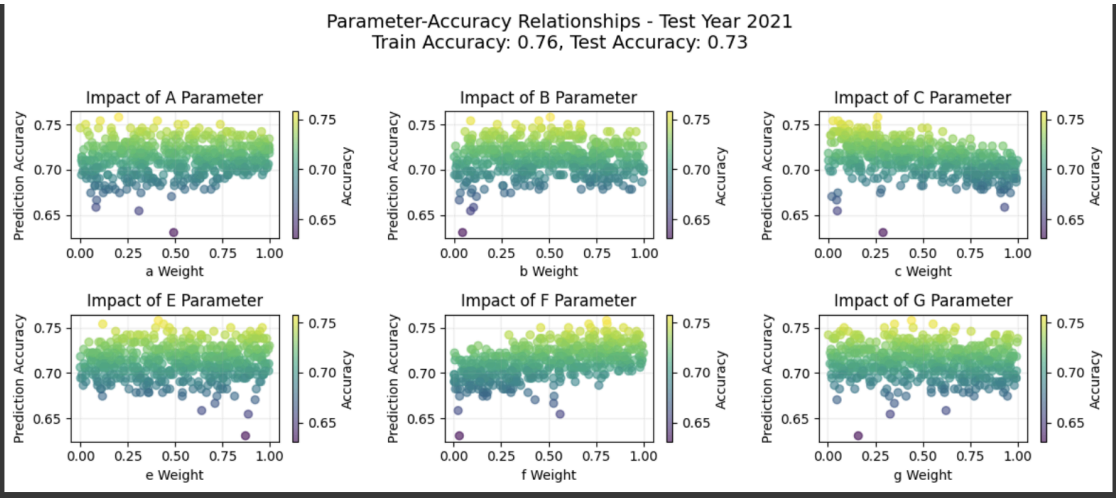
I have found that seed and defensive efficiency seem to increase as rounds go on. Talent becomes less important in later rounds. Height, 3p%, and offensive efficiency seem to be equally important early and late. This will be applied when picking the games.

I wish I would have been able to implement a better way of incorporating these results into the following equations and learn from that using specific rounds more than just being limited to early vs late.

```
   base1 = (height1_norm * c + three_pt1_norm * e + adjOE1_norm * f) * 3
   base2 = (height2_norm * c + three_pt2_norm * e + adjOE2_norm * f) * 3

   momentum1 = ((seed1_norm * a - (adjDE1_norm * g)) * round_num) +
((talent1_norm * b) * (6/round_num))
   momentum2 = ((seed2_norm * a - (adjDE2_norm * g)) * round_num) +
((talent2_norm * b) * (6/round_num))
```

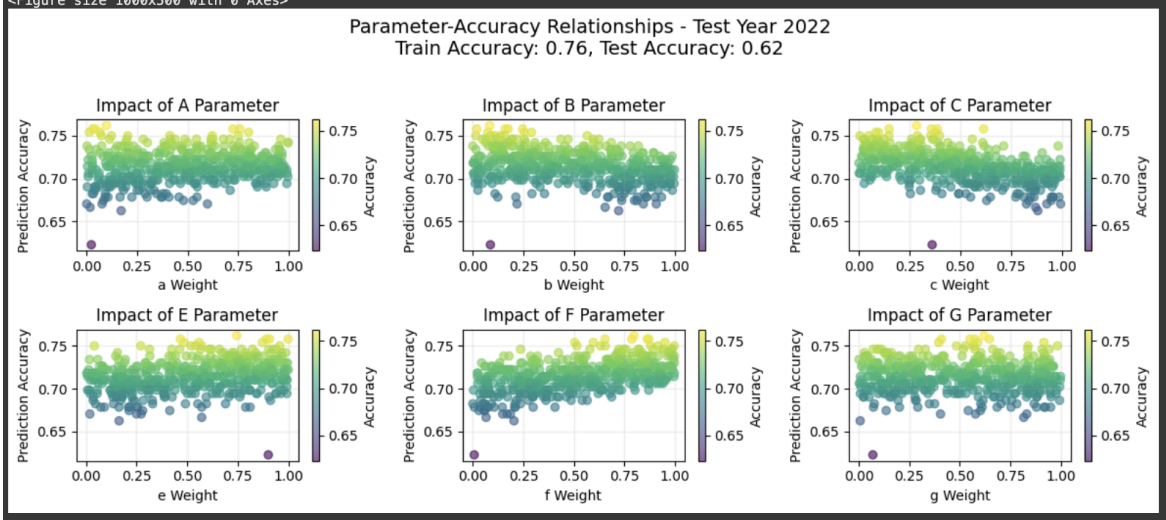
```
    return base1 + momentum1, base2 + momentum2
```

Using this equation I did a train-test split on 4 years of training and 1 year to test on (2021-2025). Simulating random weights and taking the weights of the highest accuracy simulation of the 4 training years and applying that to the test year.



