Developing the Optimal Bracket Strategy for the NCAA Men's Basketball Tournament using Monte Carlo Simulation

Abstract

The NCAA basketball tournament is a single-elimination, 64-team tournament where teams compete to win the national championship (Wilco, 2023). Colloquially known as "March Madness" it has become popular to try to pick which team will advance to each round in the tournament to beat other users' brackets (Wilco, 2023). This involves attempting to make accurate predictions of which team will advance, but mainly the goal can be seen as only needing to be more accurate than other users. This begs the question: is there a strategy for filling out brackets that consistently outperforms other users? The purpose of this study was to determine a strategy for filling out brackets to beat other users using game theory and Monte Carlo simulation, and then implement these decisions in the 2023 NCAA tournament. The simulated results of this study showed that there is a strategy for picking teams that outperform the average user's bracket, which was created by testing various strategies and taking the most accurate predictions in each round to form a combined strategy. By utilizing mathematical and analytical models, this approach for selecting teams serves as a possible powerful tool for beating other users' brackets.

Keywords: NCAA, tournament, simulation

Review of Related Literature

NCAA Tournament Background

The NCAA basketball tournament is a single-elimination, 64-team tournament where teams compete to win the national championship (Wilco, 2023). Once the teams are selected for the tournament, each one is given a seed one through 16 and placed into one of four regions (Wilco, 2023). In this seeding, a one seed is considered the best team in the region while a sixteen seed is considered the worst (Leonhart, 2015). Additionally, each of the four one seeds are ranked so that the best overall team competes in the top left region against the worst one seed in the bottom left region (Leonhart, 2015). Colloquially known as "March Madness" it has become popular to try to pick which team will advance to each round in the tournament to beat other users' brackets (Wilco, 2023). Users fill out brackets on websites such as CBS or ESPN where the point totals for their brackets are calculated in real-time and ranked against other users' brackets (Wilco, 2023). Therefore, the goal of filling out a bracket becomes to make predictions of which teams will advance to each round to gain the most points possible and beat other users.

The NCAA tournament has become known as "March Madness" due to unexpected outcomes that often occur during the tournament (Wilco, 2023). There is a roughly 1 in 9.2 quintillion chance of selecting a perfect bracket by randomly deciding each matchup, or roughly a 1 in 120 billion chance using some basic strategies to do so (Wilco, 2023). These nearly impossible odds and seemingly random outcomes that occur during the tournament are part of what makes filling out NCAA tournament brackets exciting for users. Not only is there a competitive aspect from beating other users, but excitement also comes from attempting to make correct picks that are difficult to predict. This has led to the development of strategies to attempt to correctly make bracket picks in any given year to strive for perfection (Mass, 2023; Ota, 2023; Boice & Silver, 2018; Dutta et al., 2017; Niemi et al., 2008). While devising a strategy to create a perfect bracket may be a futile task, creating one which accurately models the seeming randomness of this tournament could give users an edge.

Bracket Strategies

To fill out brackets, a number of strategies have been developed to attempt to correctly select teams in any given year that range in complexity (Mass, 2023). Notably, the simplest strategy is picking a random winner for each game using equal likelihoods, similar to flipping a coin (Mass, 2023). This strategy requires no knowledge of how basketball or the NCAA tournament work making it easy to implement, but possibly also inaccurate. Furthermore, another simple and popular strategy is to choose the higher seed to win each game (Mass, 2023). Given that the higher a team is seeded the better they should be relative to the other teams in the tournament, selecting the higher seed team in each matchup is a theoretically sound approach (Mass, 2023). Both of these strategies involve little knowledge or effort to make bracket picks, which if they provide accurate predictions could yield a simple yet effective approach.

Other strategies also exist which involve using externally calculated likelihoods for each team to advance to each round to fill out brackets (Ota, 2023; Boice & Silver, 2018). Notably, ESPN publishes a spreadsheet of the percentage of users who picked each team to advance to each round prior to the tournament beginning (Ota, 2023). Using these probabilities to make picks would ensure that a bracket does as well as the average user on ESPN, and while it involves a little more research it is still a quite simple strategy to implement. This strategy does have a notable feasibility issue, however, due to the published table being live data which is collected in real-time. Given that approximately 20 million brackets are submitted within the span of four days, however, it can be assumed that a sufficient sample size will be collected in a rather short period of time (Ota, 2023). Similarly, FiveThirtyEight, an analytical blogging site, publishes a spreadsheet before each tournament with their probabilistic forecasts for each team to advance to each round (Boice & Silver, 2018). FiveThirtyEight’s model is notably more complex than other approaches previously discussed and is based on a number of externally calculated power ratings for each team (Boice & Silver, 2018). While this approach is complex for FiveThirtyEight to model and predict, it is still simple for users to take FiveThirtyEight’s predictions and make bracket selections. Given that this approach utilizes more advanced analytics, it is possible that it could be effective for predicting the behavior of teams in the tournament. Both of the aforementioned strategies involve more research compared to the simpler approaches previously discussed, yet most of the analytics needed to implement each strategy are done by external sources yielding approaches that are still quite simple for users.

The most complex approach to making tournament selections involves internally creating an analytical model to predict the likelihood a team advances to each round, using some factors of team performance (Dutta et al., 2017). A method to do so could involve using team power ratings from a highly regarded source such as Ken Pomeroy, a college basketball statistician, to build a model which predicts matchups in the tournament (Pomeroy, 2012). Kem Pomeroy’s Ratings, sometimes known as KenPom Ratings, are considered to be highly accurate in predicting game outcomes and therefore could be a good source for building a probabilistic model in this context (Dutta et al., 2017). If KenPom ratings are truly more accurate than other rating systems and a reasonable probabilistic model can be fit using this data, then this approach could have the benefit of harnessing advanced analytics to make possible accurate predictions. This approach is by far the most complex of any discussed thus far for making bracket selections, however, as it involves internal processing and analytics before teams can be selected.

Some hybrid strategies exist as well, which utilize theories and predictions from various sources to create new approaches for filling out brackets (Niemi et al., 2008; Gibbons, 1992). Notably, game theory can be applied to exploit the possible gains of correctly picking a team with the possible losses of other users incorrectly picking a team (Gibbons, 1992). If a strategy can be developed which selects teams that exploit other users’ selections to maximize the possible gain of a correct pick, this could yield an approach that performs quite differently than the average user (Niemi et al., 2008). While this approach is slightly complex as it would involve gathering data from various sources and choosing teams based on an internally calculated metric, it could yield a high-risk, high-reward strategy (Niemi et al., 2008). Lastly, there is a chance that of all the strategies discussed thus far certain approaches may perform better in some rounds of the tournament relative to others. In this case, it is possible that the most accurate approach in each round of the tournament could be utilized to create a combined strategy that harnesses the benefit of multiple different theoretical approaches to filling out brackets. While this involves some complexity in sorting accuracies, possibly breaking ties, and combining bracket selections, the possible reward of combining the most accurate elements of various strategies is a benefit that no previously discussed strategy has. These two approaches are moderately complex and require the combination of data, theories, and approaches to make new strategies, however, the possible benefits of harnessing multiple methods into a hybrid approach may be significant.

