March Madness Seeding Analysis

By Grace Burns, Graham Dynis, Kaleb Jordan, & Isaiah Kuehl

TABLE OF CONTENTS

INTRODUCTION

The NCAA March Madness tournament is one of the most anticipated sporting events of the year, with 364 Division I college basketball teams competing for 68 coveted positions, all with the chance to win the National Championship. Each year, 31 teams earn an automatic bid by winning their respective conference tournaments. The remaining 37 at-large bids are awarded by the NCAA Selection Committee, comprised of Division I athletic directors and conference commissioners. Once the 68 teams are chosen, the Selection Committee assigns each team a seed from 1 to 16, with the four top-ranked teams receiving No. 1 seeds, the next four receiving No. 2 seeds, and so on, forming the structure of the tournament bracket.

For almost 30 years, Joe Lunardi has been the resident "Bracketologist" at ESPN, attempting to mimic the Selection Committee's decisions with his own bracket projections. While there are only a handful of professional bracketologists, there has been a rise of social media personalities and independent analysts developing their own bracketology methods, many of which can be found at bracketmatrix.com. We wanted to explore the bracketology space in depth and ultimately develop our own approach for predicting the March Madness bracket.

This project aims to develop and evaluate the most predictive models for selecting the 68 teams and determining their seeding in the tournament. By analyzing historical data collected from Bart Torvik's website, we used statistical methods to assess how accurately we can classify a team's Seed using various resume-based and quality-based metrics. This project's objective is to highlight key performance indicators, uncover existing biases, and ultimately enhance the ability to forecast future NCAA March Madness tournament seedings.

METHODS

Our data consists of two distinct datasets, the first of which contains 250 observations spanning the last five NCAA Men's March Madness tournaments (50 from each year 2021-2025), from Seeds 1-12. Using *Seed* as the target variable, we have the variables: *BPI* (Basketball Power Index), *KP* (KenPom Rankings), *KPI* (Key Performance Indicator), *NET* (NCAA Evaluation Tool), and *SOR* (Strength of Record) as potential predictors along with others relating to Quad 1 performance and the conference the teams play in. Additionally, feature engineering was conducted to identify unseen insights.

The second dataset contains 73 carefully selected observations of "bubble" teams from the past five seasons of Division I men's basketball. For these teams on the brink of qualifying as an at-large, the binary outcome *In or Out* was used as the target variable, and the same set of predictors from the previous model was considered.

An overview for some of the endogenous variables:

- **BPI:** BPI is a statistic used to determine a team's performance, rather than the result. Instead, BPI considers factors such as a team's pace (number of possessions per game) and offensive/defensive efficiency.
- **KP:** KP evaluates a team's performance by examining both offensive and defensive ratings, identifying the strengths and weaknesses of each team.
- **KPI:** KPI is a result-based statistic that evaluates team performances per game on a ±1 evaluation. It determines the margin a team wins or loses by, with 0.0 representing a near-tie.
- **NET:** NET is an in-depth measurement that involves numerous metrics to determine a team's strength and worthiness of their tournament chances.
- **SOR:** SOR is a way to determine a team's accomplishments throughout the season by investigating the difficulty of their schedule and the actual results.
- **Power Conf:** Power Conf is a binary result of a team playing in a strong conference like the Big 10, Big 12, ACC, SEC, etc.
- **W:** W is the number of wins that a team has against Quad 1 opponents in each season.
- **L:** L is the number of losses that a team has against Quad 1 opponents in each season.
- **Center within Groups:** To adjust for differences between Power Conference and Non-Power Conference teams, BPI, KPI, and SOR were centered within each group. For each variable, the group mean was subtracted from each team's value. This process standardizes the variables relative to each team's competitive environment, allowing fairer comparisons across teams with different schedules and resource levels. This ensures that a team's performance is measured relative to its peer group, preventing Power Conference teams from automatically dominating the strength-based metrics.

