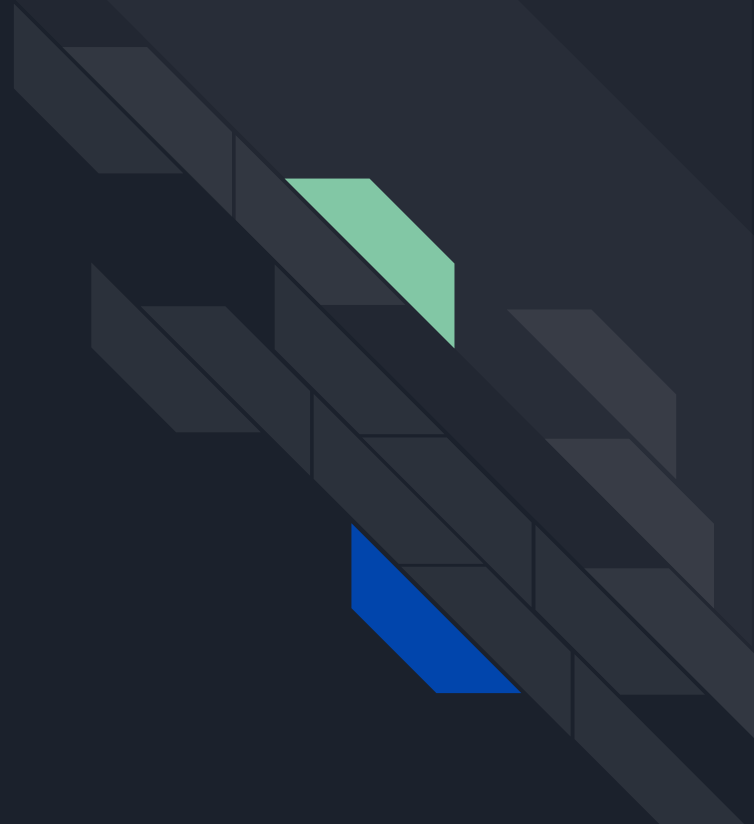


# NCAA Men's Basketball March Madness Predictions with Machine Learning

Yvon Kadeoua



# THE TOURNAMENT



# MARCH MADNESS

The tournament consist of several rounds

- 68 teams
- The first 4(teams with the lowest seeds)
- The first round( the round of 64)
- The second round(The round of 32)
- Sweet 16
- Elite 8
- Final 4
- 4 regions, 16 teams each
- Each team seeded 1-16

2017 Champion: **North Carolina Tar Heels**  
(#1 seed in the South Region)



# March Madness

## 2017



### East

|                           |    |
|---------------------------|----|
| Villanova (31-3)          | 1  |
| Mount St. Mary's (19-15)  | 16 |
| Wisconsin (25-9)          | 8  |
| Virginia Tech (22-10)     | 9  |
| Virginia (22-10)          | 5  |
| UNC Wilmington (29-5)     | 12 |
| Florida (24-8)            | 4  |
| East Tennessee St. (27-7) | 13 |
| SMU (29-4)                | 6  |
| USC (24-9)                | 11 |
| Baylor (25-7)             | 3  |
| New Mexico St. (28-5)     | 14 |
| South Carolina (22-10)    | 7  |
| Marquette (19-12)         | 10 |
| Duke (27-8)               | 2  |
| Troy (22-14)              | 15 |

|                           |    |
|---------------------------|----|
| Gonzaga (32-1)            | 1  |
| South Dakota St. (18-16)  | 16 |
| Northwestern (23-11)      | 8  |
| Vanderbilt (19-15)        | 9  |
| Notre Dame (25-9)         | 5  |
| Princeton (23-6)          | 12 |
| West Virginia (26-8)      | 4  |
| Bucknell (26-8)           | 13 |
| Maryland (24-8)           | 6  |
| Xavier (21-13)            | 11 |
| Florida St. (25-8)        | 3  |
| Florida Gulf Coast (26-7) | 14 |
| Saint Mary's (28-4)       | 7  |
| VCU (26-8)                | 10 |
| Arizona (30-4)            | 2  |
| North Dakota (22-9)       | 15 |

### West



First Four - March 14-15

|                          |    |    |                     |
|--------------------------|----|----|---------------------|
| Mt. Saint Mary's (19-15) | 16 | 16 | NC Central (25-8)   |
| New Orleans (20-11)      | 16 | 16 | UC Davis (22-12)    |
| Providence (20-12)       | 11 | 11 | Kansas St. (20-13)  |
| USC (24-9)               | 11 | 11 | Wake Forest (19-13) |