Purpose of the Study

The purpose of this study was to determine a strategy for filling out brackets to consistently beat other users. Seven approaches were constructed, their performance on historical data was tested using Monte Carlo simulation, and their approaches were implemented for the 2023 NCAA Tournament. Specific research questions included:

1. Which historically simulated strategy performs the best? Specifically, which strategy is expected to yield the most points?
2. Does the historical simulated performance of the best strategy significantly differ from that of an average user's bracket?
3. How does each of these strategies perform in the 2023 NCAA Tournament?

Methodology

Strategies

The first strategy that was used to select teams to advance in the tournament was randomly choosing the winner of each game using equal likelihoods. For this approach, the first-round matchups were set, and a random number $U \sim \text{Unif}(0,1)$ was selected for each matchup. When visualizing the bracket, if $U \geq 0.5$ then the upper team advanced and if $U < 0.5$ the lower team advanced.

The second strategy used was choosing the higher-seeded team to win each game, known as a "chalk" strategy. The higher-seeded team was advanced for each matchup in each round, and once the final four one seeds remained the teams in the upper quadrants advanced and the team from the upper left quadrant won. This was chosen because one seeds are ranked from highest to lowest in the upper left, upper right, bottom right, and bottom left quadrants, respectively (Wilco, 2023).

The third strategy implemented was utilizing ESPN's percentages of users who picked each team to advance to each round, known as the "Who Picked Whom" (ESPN, 2023). For this approach, the first-round matchups were set, and the team with the higher chosen percentage in each round was advanced to the next round.

The fourth strategy implemented was utilizing FiveThityEight's predictions of the probability of each team to advance to each round (FiveThirtyEight, 2023). For this approach, the first-round matchups were set, and the team with the higher probability from FiveThirtyEight in each round was advanced to the next round.

The fifth strategy that was implemented was utilizing KenPom's ratings to fit a logistic regression model to predict the probability a team would beat another team. To fit this logistic regression model, KenPom ratings from the past four years were used as the predictor variables. Additionally, the outcome of tournament games from the same four years was used as the respnse variable. Training and testing data sets were created, and the logistic regression model was fit on the training set using elastic net regression to regularize a large number of parameters in the logistic regression model. Next, predictions were made for each matchup using the trained logistic regression model and if the first team had a probability of ≥ 0.5 then they advanced to the next round, otherwise, the second team advanced.

The sixth strategy used to make picks was finding the biggest leverage between FiveThirtyEight's predicted probability and ESPN's user percentages for each team in each round, weighted by FiveThirtyEight's predicted probability. This can be viewed using the following equation:

$$Leverage = (p(FiveThirtyEight) - p(ESPN)) * p(FiveThirtyEight)$$

This value finds the biggest difference between FiveThirtyEight's predicted probability and ESPN's user percentages as possible leverage points to gain an edge over the average ESPN user. This value is then weighted by FiveThirtyEight's predicted probability to balance points of high leverage with predictions of high likelihood from FiveThirtyEight's predictions. Once the leverage for each team in each round had been calculated, picks were made in reverse order starting by choosing the champion with the highest leverage and working through preceding rounds. This was done since picks in further rounds are worth more points, making choosing the winner of the national championship game with the highest leverage the most valuable pick.

The final strategy used was implemented after the picks from the first six strategies had already been run through the Monte Carlo simulation. The simulated accuracy of each strategy in each round was calculated, and the picks from the most accurate strategy in each round were selected to create a combined approach. Since choosing the winner of the national championship is the most valuable pick, this process started by choosing the strategy with the highest accuracy in the national championship game and working backward until all picks had been made. In the case of ties in accuracies, a priority of strategies was set to select strategies that tended to differ from the average ESPN user's picks the most: the difference, FiveThirtyEight, KenPom, random, chalk, and then ESPN strategies.

Data Sets

The first data set required was ESPN's "Who Picked Whom" data set, which was retrieved as historically far back as possible: the 2023, 2022, 2021, 2019, and 2018 tournaments. This data was readily available on ESPN's website to be converted into a .csv file, but since the start of the 2023 tournament is no longer available online past 2022 and has been stored locally. This data set (see Appendix A) contains the team seed, name, and percentage of users who selected the team to advance in each round (ESPN, 2023).

The second data set required was FiveThirtyEight's predicted probabilities data set, which was retrieved for the 2023, 2022, 2021, 2019, and 2018 tournaments. This data was readily available on FiveThirtyEight's website to be converted into a .csv file and was only retrieved from 2018-2023 since this is the range available for EPSN's data set. This data set (see Appendix B) contains the team seed, name, predicted probability to advance in each round, the date the forecast was created, and other miscellaneous information (FiveThirtyEight, 2023).

The third data set required was KenPom's team ratings data set, which was retrieved for the 2023, 2022, 2021, 2019, 2018, and 2017 tournaments. A web scraper was created internally to easily access this data and was retrieved from 2017-2023 as this was as long as Ken Pomeroy's has published the same types of ratings every year. These ratings (see Appendix C) consisted of a team's adjusted offensive efficiency, adjusted defensive efficiency, adjusted tempo, luck, adjusted offensive efficiency of opposing offenses, adjusted defensive efficiency of opposing offenses, and non-conference strength of schedule rating (KenPom, 2023).

The fourth data set required was the historical tournament matchups for the 2022, 2021, 2019, 2018, and 2017 tournaments. A web scraper created internally using the website SportsReference was required to easily access this data and was retrieved from 2017-2022 to match the range of KenPom's data to fit the logistic regression model (Sports Reference, 2022). This data set (see Appendix D) consisted of a team's seed, name, and opponent in each round.

The final data set required was the historical win probabilities of each seed versus every seed they have faced. This data was readily available on the website Mcubed.net to be converted into a .csv file (Mcubed.net, 2023). This data set (see Appendix E) consisted of a first seed, a second seed, and the first seed's win percentage over the second seed (Mcubed.net, 2023).

Data Analysis

To determine which strategy performs the best, the picks from each strategy for the 2018-2023 tournaments were simulated using Monte Carlo simulation. For this simulation model, a random number $U \sim \text{Unif}(0,1)$ was selected for each matchup. If U was ≥ the probability that the higher seeded team beat the lower seeded team ($U \geq$ p(higher seed winning) then the lower seeded team advanced, and if $U <$ p(higher seed winning) then the higher seeded team advanced. In the case of ties, if $U \geq 0.5$ then the upper team advanced, and if $U < 0.5$ then the lower team advanced. Once all the bracket picks had been simulated, the total points each strategy would have gained for the simulated outcomes and their accuracy in each round were calculated. A correct pick in the first round was worth 10 points, and this amount doubled for every subsequent round. This process was repeated 1000 times to gather a sufficient sample size.