We used numerous algorithms and statistical techniques to determine the model that best represents the data. After careful analysis, we identified algorithms that could be useful for a project like this: a **simple linear regression** and two **ordinal logistic regression** models. These models have their own pros and cons. A simple linear regression model could be applied to ordinal outcomes like Seed, and it would be straightforward to understand, but it treats the response as continuous and assumes an equal distance between Seed thresholds. This can lead to biased or uninterpretable results (e.g., Seed 0 or Seed 13). Instead, a linear regression model will be treated as our baseline. A more appropriate approach is to use an ordinal regression model, which accounts for the ranked nature of the outcome and provides more meaningful probability-based predictions. Ordinal regression models are easy to interpret (e.g., one-unit increase in BPI affects Seed) and capture the ordinal nature of the data (Seed 1 is better than Seed 2, etc.). Ordinal regression has some disadvantages over the others, such as its flexibility, as it does not capture non-linearities or interactions. Since it captures the proportional odds, the relationship between the response variable and the predictors remains constant for all thresholds. This may be unrealistic for our data. The proportional odds model assumes that the relationship between predictors and the log-odds of being a lower seed is constant across thresholds. This assumption simplifies interpretation but may not hold if certain metrics disproportionately affect top or bottom seeds. Below is a diagram of the formula for an ordinal regression model:

An ordinal model gives you cumulative probabilities of the form:

$$\mathbf{P(Y \leq j) = 1 / (1 + \exp[-(\tau j - \eta)])}$$

To get the probability of an exact category, you take:

$$\mathbf{P(Y = j) = P(Y \leq j) - P(Y \leq j - 1)}$$

To model the selection of bubble teams, a **binomial generalized linear model (GLM)** was employed. This model is well-suited for predicting the probability of an event with two possible outcomes, in this context either making or missing the NCAA Tournament. The model estimates the likelihood of making the tournament as a probability between 0 to 1, based on various combinations of metrics.

RESULTS

A baseline linear regression model was fitted to predict NCAA tournament seedings using three predictors: BPI, KPI, and SOR. All three variables were statistically significant predictors of Seed placement, with positive coefficients indicating that higher BPI, KPI, and SOR values (e.g., worse relative rankings) were associated with higher seed numbers (worse seeds). Specifically, SOR had the largest coefficient ($\beta = 0.091$, $p < 0.05$), followed by KPI ($\beta = 0.062$, $p < 0.05$) and BPI ($\beta = 0.036$, $p < 0.05$). The model explained approximately **86% of the variance** in Seed outcomes ($R^2 = 0.8634$), suggesting a strong linear relationship between the predictor metrics and Seed assignment. However, the model's predictive accuracy was relatively limited when Seed values were rounded to the nearest whole number: only **32% (64 / 200)** of Seeds in the training data were correctly predicted.

Figure 1: Linear Regression Table of Coefficients

Variable	**Coefficient**	**P-Value**
Intercept	1.354489	< 0.05 (Significant)
BPI	0.036234	< 0.05 (Significant)
KPI	0.061527	< 0.05 (Significant)
SOR	0.090935	< 0.05 (Significant)