East

South

### Midwest

|                          |    |
|--------------------------|----|
| Kansas (28-4)            | 1  |
| UC Davis (22-12)         | 16 |
| Miami (21-11)            | 8  |
| Michigan St. (19-14)     | 9  |
| Iowa St. (23-10)         | 5  |
| Nevada (28-6)            | 12 |
| Purdue (25-7)            | 4  |
| Vermont (29-5)           | 13 |
| Creighton (25-9)         | 6  |
| Rhode Island (24-9)      | 11 |
| Oregon (29-5)            | 3  |
| Iona (22-12)             | 14 |
| Michigan (24-11)         | 7  |
| Oklahoma St. (20-12)     | 10 |
| Louisville (24-8)        | 2  |
| Jacksonville St. (20-14) | 15 |

|                           |    |
|---------------------------|----|
| North Carolina (27-7)     | 1  |
| Texas Southern (23-11)    | 16 |
| Arkansas (25-9)           | 8  |
| Seton Hall (21-11)        | 9  |
| Minnesota (24-9)          | 5  |
| Mid. Tennessee (30-4)     | 12 |
| Butler (23-8)             | 4  |
| Winthrop (26-6)           | 13 |
| Cincinnati (29-4)         | 6  |
| Kansas St. (20-13)        | 11 |
| UCLA (29-4)               | 3  |
| Kent St. (22-13)          | 14 |
| Dayton (24-7)             | 7  |
| Wichita St. (30-4)        | 10 |
| Kentucky (29-5)           | 2  |
| Northern Kentucky (24-10) | 15 |

### South

THE BRACKET

- 19 millions brackets were submitted to ESPN
- 9.2 quintillion possible brackets
- The majority of fans submit 2 different brackets
- Nobody has ever submitted a perfect bracket

- The approximate retail value of all prizes on ESPN is \$22,950
- Warren Buffett offers \$1 million for life for a perfect Sweet 16 field



# \$1 Million

## Perfect Bracket Challenge

Finally a bracket contest you can't lose

| PLACE                | PRIZE             |
|----------------------|-------------------|
| PERFECT BRACKET      | \$1,000,000       |
| 1st                  | \$10,000 CASH     |
| 2nd                  | \$5,000 CASH      |
| 3rd                  | \$2,500 CASH      |
| 4th - 100th          | 15,000 BetPoints™ |
| 101st - 150th        | 7,500 BetPoints™  |
| 151st - 300th        | 3,500 BetPoints™  |
| 301st - 500th        | 1,000 BetPoints™  |
| 501st - 1,000th      | 500 BetPoints™    |
| 1,001st - 2,500th    | 100 BetPoints™    |
| 2,501st - 5,000th    | 75 BetPoints™     |
| 5,001st - Last Place | 50 BetPoints™     |



# Overview of March Madness Project And Questions of Interest

- Can we use NCAA historical data and Machine Learning to predict the games outcome?
- Based on the probabilities can we predict what team will move forward?
- How accurate can these statistical results be compared to the brackets?



# PREVIOUS WORK

- Most previous work have been developed in the context of Statistical Modeling rather than Machine learning
- Existing work which employs machine learning suffered from poor feature selection
- Models built based on Efficiency metrics and Las Vegas Point Spreads



## HOW I PLAN TO FIX THIS:

- Compute moving averages and the “4 factors”
- Analyze a wide range of feature selection methods
- Pick relevant features
- Analyze a broad range of Machine Learning techniques
- Select the model with the best accuracy scores





## THE DATA

- field goals made
- field goals attempted
- three pointers made
- three pointers attempted
- free throws made
- free throws attempted
- offensive rebounds
- defensive rebounds
- Assists
- turnovers committed
- Steals
- Blocks
- personal fouls committed
- offensive and defensive efficiencies
- Effective field goal percentage
- Turnover percentage
- Offensive Rebound Percentage
- Free throw rate



## SELECTED FEATURES

- Defensive efficiencies
- Effective field goal percentage
- Free throw rate
- free throws attempted
- defensive rebounds
- turnovers committed
- Steals
- personal fouls committed



# MODELS

- Adaptive Boosting
- Gradient Boosting
- Neural Network
- Support Vector Machine
- Random Forest



# LEARNING TECHNIQUES

Trained model for classification for picking loser and winner for each matchup and determining the likelihood for each outcome

# NET RANKINGS

NCAA EVALUATION TOOL

- 1  VIRGINIA
- 2  DUKE
- 3  MICHIGAN
- 4  TEXAS TECH
- 5  TENNESSEE
- 6  KANSAS
- 7  GONZAGA
- 8  NEVADA
- 9  MICHIGAN STATE
- 10  WISCONSIN



## METHOD OF EVALUATION

- Accuracy scores
- Create Brackets
- Bracket scores

# ACCURACY SCORE

| Model               | Acc  | AUC  |
|---------------------|------|------|
| Random Forest       | 0.87 | 0.94 |
| Neural Network      | 0.87 | 0.93 |
| Gradient Boosting   | 0.87 | 0.93 |
| Adaptative Boosting | 0.85 | 0.92 |
| SVM                 | 0.79 | 0.82 |