After the simulations had been completed for each year, the point totals for each strategy in each year were tested for normality before further statistical inferences were made. For this test, the null hypothesis was that the data came from a normal distribution, while the alternative hypothesis was that the data did not come from a normal distribution. For this test, a significance level of 0.05 was set, and if the associated p-value was below this level we would reject the null hypothesis and conclude that the data was non-normal. The results of this test were used to determine the proper statistical procedure to use for further inferences.

The mean and standard deviation of the total simulated points from each strategy were calculated, along with the probability that the simulated points from each strategy was greater than that of the average ESPN user: p(E[Strategy] > E[ESPN]). If the distribution of the data was determined to be non-normal, then 95% confidence bootstrap intervals were calculated for the mean and standard deviation. Additionally, in this case the cumulative density function from the skew-normal distribution was used instead of the normal distribution. The results from these calculations would reveal the simulated performance of each strategy, as well as the probability that they would outperform the average ESPN user.

The final test that was done was a paired t-test to determine if the simulated performance of the best-performing strategy was significantly different from that of the average ESPN user for each year. If it was determined that the distribution of the data was non-normal, then a Wilcoxon signed-rank test was used. For this test, the null hypothesis was that the mean difference between the sample means was zero, while the alternative hypothesis was that the mean difference was not zero. For this test, a significance level of 0.05 was set, and if the associated p-value was below this level we would reject the null hypothesis and conclude that the two-sample means were statistically different.

Results

Each strategy was implemented, and the simulation was run for historical NCAA Tournaments, namely the 2018, 2019, 2021, and 2022 tournaments. The 2020 tournament was skipped, as it was cancelled due to COVID-19. For each of these tournaments, a histogram was created of the distribution of simulated point totals, the mean, standard deviation, normality test p-value, probability of scoring higher than the average user’s bracket, and Wilcoxon Signed-Rank test p-value were reported for each strategy. Additionally, a plot of the accuracy of each strategy in each round was chosen, along with which strategies were used in the combination approach. Furthermore, the simulated point totals were compared to the actual point totals each of these strategies would have received in the real tournament. The simulated results from the four years examined can be seen below:

