**Merkle March Madness Bracket**

William and Mary MSBA

We used data provided from the Kaggle.com March Madness Contest because it went all the way back to 2003 and contained more information. We attached the csv files used in the email with this write up. The first thing we did was summarize the statistics for each team for each season. For example, we had the average points scored, the average offensive rebounds, etc. for Duke in 2003, 2004, 2005, etc. Then we added each team’s win % in each season, as well as their average opponent win % and average opponents opponent win % to account for strength of schedule.

We trained a logistic regression model on tournament games from 2003 to 2017. For each tournament game, we took the average stats of team 1 for that season minus the average stats for team 2 for that season. For example, if team 1 scored an average of 70 points a game and team 2 scored an average of 65 points a game, the observation would have a value of 5 for points scored to indicate that team 1 scored 5 points higher per game on average. The y value for the model was either a 0 or 1 depending on which team won the game. This allowed the model to see which stats were significant in predicting wins as well as the degree to which they were significant (size of the coefficient and p-value).

**The full list of variables included are:**

Points Scored, Points Allowed, Field Goals Made, Field Goal Attempts, 3 Pointers Made, 3 Pointers Attempted, Free Throws Attempted, Offensive Rebounds, Defensive Rebounds, Assists, Turnovers, Steals, Blocks, Personal Fouls, Field Goals Made Allowed, Field Goal Attempts Allowed, 3 Pointers Made Allowed, 3 Pointers Attempted Allowed, Free Throw Attempts Allowed, Offensive Rebounds Allowed, Defensive Rebounds Allowed, Assists Allowed, Turnovers Allowed, Steals Allowed, Blocks Allowed, Personal Fouls Allowed, WinLoss %, Opponent WinLoss %, and Opponents Opponent WinLoss%.

The most significant variable was found to Opponent WinLoss %, followed by WinLoss %, Steals, 3 Pointers Made Allowed, Assists, Field Goals Attempted Allowed, Turnovers Allowed, and Personal Fouls. All of these coefficients had p-values below .05.

After the model was trained, we summarized the season stats for all teams for the 2018 season. Then we created an outline for the bracket and predicted the probability of victory for each team in each matchup (from 0 to 1). We chose the team with the higher probability of victory and moved them on to the next round.

The probabilities for each tournament game are below and can be seen at the end of the script submitted:

|  |  |  |  |
| --- | --- | --- | --- |
| **Team1** | **Team2** | **Probability of Team 1 Win** | **Winner** |
| Villanova | Radford | 0.985666 | Villanova |
| Purdue | CS Fullerton | 0.95285 | Purdue |
| Texas Tech | SF Austin | 0.82942 | Texas Tech |
| Wichita St | Marshall | 0.852546 | Wichita St |
| West Virginia | Murray St | 0.550977 | West Virginia |
| Florida | St Bonaventure | 0.425806 | St Bonaventure |
| Arkansas | Butler | 0.527006 | Arkansas |
| Virginia Tech | Alabama | 0.440819 | Alabama |
| Kansas | Penn | 0.961143 | Kansas |
| Duke | Iona | 0.960185 | Duke |
| Michigan St | Bucknell | 0.838219 | Michigan St |
| Auburn | Col Charleston | 0.803698 | Auburn |
| Clemson | New Mexico St | 0.743215 | Clemson |
| TCU | Syracuse | 0.508157 | TCU |
| Rhode Island | Oklahoma | 0.791816 | Rhode Island |
| Seton Hall | NC State | 0.626074 | Seton Hall |
| Virginia | UMBC | 0.988613 | Virginia |
| Cincinnati | Georgia St | 0.949273 | Cincinnati |
| Tennessee | Wright St | 0.883876 | Tennessee |
| Arizona | Buffalo | 0.664362 | Arizona |
| Kentucky | Davidson | 0.817294 | Kentucky |
| Miami FL | Loyola-Chicago | 0.490394 | Loyola-Chicago |
| Nevada | Texas | 0.761538 | Nevada |
| Creighton | Kansas St | 0.513223 | Creighton |
| Xavier | TX Southern | 0.995708 | Xavier |
| North Carolina | Lipscomb | 0.947758 | North Carolina |
| Michigan | Montana | 0.879805 | Michigan |
| Gonzaga | UNC Greensboro | 0.794707 | Gonzaga |
| Ohio St | S Dakota St | 0.767676 | Ohio St |
| Houston | San Diego St | 0.738322 | Houston |
| Texas A&M | Providence | 0.489019 | Providence |
| Missouri | Florida St | 0.580446 | Missouri |
| Villanova | Alabama | 0.910509 | Villanova |
| Purdue | Arkansas | 0.738841 | Purdue |
| Texas Tech | St Bonaventure | 0.518268 | Texas Tech |
| Wichita St | West Virginia | 0.648285 | Wichita St |
| Kansas | Seton Hall | 0.840875 | Kansas |
| Duke | Rhode Island | 0.678576 | Duke |
| Michigan St | TCU | 0.558904 | Michigan St |
| Auburn | Clemson | 0.492624 | Clemson |
| Virginia | Creighton | 0.943459 | Virginia |
| Cincinnati | Nevada | 0.626401 | Cincinnati |
| Tennessee | Loyola-Chicago | 0.585785 | Tennessee |
| Arizona | Kentucky | 0.439384 | Kentucky |
| Xavier | Missouri | 0.89592 | Xavier |
| North Carolina | Providence | 0.723888 | North Carolina |
| Michigan | Houston | 0.590186 | Michigan |
| Gonzaga | Ohio St | 0.463782 | Ohio St |
| Villanova | Wichita St | 0.821526 | Villanova |
| Purdue | Texas Tech | 0.692886 | Purdue |
| Kansas | Clemson | 0.719887 | Kansas |
| Duke | Michigan St | 0.72544 | Duke |
| Virginia | Kentucky | 0.858743 | Virginia |
| Cincinnati | Tennessee | 0.659225 | Cincinnati |
| Xavier | Ohio St | 0.710304 | Xavier |
| North Carolina | Michigan | 0.598866 | North Carolina |
| Villanova | Purdue | 0.738671 | Villanova |
| Kansas | Duke | 0.574028 | Kansas |
| Virginia | Cincinnati | 0.791858 | Virginia |
| Xavier | North Carolina | 0.625227 | Xavier |
| Villanova | Kansas | 0.601869 | Villanova |
| Virginia | Xavier | 0.744321 | Virginia |
| Villanova | Virginia | 0.360298 | Virginia |