Figure 2: Linear Regression Formula

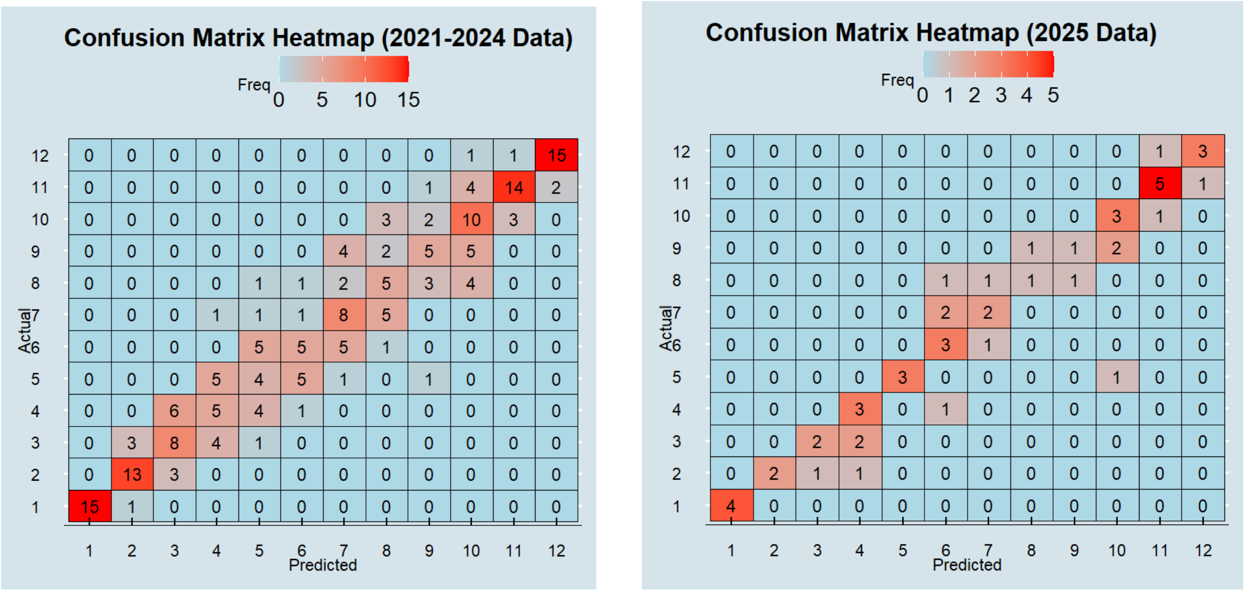
$$\hat{Y} = 1.3545 + (0.0362 \cdot BPI) + (0.0615 \cdot KPI) + (0.0909 \cdot SOR) + \in$$

When the baseline model was applied to teams from the 2025 season, prediction accuracy remained similar, correctly matching **30% (15 / 50)** of actual Seeds. This consistency suggests the baseline model's generalization performance was stable across both training and test sets.

Improving on the baseline linear model, an ordinal logistic regression model (cumulative link model) was fitted using the same predictors. Again, BPI, KPI, and SOR were all statistically significant ($p < 0.05$), with SOR ($\beta = 0.155$) demonstrating the most substantial relationship to Seed placement. The ordinal model achieved better classification performance than the baseline linear regression. The confusion matrix conveyed a stronger agreement between predicted and actual Seed values, particularly among the top Seeds (e.g., 15 out of 16 teams were correctly classified as Seed 1) in the training data and a true accuracy of **44% (88 / 200)**. Although some misclassifications occurred, most predictions fell within ±1 of the actual Seed **86.5%** of the time, or **173 out of the 200 teams**, highlighting the ordinal model's strength in capturing the ordered nature of the response variable, Seed.

Our final model is much more complex. After running a likelihood ratio test and calculating McFadden's R^2, the final complex model demonstrated a significant improvement over the original ordinal model, explaining approximately **16% more variation** in Seed assignment. This ordinal logistic regression model contains numerous variables including interaction terms and feature engineered variables. The variables: NET, KPI, SOR, W, a center within group term BPI_cwg, interaction terms W:Opp (Quad 1 Wins and Quad 1 Opportunities), KP:`Conf Champ` (KenPom ranking and Conference Champions), and lastly, KP:SOR_cwg (KenPom and center within group for SOR). All variables in the model were statistically significant ($p < 0.05$), with SOR (z value |8.626|) representing the strongest relationship to Seed. Compared to the baseline linear model and ordinal model, this more complex model showed a greater performance. The confusion matrix for this model had a true accuracy of **53.5% (107 / 200)**. The highest accuracy amongst the Seeds were for Seeds 1 and 2. The confusion matrix for the predictions falling within ±1 of the actual Seed is **89% (178 / 200)**. To assess the generalizability of the final model, a 5-fold cross-validation procedure was performed on the training dataset. In this approach, the data was partitioned into five approximately equal folds, with the model trained on four folds and evaluated on the remaining fold. This process was repeated five times, each time using a different fold as the validation set. The average cross-validated accuracy for exact Seed prediction was approximately **47%**, while the accuracy for predictions within ±1 Seed of the actual value was approximately **85%**. These results suggest that the model maintains strong performance across different subsets of the data, demonstrating robustness and also reducing concerns about overfitting the training data. The accuracy of the ordinal model for the 2025 dataset was **54% (27 / 50)**, but when allowing for a ±1 Seed margin, accuracy increased substantially to **94% (47 / 50)**. For the 2025 data, we still see a better performance with a true accuracy of **64%** or **32 out of 50**. For the classifications of ±1 Seed margin the accuracy performs like the simpler ordinal model with an accuracy of **92% (46 / 50)**.