Figure 1
Simulated Results from the 2018 NCAA Tournament

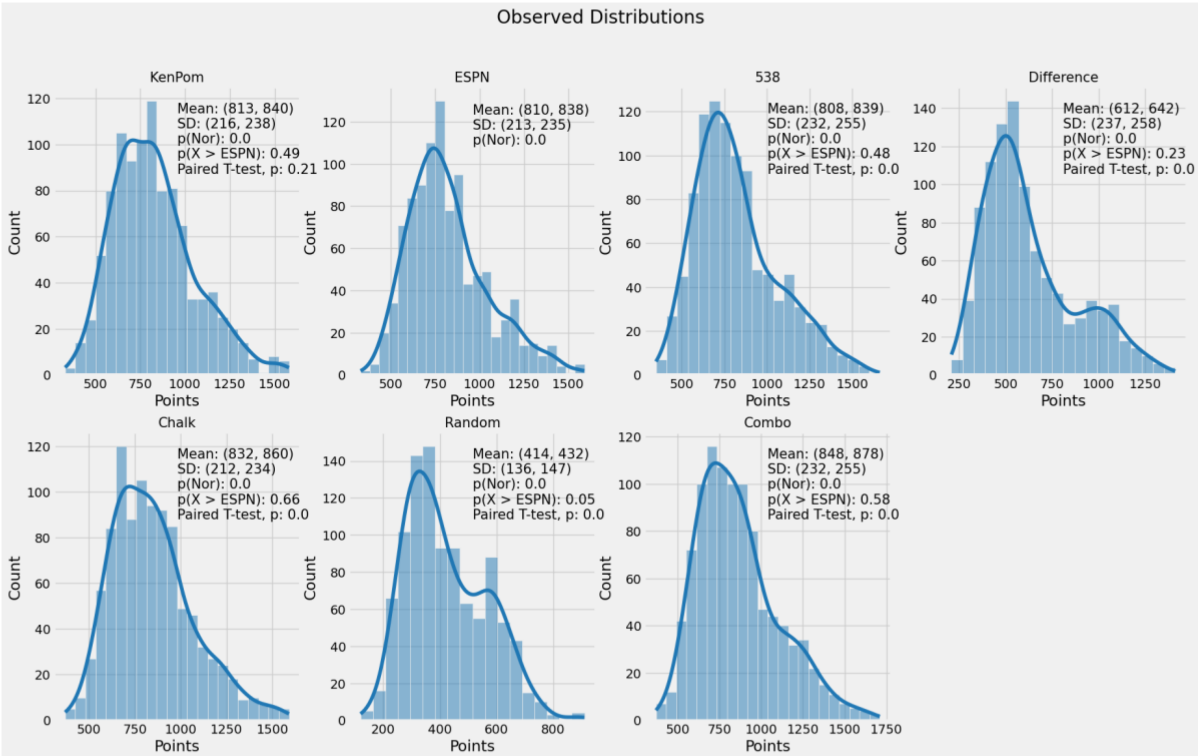


Figure 2
Simulated Results from the 2019 NCAA Tournament

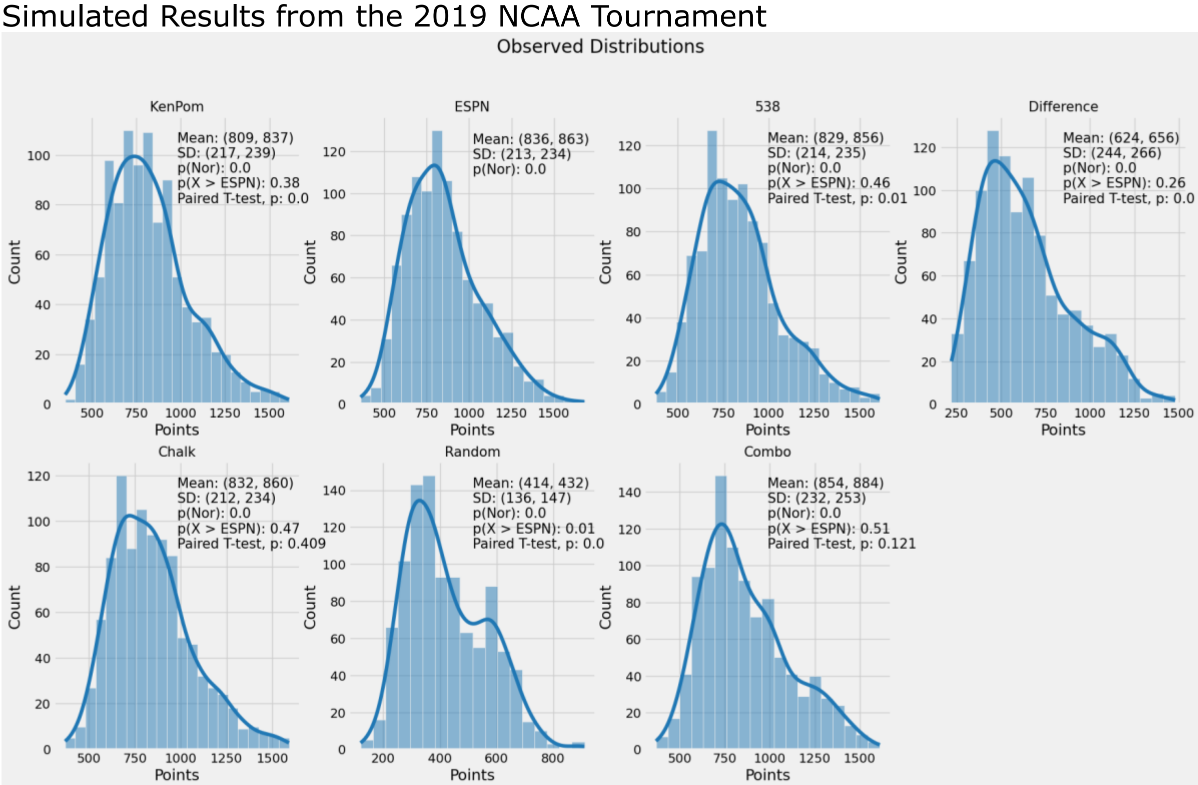


Figure 3
Simulated Results from the 2021 NCAA Tournament

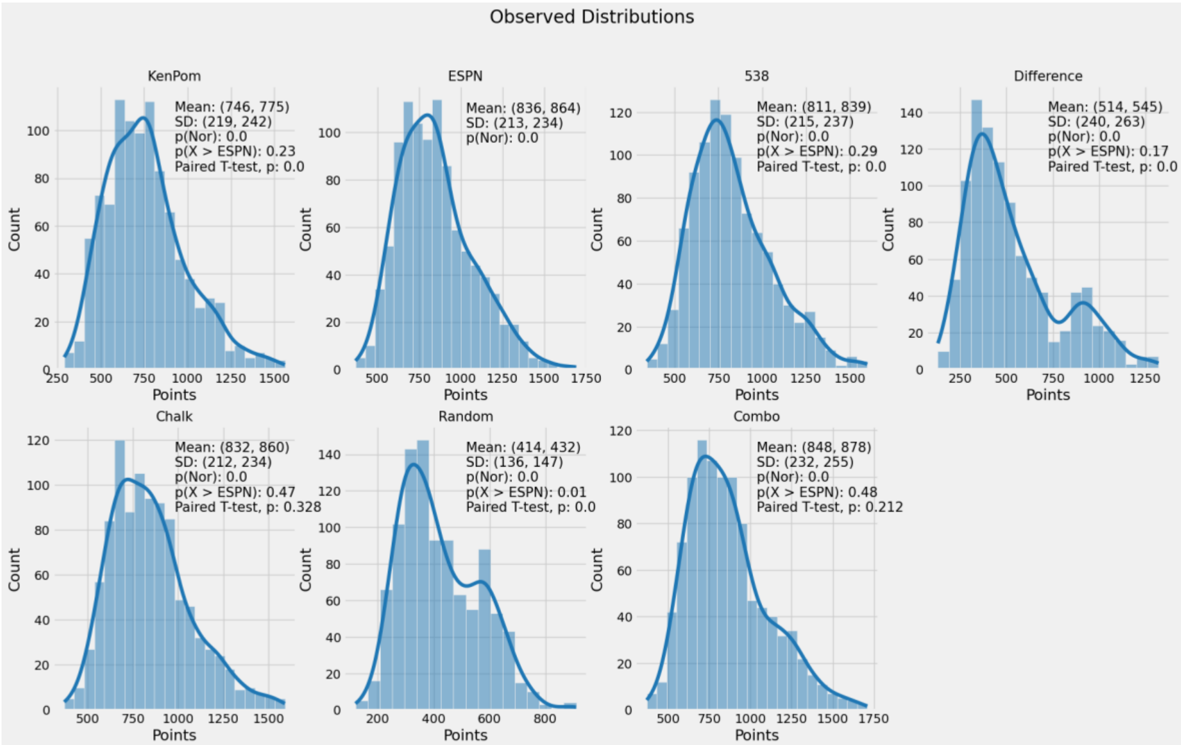
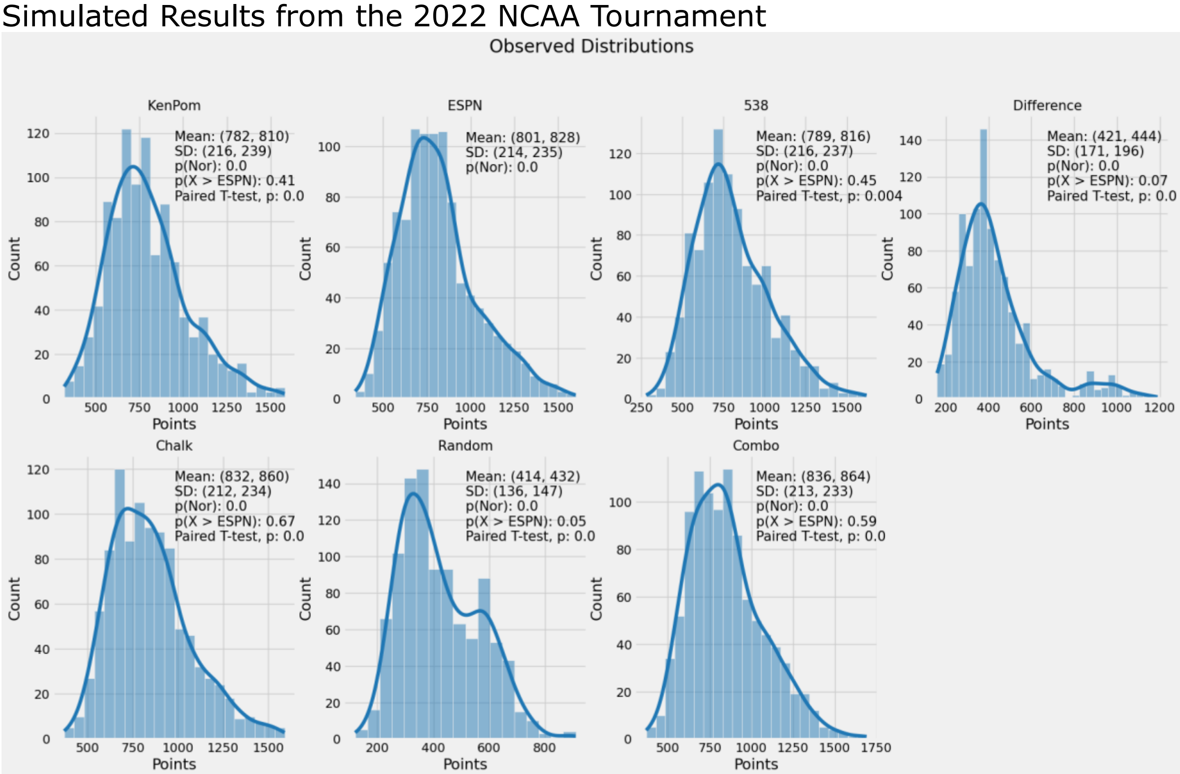
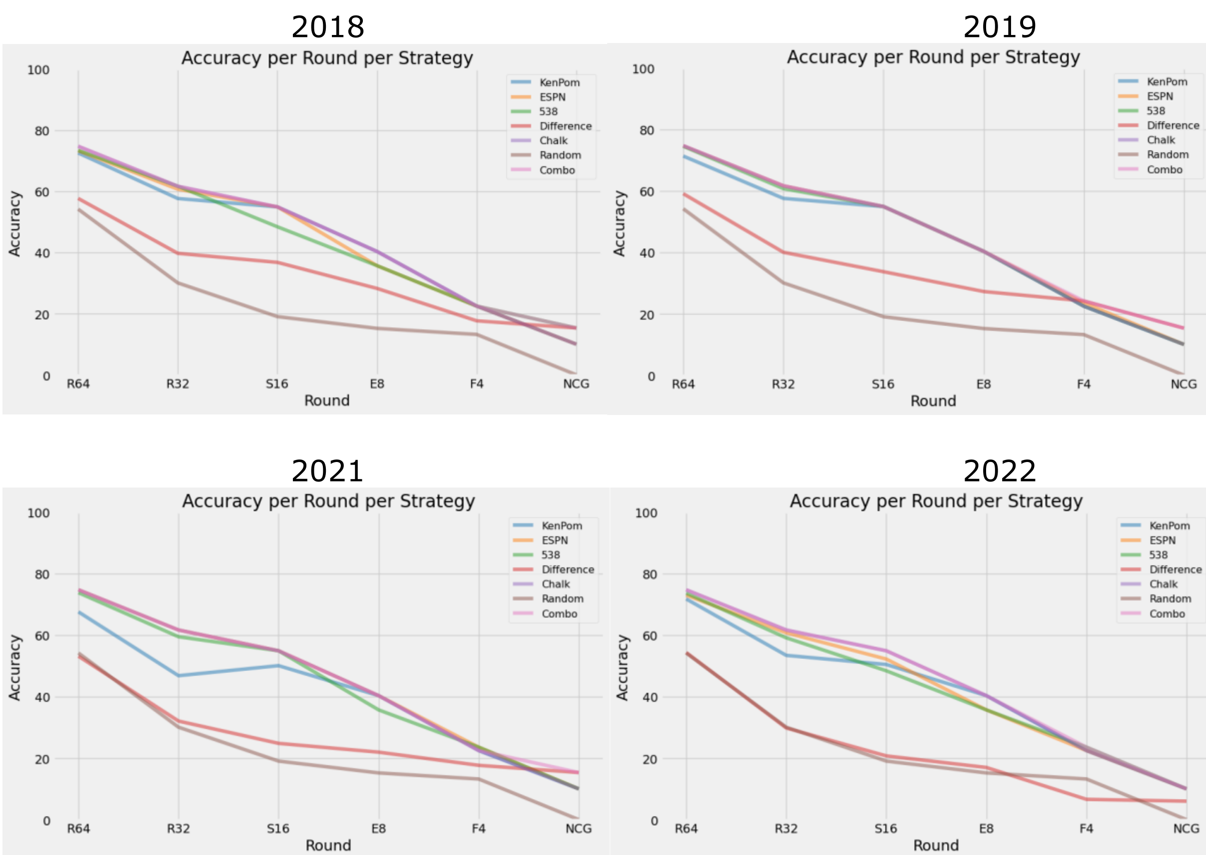