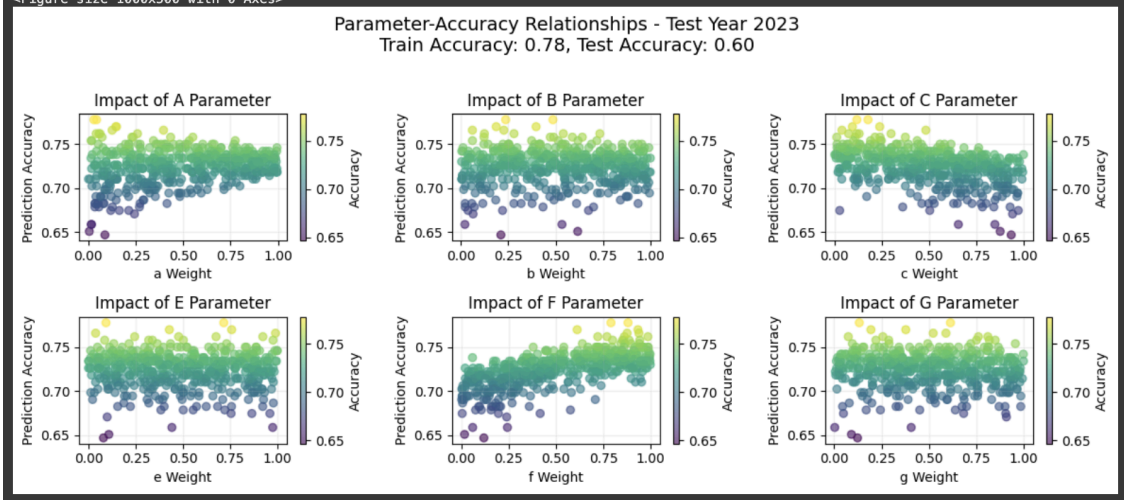

```
========================================================================

Best parameters for 2021:
{'a': 0.20343656865317172, 'b': 0.506885433766698, 'c': 0.26243388020839564, 'e': 0.41186249630816263, 'f': 0.8051239175757872, 'g': 0.4404969084973849}
Training Accuracy: 0.7579
Test Accuracy: 0.7302

Key: a=seed b=talent c=height e=3pt% f=adjOE g=adjDE
```

```
=== Processing test year: 2022 ===
<Figure size 1000x500 with 0 Axes>
```



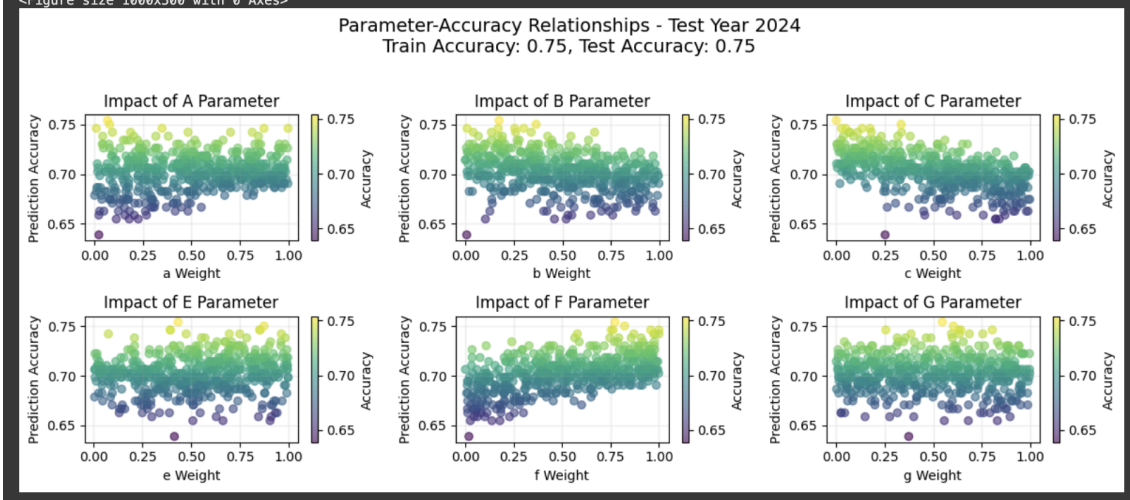

```
================================================================================

Best parameters for 2022:
{'a': 0.09895629451689936, 'b': 0.0836096645725759, 'c': 0.2848180231802159, 'e': 0.7445051173379802, 'f': 0.7980274709541819, 'g': 0.620498144947443}
Training Accuracy: 0.7619
Test Accuracy: 0.6190

Key: a=seed b=talent c=height e=3pt% f=adjOE g=adjDE
```

```
=== Processing test year: 2023 ===
<Figure size 1000x500 with 0 Axes>
```