Figure 3: Confusion Matrix Heatmaps



Confusion Matrix Heatmap (2021-2024 Data)

	1	2	3	4	5	6	7	8	9	10	11	12
12	0	0	0	0	0	0	0	0	0	1	1	15
11	0	0	0	0	0	0	0	0	1	4	14	2
10	0	0	0	0	0	0	0	3	2	10	3	0
9	0	0	0	0	0	0	4	2	5	5	0	0
8	0	0	0	0	1	1	2	5	3	4	0	0
7	0	0	0	1	1	1	8	5	0	0	0	0
6	0	0	0	0	5	5	5	1	0	0	0	0
5	0	0	0	5	4	5	1	0	1	0	0	0
4	0	0	6	5	4	1	0	0	0	0	0	0
3	0	3	8	4	1	0	0	0	0	0	0	0
2	0	13	3	0	0	0	0	0	0	0	0	0
1	15	1	0	0	0	0	0	0	0	0	0	0

Confusion Matrix Heatmap (2025 Data)

	1	2	3	4	5	6	7	8	9	10	11	12
12	0	0	0	0	0	0	0	0	0	0	1	3
11	0	0	0	0	0	0	0	0	0	0	5	1
10	0	0	0	0	0	0	0	0	0	3	1	0
9	0	0	0	0	0	0	0	1	1	2	0	0
8	0	0	0	0	0	1	1	1	1	0	0	0
7	0	0	0	0	0	2	2	0	0	0	0	0
6	0	0	0	0	0	3	1	0	0	0	0	0
5	0	0	0	0	3	0	0	0	0	1	0	0
4	0	0	0	3	0	1	0	0	0	0	0	0
3	0	0	2	2	0	0	0	0	0	0	0	0
2	0	2	1	1	0	0	0	0	0	0	0	0
1	4	0	0	0	0	0	0	0	0	0	0	0

Figure 4: Ordinal Logistic Regression Table of Coefficients

Variable	**Coefficient**	**P-Value**
NET	0.06724	< 0.05 (Significant)
KPI	0.06832	< 0.05 (Significant)
SOR	0.23140	< 0.05 (Significant)
W	-1.12937	< 0.05 (Significant)
BPI_cwg	0.06147	< 0.05 (Significant)
W:Opp	0.03512	< 0.05 (Significant)
KP:`Conf Champ`	0.04984	< 0.05 (Significant)
KP:SOR_cwg	-0.00252	< 0.05 (Significant)

Figure 5: Thresholds Coefficient Table

Threshold	Estimate
1 | 2	-4.929
2 | 3	-1.978
3 | 4	0.450
4 | 5	2.557
5 | 6	4.535
6 | 7	6.202
7 | 8	7.732
8 | 9	9.159
9 | 10	10.663
10 | 11	12.814
11 | 12	17.745

For the bubble team selection model, we tested ten different combinations of predictors to identify the most effective subset for final model selection. Each combination yielded varying degrees of predictive accuracy. The initial model used the same baseline predictors as the seeding model. In this configuration, SOR emerged as a statistically significant predictor ($p < 0.05$). However, unlike in the original model, both BPI and KPI were not statistically significant, with their estimated coefficients near zero, which indicates little to no influence. This suggested that stronger models were likely achievable beyond the baseline.