Figure 4
Simulated Results from the 2022 NCAA Tournament



As it can be seen, for each of the strategies in each year the p-value associated with the test that the distribution of the data was normal was approximately 0. Because of this, it was concluded that each distribution was non-normal which can be clearly seen in the associated histograms as most of the distributions are right-skewed. Since these distributions were non-normal, 95% confidence bootstrap intervals were created for the mean and standard deviation of each strategy's simulated point totals. For every year examined, it can be seen that the combination strategy had the highest expected point total, but also the highest standard deviation of the strategies tested. This strategy had a 58%, 51%, 48%, and 59% chance of scoring higher than the average user represented as the ESPN strategy in each year, respectively. Using the Wilcoxon Signed-Rank test to test the significance of this difference, the p-value associated with the difference between the combination and ESPN strategy for each year was ~0, 0.121, 0.212, and ~0, respectively. Additionally, the chalk strategy had the second highest expected point total with a lower standard deviation than the combination strategy. This strategy had a better chance at scoring better than the ESPN strategy than the combination approach in 2018 and 202022 due to its comparable expected point total but lower standard deviation and was statistically different from the ESPN strategy in both years. The simulated accuracies of each year, model, and round can be seen below:

Figure 5
Simulated Accuracies



From this, it can be seen that the KenPom, ESPN, FiveThirtyEight, chalk, and combination strategies all performed quite similarly, with the difference and random strategies performing noticeably worse. The only exception to this, however, is the difference strategy's accuracy in the national championship game, which was the highest in ever year except 2022. Using these accuracies, the following combination strategies were used:

2018
R64: Chalk, R32: 538, S16: KenPom, E8: KenPom, F4: 538, NCG: Difference.
2019
R64: Chalk, R32: Chalk, S16: 538, E8: 538, F4: Difference, NCG: Difference.
2021
R64: Chalk, R32: Chalk, S16: 538, E8: KenPom, F4: 538, NCG: Difference.
2022
R64: Chalk, R32: Chalk, S16: Chalk, E8: KenPom, F4: 538, NCG: 538.

The actual results for each strategy in each year can be seen below:

Table 1
Observed Results

Strategy	2018 Points	2019 Points	2021 Points	2022 Points
KenPom	850	1020	1030	650
ESPN	830	770	790	590
FiveThirtyEight	1140	950	820	790
Difference	1050	1290	980	280
Chalk	810	920	950	620
Random	660	370	150	590
Combination	1130	1240	1270	780

When considering a team over or under performing as being outside of their mean +/- one standard deviation, it can be seen that in 2018 the FiveThirtyEight, difference, and combination strategies overperformed, in 2019 and 2021 the combination and difference strategies overperformed, and in 2022 random strategy overperformed while the chalk strategy underperformed. Notably, in three out of four years the difference and combination strategies overperformed, while other strategies were more predictable. Additionally, every strategy other than the random strategy had its lowest point total in 2022.

Discussion

The results of this study provided useful insight into the performance of various strategies for selecting teams in the NCAA Basketball Tournament. The expected performance of each strategy was examined, as well as how much they tend to vary, and their actual performance in a small sample of four years. From this, an approach can be implemented to fill out a bracket in the 2023 Tournament which has been tested and examined for predictive power.