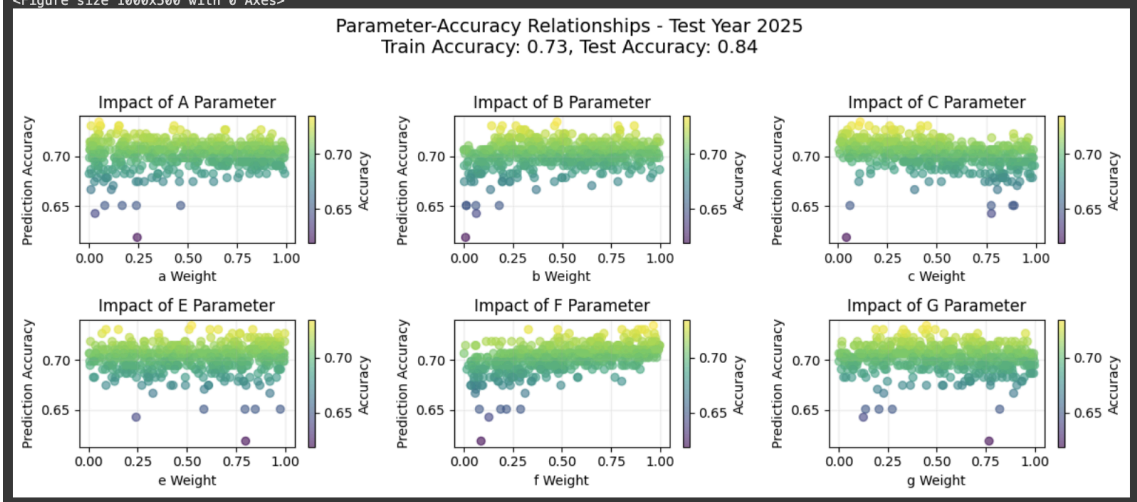

```
================================================================================

Best parameters for 2023:
{'a': 0.04374583347257399, 'b': 0.48574030851689154, 'c': 0.17771744507716747, 'e': 0.0925779533558827, 'f': 0.8808288576642733, 'g': 0.13255175837202016}
Training Accuracy: 0.7778
Test Accuracy: 0.6032

Key: a=seed b=talent c=height e=3pt% f=adjOE g=adjDE
```

```
=== Processing test year: 2024 ===
<Figure size 1000x500 with 0 Axes>
```




```
================================================================================

Best parameters for 2024:
{'a': 0.06779926153885385, 'b': 0.1732227186677735, 'c': 0.000474942028054981, 'e': 0.43304711246278715, 'f': 0.7729634860816652, 'g': 0.5465560288992112}
Training Accuracy: 0.7540
Test Accuracy: 0.7460

Key: a=seed b=talent c=height e=3pt% f=adjOE g=adjDE
```

```
=== Processing test year: 2025 ===
<Figure size 1000x500 with 0 Axes>
```




```
================================================================================

Best parameters for 2025:
{'a': 0.05048705893422156, 'b': 0.4742002831004182, 'c': 0.11339687455267156, 'e': 0.5239344648036627, 'f': 0.9669489719594054, 'g': 0.44852411262208935}
Training Accuracy: 0.7341
Test Accuracy: 0.8413
```

Major observations: There is a wide variance when seed(A) is weighted lower showing it's a stable metric but has a limited ceiling. Height(C) seems to be a weaker metric

which seems to make sense in the context of modern basketball considering how basketball is becoming a more perimeter centric game and going away from big man play. Offensive efficiency(F) seems to be the strongest metric. Also found it interesting the accuracy was higher for the test data in 2025.

To answer the research if we can pick the outcomes of march madness games: The accuracy I came up with seems to be around 75% depending on the year

and to answer the question if offense or defense are more important: offense seems to be the much more important side according to the research. Limitations include that my model ignores in-game variance factors: injuries, coaching adjustments, and game-day randomness, which could further refine predictions. I learned a lot in this class and am excited to continue to learn and work on improving my models.

Reflection on Challenges

One of the biggest challenges I faced was getting the data merged because the team names weren't always exactly the same(e.g., "San Diego St." vs. "San Diego State"). Then make sure I typed the names correctly when I put in the game results data also added to that. Another challenge I faced was the simulation time. Some of these simulations took over an hour to complete, and when I wanted to make changes I had to resimulate new results.

Lastly I thought coming up with ideas like how I was going to get the equations and which metrics I should use was very hard to come up with. I still feel like there are

many things I could do with this and improve and would love to spend time to continue to improve my metrics, methods, and keep working on this model in the future.

Citations

Dean, J. (2014, March 21). *March madness and predictive modeling*. The SAS Data Science Blog. https://blogs.sas.com/content/subconsciousmusings/2014/03/21/march-madness-and-predictive-modeling/?utm_

McIver, C., Avalos, K., & Nayak, N. (2025, March 17). *March madness tournament predictions model: A mathematical modeling approach*. arXiv.org. https://arxiv.org/abs/2503.21790?utm_

NCAA.com. (2025, April 16). *Records for every seed in March Madness from 1985 to 2025*. https://www.ncaa.com/news/basketball-men/article/2025-04-16/records-every-seed-march-madness-1985-2025

Pomeroy, ken. (n.d.). 2025 Pomeroy College Basketball Ratings. https://kenpom.com/