Figure 6: Binomial Baseline Model Table of Coefficients

Variable	Coefficient	P-Value
Intercept	7.681585	< 0.05 (Significant)
BPI	0.005474	0.79394
KPI	-0.017352	0.55227
SOR	-0.143697	< 0.05 (Significant)

A second approach involved feature engineering, introducing two new composite variables: *R.AVG* (Resume Metric Average) and *Q.AVG* (Quality Metric Average). Metrics in the dataset were categorized into two groups: resume-based (SOR, KPI, WAB) and quality-based (BPI, KP, SAG, TRK). By averaging the metrics within each group, we aimed to capture a more comprehensive yet simplified representation of a team's profile. In this model, R.AVG was a significant predictor ($p < 0.05$), while Q.AVG had little effect on bubble team selection.

Figure 7: Binomial Metric Averages Model Table of Coefficients

Variable	Coefficient	P-Value
Intercept	7.25013	< 0.05 (Significant)
R.AVG	-0.13366	< 0.05 (Significant)
Q.AVG	-0.01085	0.66789

Subsequent models incorporated a Major Conference dummy variable, Quad 1 Wins and Quad 1 Win Percentage as additional predictors. However, these models did not demonstrate improved performance over the earlier approaches. In a final evaluation, we assessed the seeding model's utility in predicting the bubble by using the same set of predictors, excluding KP:`Conf Champ`, since none of the bubble teams in the dataset won their conference tournament. In this approach, only Quad 1 Wins and the interaction term between Quad 1 Wins and Quad 1 Games were significant at the 0.05 level.

Figure 8: Binomial Final Model Table of Coefficients

Variable	Coefficient	P-Value

Intercept	5.153856	0.5494
NET	-0.089878	0.0893
KPI	-0.019376	0.5964
SOR	-0.042867	0.8074
Q1.W	1.765240	< 0.05 (Significant)
BPI_cwg	0.063601	0.1316
Q1.W:Q1.Opp	-0.097257	< 0.05 (Significant)
KP:SOR_cwg	-0.001753	0.6302

To evaluate the performance of each model, we calculated the **Brier Score**, which measures the mean squared error between the predicted probabilities and actual outcomes in the test set of 2025 bubble teams:

$$Brier\ Score = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2$$

where p_i is the predicted probability, y_i is the actual outcome (1 if the team made the tournament, 0 otherwise), and N is the total number of observations (N = 13). A Lower Brier Score indicates better predictive accuracy, with a score of 0 representing perfect prediction. Among all the models tested, the **Metric Averages** model (described in Figure 7) achieved the lowest Brier Score, displaying the best performance in predicting the bubble.

Figure 9: Brier Score Results

Model	**Brier Score**
Baseline	0.206
Metric Averages	0.179
Seeding Final	0.214

Across all ten bubble selection models tested, each correctly identified **6 of 8 (75%)** available at-large bubble spots by selecting the teams with the highest predicted probabilities of making the tournament. When combining the results of both the seeding and bubble selection models, the overall approach correctly predicted **66 of the 68 (97%)** teams that made the NCAA Tournament in 2025.

DISCUSSION

The analysis conducted offers insight that should be considered when seeding teams for the March Madness tournament, a process that has not relied too heavily on statistics in the past. Our models address the concerns that college basketball fans have about the Selection Committee's decisions as well as explores the ever-growing field of bracketology. The models we chose to investigate use objective, quantifiable metrics to predict, with accuracy, tournament seedings as well as at-large bids. The controversial topic of whether to rely more on statistics versus human judgement or a 'feel' in the realm of sports is what makes this analysis important. The analysis highlights the necessity of utilizing a focused set of statistical variables when determining tournament seeds while also acknowledging their limitations.

Two resume-based metrics (SOR and KPI) as well as one quality-based metric (NET) stood out as significant predictors of seed number in the tournament. The simple regression model that consisted of these three variables had a relatively solid performance, while the more complex ordinal logistic model boasted even *more* success when incorporating these three variables—achieving 94% accuracy within a ±1 seed margin for the 2025 data.