The simplest strategy used was the random strategy, where a team was randomly chosen with equal odds to win each game. Over the four-year period examined, this strategy had the same expected point total every year with 423 points and the same expected standard deviation of 141.5 points. The fact that neither of these estimates varied from year to year was to be expected, as implementing a static strategy would never garner different simulated results if the probabilities powering the simulation never changed. Notably, this strategy had the lowest estimated point total, standard deviation, and simulated accuracy in each round. While this strategy had a poor performance, this is somewhat to be expected, as there is a 1 and 9.2 quintillion chance of selecting a perfect bracket when choosing teams randomly (Wilco, 2023). While it is possible that a random bracket could perform well, the odds of doing so are slim given the nature of this tournament. Examining this strategy's actual performance over the four-year period confirms these findings as it averaged 442.5 points with a standard deviation of 231 points, which was the lowest point total with the second highest standard deviation. Additionally, this strategy never accomplished the goal of outscoring the average user's bracket.

The next strategy tested was the chalk strategy, where the higher-seeded team was chosen to win each game. Over the four-year period examined, this strategy had the same expected point total every year with 846 points and the same expected standard deviation of 234 points. Notably, this strategy had the second highest estimated point total and standard deviation, and was projected to beat the average user two out of four years, both of which were statistically significant. This strategy appeared to perform quite well on the simulated results as it was expected to perform better than multiple more complex approaches, and with a lower standard deviation. The simulated success of this approach can be explained by the theory on which the tournament is built; teams are seeded better because they are supposed to be better teams relative to the rest of the competition, and therefore selecting better seeds teams to always win should result in sound strategy (Mass, 2023). When examining this strategy's actual performance over the four-year period it can be seen that it averaged 825 points, the fifth highest of the strategies tested, with a standard deviation of 149 points, the fifth lowest. This shows that this strategy underperformed relative to the other strategies tested for this range of years used. While this sample size is small and could partially explain this result, it can also be assumed the often unexpected outcomes of this tournament could also explain why choosing the better seed to win every game might not be the best approach (Wilco, 2023).

The next strategy tested was the baseline, the ESPN strategy which took the average user's bracket each year. Over the four-year period examined, this strategy averaged an expected point total of 834.5 points and averaged a standard deviation of 224 points. Notably, this strategy had the third-highest estimated point total and the fifth-lowest standard deviation. This strategy performed well on the simulated results and outperformed multiple more complex approaches, with a lower standard deviation. The simulated success of this approach is similar to that of the chalk strategy, due to the fact that the average user's bracket largely follows a chalk strategy. This showed that picking the average user's bracket is a reasonable strategy for scoring well on the simulated results, however, in practice it will not give a high likelihood of beating other users. When examining this strategy's actual performance over the four-year period it can be seen that it averaged 745 points with a standard deviation of 106 points, which were the second lowest and lowest, respectively, of the strategies tested. This strategy underperformed relative to the other strategies tested, given a small sample size. This result can be partially explained using the same reasoning as the chalk strategy, with an additional consideration that some users may not use sound approaches for picking teams such as picking favorite teams, coolest names, or best mascot, for example.

The next strategy tested was the FiveThirtyEight strategy, where the team with the highest probability to win each game according to FiveThirtyEight was chosen. Over the four-year period examined, this strategy averaged an expected point total of 823 points and averaged a standard deviation of 230 points. Notably, this strategy had the fourth-highest estimated point total, the third-highest standard deviation, and was never predicted to outperform the ESPN strategy. The lack of simulated success when using this strategy is hard to determine without further examination of how FiveThirtyEight construct its probabilistic model, but it can be assumed that this shows the difficulty with trying to build a model to predict a tournament with highly random outcomes at times. When examining this strategy's actual performance over the four-year period it can be seen that it averaged 925 points, the second highest of the strategies tested, with a standard deviation of 159 points, the third lowest. This shows that this strategy overperformed relative to its simulated results, though given a small sample size and highly random outcomes, it is expected that results may vary. This is the best-performing model thus far on the observed data, showing that this strategy may be a reasonable one to choose, though further testing is required given the discrepancy between the simulated and observed results.

The next strategy tested was the KenPom strategy, where a logistic regression model was built to predict the outcomes of games using KenPom ratings. The logistic regression model fit had a test accuracy of 81%, and over the four-year period examined this strategy averaged an expected point total of 801.5 points and averaged a standard deviation of 228 points. Notably, the strategy had the third lowest estimated point total, the fourth highest standard deviation. This strategy appeared to perform worse than a majority of other strategies tested with a lower expected point total and a higher standard deviation, and again showed the difficulty with trying to build a model to predict a tournament with highly random outcomes at times. Additionally, it is possible the logistic regression model built for this study could be improved which may change the simulated results. When examining this strategy's actual performance over the four-year period it can be seen that it averaged 888 points with a standard deviation of 178.5 points, both of which were the fourth highest. This shows that this strategy overperformed relative to its simulated results, explained by a small sample size and highly random outcomes. Given the complexity of this strategy and that it performed worse than the simpler FiveThirtyEight approach in both the simulated and observed results, however, using this strategy may not be the most reasonable approach.
The next strategy tested was the difference strategy, where the largest difference between FiveThirtyEight's predictions and ESPN's percentages determined which

picks were made. Over the four-year period examined, this strategy averaged an expected point total of 557 points and averaged a standard deviation of 234 points. Notably, the strategy had the second-lowest estimated point total, and the second-highest standard deviation. This strategy appeared to perform worse than a majority of other strategies tested with a lower expected point total and showed that the implementation of game theory in this manner may not be the best approach. It is possible that a different calculation of the leverage value which balances possible gain and loss better than the calculation used here which is rather high-risk high-reward could yield a better result. Uniquely however, this strategy yielded the highest accuracy in the national championship game in three out of four years, and correctly predicted the national champion in each of those years. When examining this strategy's actual performance over the four-year period it can be seen that it averaged 900 points, the third highest of the strategies tested, with a standard deviation of 434 points, the highest. This shows that this strategy overperformed relative to its simulated results but with a large standard deviation. While the small sample size could partially explain this result it can also be assumed that the high-risk high-reward approach previously explained accounted for some of this variance. Additionally, the fact that this strategy correctly picked the national champion three out of four years is notable, as this pick alone is worth 320 points in a given year. Furthermore, it should be noted that the simulated point total of this strategy and the observed point total had a correlation of 0.93 over the four years tested, which shows that these totals tend to move together. This strategy may be suitable if a high-risk approach is desired and simulations are bullish, but given the other strategies examined it may not be best.

The final strategy tested was the combination approach, where the picks from the strategy with the highest accuracy in each round were used. Over the four-year period examined, this strategy averaged an expected point total of 861 points and averaged a standard deviation of 238 points. Notably, the strategy had the highest estimated point total and standard deviation and was projected to beat the average user in two out of four years with both being statistically significant. This strategy performed better than all other strategies on the simulated results, but also with the highest variability. Notably, this strategy used chalk, FiveThirtyEight, KenPom, and difference approaches throughout the four years tested. When examining this strategy's actual performance over the four-year period it can be seen that it averaged 1105 points, the highest of the strategies tested, with a standard deviation of 223 points, the third highest. This confirmed that this strategy appeared to be the best performing one, given a small sample size, and it beat the ESPN approach in every year tested. In fact, the simulated and observed point total had a correlation of 0.93 over the four years tested, and if the only year this strategy predicted a point total of less than 860 is omitted then this approach averaged 1213 points with a standard deviation of 73 points. Overwhelmingly, this strategy appears to be the best option for beating the average user's bracket, especially when the simulation is bullish. When considering the first and second research question of which strategy performs the best and whether it is significantly better than the average user's bracket, it would appear that this strategy is the most powerful to outperform the ESPN strategy.

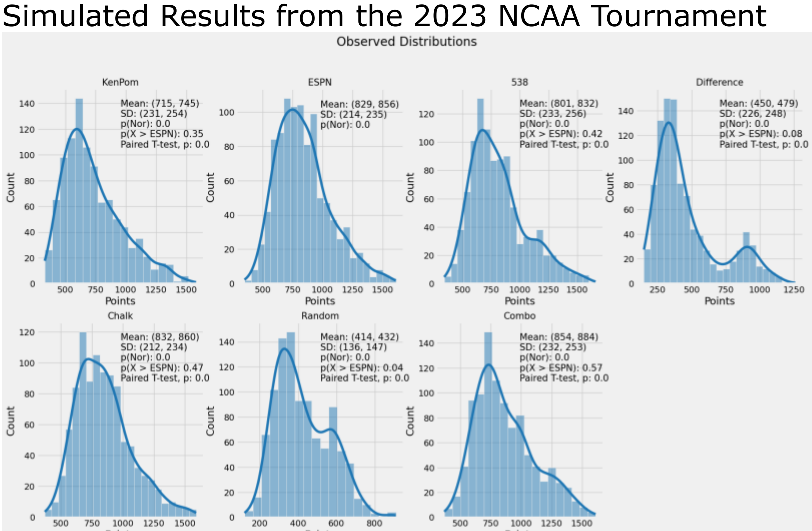
Limitations
Before drawing conclusions regarding the results of this study or implementing its findings for the 2023 NCAA Tournament, the limitations of this study must be addressed to provide context to the findings presented. Most notably, more data is required to fully confirm the performance of each of these strategies with a higher degree of confidence. The limitation of only having four years of ESPN data made it difficult to test all of these approaches for a larger time period, and if it was possible to do so the results of this study could change. Additionally, having a larger set of KenPom ratings could help build a more accurate model to predict win probabilities using logistic regression.

Similarly, the use of more rating metrics and building more sophisticated models for predicting game outcomes could yield better results than the implementation used here. The use of other strategies not tested here could yield different results as well, and more approaches should be examined to find strategies with reasonable performance. In general, the findings of this study should not be considered absolute. Any findings of this study which are divergent from previous literature require the examination of methodologies to understand how different studies were constructed. This study focused on broader findings across multiple approaches which may not be the case for other studies. The findings of this study provided more context into strategies for picking teams in the NCAA Tournament which could benefit future research and knowledge on the topic as a whole, and these results should be viewed as potentially useful information, not definitive findings.

Application & Implementation

For use in creating the optimal bracket in the 2023 NCAA Tournament, the following simulated results were used:

<u>Figure 6</u>
Simulated Results from the 2023 NCAA Tournament

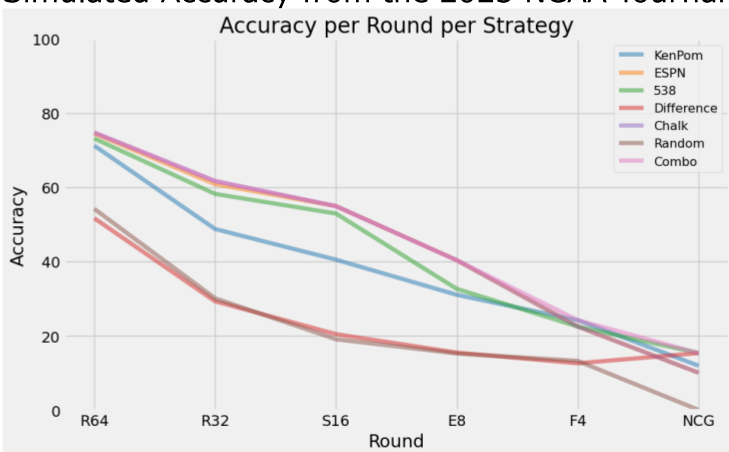


eIt can be seen that the combination strategy had the highest expected point total, with a probability of scoring higher than the average user at 55%. Additionally, the p-value for the Wilcoxon Signed-Rank test for the difference with ESPN was approximately 0, showing that the combination strategy was statistically different from ESPN.

Each strategies accuracy per round can be seen in the figure below:

<u>Figure 7</u>

Simulated Accuracy from the 2023 NCAA Tournament



From this, it can be seen that the KenPom, ESPN, FiveThirtyEight, chalk, and combination strategies all performed quite similarly, with the difference and random strategies performing noticeably worse. The difference strategy's accuracy in the national championship game was the highest at 14.2%, however. Using these accuracies, the combination strategy implemented the following approaches:

R64: Chalk, R32: Chalk, S16: Chalk, E8: Chalk, F4: KenPom, NCG: Difference.

This bracket was submitted to ESPN's tournament challenge, where it had 470 points and finished in the 66th percentile. Additionally, it did beat the average user's bracket for this year as the ESPN bracket finished with 450 points, but was not the highest scoring strategy tested as the FiveThirtyEight approach had 520 points. Notably, this was the worst performance of every strategy used across the four years tested. The full results can be seen below:

<u>Table 2</u>

Observed Results

Strategy	**2023 Points**
KenPom	440
ESPN	450
FiveThirtyEight	520
Difference	460
Chalk	470
Random	310
Combination	470

Conclusion & Recommendations for Future Research

The results of this study showed that a combined strategy for picking teams in the NCAA Basketball Tournament which utilizes the chalk, KenPom, FiveThirtyEight, and difference approaches outlined in this study in any given year is the most powerful approach. Notably, this strategy had the highest expected point total of all the strategies tested as well as the highest average observed point total in the four years examined. This approach did have the second highest standard deviation of the strategies tested for the simulated and observed results, but the possible benefit of using this approach outweighs its possible variance. Most notably, this strategy had a greater than 50% chance of outperforming the average user’s bracket in three of the four years examined and beat the average user’s bracket in all of the actual tournaments for this range of years. When this approach was implemented for the 2023 tournament, it finished in the 66th percentile of users and beat the average user’s bracket.