The models focused on bubble teams also provided insight for seeding teams, specifically the last four in and first four out—another ambivalent aspect of bracketology. Through the analysis of binomial models, we discovered that resume-based metrics, especially composite averages of key indicators, were the best in predicting seeds for teams on the brink. Although it is near impossible to forecast human decisions, the strong performance of the Metric Averages model (lowest Brier Score) proves that even the most disputed selections are *capable* of being predicted.

An overview of the results concludes that seed assignment can be explained by a combination of both resume-based and quality-based metrics, with models improving when nuanced features (adjustment for strength of conference, interaction terms, etc.) are added. Although, unfortunately, bias can never be completely eradicated and will always play an interesting role in this topic, we observed that it can be significantly reduced with ample data analysis. Finally, we noted that predictive modeling possesses consistency across different seasons, meaning that the factors that the Selection Committee values the most have not changed in the past few years. Shifting the trajectory of the committee ever so slightly so that data analysis is considered even just *a bit* more in the future will hopefully lead to achieving as much fairness and transparency as possible in the madness that surrounds March.

IMPLICATIONS

This predictive model has much upside potential. This is a resource that the committee could utilize in a more accurate and statistically-based way to select the seeds of teams—especially which teams are the last four in or first four out in the NCAA tournament. While we are in no way implying that the committee does a poor job of seeding the March Madness tournament, both die-hard and casual fans would agree that certain decisions could be improved upon. This discontent between fans and the committee could be resolved, to some extent, by using a predictive metric to aid in the seeding of the tournament.

LIMITATIONS

The largest limitation to this sort of question seems somewhat apparent; however, there is more than just one. When predicting seeding for the March Madness tournament, we are using physical data and numbers to foresee human decisions. While this has proved itself to be very accurate, we do acknowledge where the predictions fall short. We were able to correctly predict the seed to a ±1 seed margin about 94% of the time with our best model. However, especially when it comes to bubble teams, the accuracy dips. The standards of models and die-hard fans chalk this decrease in accuracy up to a team getting "snubbed" by the Selection Committee. We added more statistics in an attempt to account for this, but even then, they were unsuccessful. This leads us to the other limitation we had in this project, which was that there were only so many variables we could use to predict the seeding of the NCAA tournament. While we found an adequate number of predictive variables, still some of those variables could be viewed as biased. For example, the AP Poll is also determined by a committee of people, posing the concern of impartiality.

FUTURE DIRECTIONS

Instead of predicting seeds with a ±1 seed margin of error, we want this project to eventually achieve perfect seed projection. While a ±1 seed margin should not be frowned upon, it would be ideal if a model could calculate the seed of each team, including the bubble teams, with complete accuracy. Another aspect of this project that could be improved in future models is the incorporation of the prediction of 13-16 seeds. However, because these seeds are more difficult to predict because of their tendency to be on the bubble, this could impact the accuracy of the project. It would be both important and necessary to adjust the models, specifically the bubble models, accordingly if planning to include 13-16 seeds.

The models included in this project could also be combined or improved by creating new models using random forest, gradient boosting machines, or neural networks. These methods, as well as the ones we utilized, possess even more room for improvement if a larger data set were to be at our disposal, increasing sample size and leading to better model generalizability. Finally, along with data including more seasons, expanding the project with the integration of more variables like win and loss margins, wins versus Top 25 teams, and other metrics contribute to the reliability of the project.

SOURCES

- Torvik, Bart. "NCAA Teamsheet Ranks." barttorvik.com, https://www.barttorvik.com/teamsheets.php?year=2025, Accessed April 10, 2025.

- Coleman, B. & DuMond, J.. (2016). An easily implemented and accurate model for predicting NCAA tournament at-large bids. Journal of Sports Analytics. 2. 1-12. https://www.researchgate.net/publication/307615009_An_easily_implemented_and_accurate_model_for_predicting_NCAA_tournament_at-large_bids.

- Paul, R. J., & Wilson, M. (2012). Political Correctness, Selection Bias, and the NCAA Basketball Tournament. Journal of Sports Economics, 16(2), 201-213. https://doi.org/10.1177/1527002512465413 (Original work published 2015)