These findings show that a reasonable approach for filling out brackets can be achieved to score relatively high and beat the average user’s bracket. This approach should be tested on more historical NCAA tournaments to obtain a higher degree of confidence in this strategy’s performance. Additionally, more strategies should be developed and implemented to determine what other approaches are reasonable and perform well. Furthermore, the use of more sophisticated models and other rating metrics with possibly a larger data set should be examined to determine if a more accurate model can built for prediciton. Choosing outcomes of the NCAA Basketball Tournament is a difficult task, but this challenge is what pushes the larger sports and analytics community to continue to push the boundary of what can and cannot be predicted in “March Madness”.

References

Boice, J., & Silver, N. (2018, March 11). *How our march madness predictions work*. FiveThirtyEight. https://fivethirtyeight.com/features/how-our-march-madness-predictions-work/

Dutta, S., Jacobson, S. & Sauppe, J. (2017). Identifying NCAA tournament upsets using Balance Optimization Subset Selection. Journal of Quantitative Analysis in Sports, 13(2), 79-93. https://doi.org/10.1515/jqas-2016-0062

ESPN. (2023). *Whom Picked Whom*. ESPN. https://fantasy.espn.com/tournament-challenge-bracket/2023/en/whopickedwhom

FiveThirtyEight. (2023). *2023 march madness predictions*. FiveThirtyEight. https://projects.fivethirtyeight.com/2023-march-madness-predictions/

Gibbons R. (1992). *Game theory for applied economists*. Archetype Publishing Inc.

Katz, J. (2015, March 15). *Here's how our N.C.A.A. bracket works*. The New York Times. https://www.nytimes.com/2015/03/16/upshot/heres-how-our-ncaa-bracket-works.html

KenPom. (2023). *2023 Pomeroy Ratings*. KenPom. https://kenpom.com/

Leonhart, D. (2015, March 15). *March madness 2015: Welcome to a difference kind of bracket*. The New York Times. https://www.nytimes.com/2015/03/16/upshot/march-madness-welcome-to-a-different-kind-of-bracket.html

Mass, A. J. (2023, March 2). *How to fill out a march madness tournament bracket: all the basics so you can join the madness.* ESPN. https://www.espn.com/fantasy/basketball/story/_/id/26103343/how-fill-tournament-bracket-all-basics-join-madness

Mcubed.net. (2023, January 13). *Men's NCAA basketball tournament records by seed*. Mcubed.net. https://mcubed.net/ncaab/seeds.shtml

Niemi J. B., Carlin B. P., Alexander J. M. (2008). Contrarian strategies for NCAA tournament pools: A cure for march madness? *CHANCE 21*, 35-42. https://doi.org/10.1007/s00144-008-0009-3

Ota, K. (2023, March 16). *ESPN tournament challenge sets new all-time record: 20 million brackets*. ESPN Press Room. https://espnpressroom.com/us/press-releases/2023/03/espn-tournament-challenge-sets-new-all-time-record-20-million-brackets/

Pomeroy, Ken. (2012, June 8). *Ratings Glossary*. KenPom. https://kenpom.com/blog/page/50/

Sport Reference. (2022). *College basketball stats and history.* Sports Reference. https://www.sports-reference.com/cbb/

Wilco, D. (2023, March 15). *What is march madness: The NCAA tournament explained.* NCAA. https://www.ncaa.com/news/basketball-men/bracketiq/2023-03-15/what-march-madness-ncaa-tournament-explained#:~:text=The%20NCAA%20Division%20I%20men's%20basketball%20tournament%20is%20a%20single,only%20four%20teams%20are%20left.

Appendix A

ESPN's "Whom Picked Whom" 2023 Data Set

R64	R32	S16	E8	F4	NCG
1Alabama-96.8%	1Alabama-87.8%	1Alabama-74.8%	1Alabama-53.5%	1Alabama-36.5%	1Alabama-19.5%
1Purdue-96.5%	1Houston-83.5%	1Houston-67.6%	1Houston-45.1%	1Houston-27.7%	1Houston-15.8%
1Kansas-95.9%	2UCLA-82.0%	1Kansas-60.1%	1Kansas-35.4%	1Kansas-19.1%	1Kansas-10.4%
1Houston-95.9%	1Kansas-81.7%	2Texas-54.3%	1Purdue-32.7%	1Purdue-15.4%	1Purdue-8.2%
2UCLA-94.3%	2Arizona-78.9%	2Arizona-49.6%	2Texas-26.5%	2Texas-13.0%	2UCLA-6.3%
...	...	...	...	...	...

Appendix B

Five Thirty Eight's "March Madness Predictions" 2023 Data Set

forecast_date	rd2_win	rd3_win	rd4_win	rd5_win	rd6_win	rd7_win	team_name
2023-03-15	0.971047546794	0.74148923198	0.588248323596	0.407246851028	0.309131289477	0.221185763004	Houston
2023-03-15	0.98578664939	0.819327989866	0.654064522752	0.452827125417	0.304872520223	0.161726456782	Alabama
2023-03-15	0.921391658573	0.64643155298	0.458320103082	0.224545505921	0.141317043999	0.083376089216	Texas
2023-03-15	0.983293160382	0.693803981921	0.406725654127	0.253038024404	0.11974324011	0.052939289831	Purdue
2023-03-15	0.97820868499	0.656920531977	0.389980839021	0.202425028359	0.091173642905	0.049506039049	Kansas

Appendix C

Ken Pomeroy's "Pomeroy Ratings" 2017-2023 Data Set

Name	AdjO	AdjD	AdjT	Luck	SOS OppO	SOS OppD	NCSOS	Year
Gonzaga	120.7	87.7	69.9	0.016	104.9	104.3	0.73	2017
Villanova	122.8	92.0	63.9	0.026000000000000000	110.2	100.2	4.26	2017
North Carolina	122.1	94.1	70.5	0.006	111.7	99.5	4.16	2017
Kentucky	119.2	92.4	72.7	0.009000000000000000	109.2	99.6	3.93	2017
West Virginia	116.5	89.7	69.7	-0.043	109.3	99.5	-7.89	2017
...	...	...	...	...	...	...	...	...

Appendix D
Sports Reference's "Historical Tournament Matchups" 1985-2022 Data Set

Seed	Opp_1	Opp_2	Opp_3	Opp_4	Opp_5	Opp_6	Team	Year
15	Kentucky						Abilene Christian	2019
14	Texas	UCLA					Abilene Christian	2021
11	North Carolina						Air Force	2004
13	Illinois						Air Force	2006
13	Gonzaga						Akron	2009
…	…	…	…	…	…	…	…	…

Appendix E
Mcubed.net's "NCAA Basketball Tournament Records by Seed" Data Set

Seed 1	Seed 2	Seed 1 Win Prob
1	1	0.5
1	2	0.545
1	3	0.634
1	4	0.705
1	5	0.825
…	…	…