## BRACKET SCORES

|                     |     |
|---------------------|-----|
| Neural Network      | 270 |
| Gradient Boosting   | 230 |
| Adaptative Boosting | 210 |
| Random Forest       | 200 |
| SVM                 | 200 |





# k Bracket

| First Round                      | Second Round                 | Sweet 16                     | Elite Eight | Final Four                     | NATIONAL CHAMPIONSHIP | Final Four                       | Elite Eight                      | Sweet 16                         | Second Round                   | First Round                   |
|----------------------------------|------------------------------|------------------------------|-------------|--------------------------------|-----------------------|----------------------------------|----------------------------------|----------------------------------|--------------------------------|-------------------------------|
| <b>Y01 Kansas</b>                |                              |                              |             | <b>Z11b Wake Forest</b>        |                       |                                  |                                  |                                  |                                | <b>W01 Villanova</b>          |
| <b>Y16b UC Davis 53.42%</b>      | <b>Y16b UC Davis</b>         |                              |             | <b>Z11a Kansas St 75.45%</b>   | OPENING ROUND GAME    |                                  |                                  |                                  | <b>W16b New Orleans 78.07%</b> | <b>W16b New Orleans 90.1</b>  |
| <b>Y08 Miami FL 76.12%</b>       |                              | <b>Y08 Miami FL 66.37%</b>   |             |                                |                       |                                  |                                  | <b>W16b New Orleans</b>          | <b>W08 Wisconsin</b>           | <b>W08 Wisconsin</b>          |
| <b>Y09 Michigan St</b>           | <b>Y08 Miami FL 57.83%</b>   |                              |             |                                |                       |                                  |                                  |                                  | <b>W09 Virginia Tech</b>       | <b>W09 Virginia Tech 82.</b>  |
| <b>Y04 Purdue 95.00%</b>         |                              | <b>Y08 Miami FL</b>          |             |                                |                       |                                  | <b>W12 UNC Wilmington</b>        |                                  | <b>W04 Florida</b>             | <b>W04 Florida</b>            |
| <b>Y13 Vermont</b>               | <b>Y04 Purdue 89.83%</b>     |                              |             |                                |                       |                                  |                                  | <b>W13 ETSU</b>                  | <b>W13 ETSU 93.98%</b>         | <b>W13 ETSU 93.98%</b>        |
| <b>Y05 Iowa St</b>               | <b>Y04 Purdue</b>            |                              |             |                                |                       |                                  | <b>W12 UNC Wilmington 71.88%</b> |                                  | <b>W05 Virginia</b>            | <b>W05 Virginia</b>           |
| <b>Y12 Nevada 83.56%</b>         | <b>Y12 Nevada</b>            |                              |             |                                |                       |                                  |                                  | <b>W12 UNC Wilmington 81.00%</b> | <b>W12 UNC Wilmington 95</b>   | <b>W12 UNC Wilmington 95</b>  |
| <b>Y06 Creighton 69.73%</b>      |                              | <b>X02 Louisville</b>        |             |                                |                       | <b>W07 South Carolina</b>        |                                  | <b>W06 SMU</b>                   | <b>W06 SMU</b>                 | <b>W06 SMU</b>                |
| <b>Y11 Rhode Island</b>          | <b>Y06 Creighton 73.98%</b>  |                              |             | <b>X15 North Dakota</b>        |                       |                                  |                                  | <b>W11a Providence</b>           | <b>W11a Providence 55.53%</b>  | <b>W11a Providence 55.53%</b> |
| <b>Y03 Oregon 59.44%</b>         | <b>Y06 Creighton</b>         |                              |             |                                |                       |                                  |                                  | <b>W03 Baylor</b>                | <b>W03 Baylor 95.00%</b>       | <b>W03 Baylor 95.00%</b>      |
| <b>Y14 Iona</b>                  | <b>Y03 Oregon</b>            |                              |             |                                |                       |                                  |                                  | <b>W03 Baylor 65.09%</b>         | <b>W14 New Mexico St</b>       | <b>W14 New Mexico St</b>      |
| <b>Y07 Michigan</b>              |                              | <b>Y02 Louisville 95.00%</b> |             |                                |                       | <b>W07 South Carolina 55.06%</b> |                                  | <b>W07 South Carolina 87</b>     | <b>W07 South Carolina 87</b>   | <b>W07 South Carolina 87</b>  |
| <b>Y10 Oklahoma St 56.55%</b>    | <b>Y10 Oklahoma St</b>       |                              |             |                                |                       |                                  |                                  | <b>W10 Marquette</b>             | <b>W10 Marquette</b>           | <b>W10 Marquette</b>          |
| <b>Y02 Louisville 91.26%</b>     | <b>Y02 Louisville 80.20%</b> |                              |             |                                |                       | <b>W07 South Carolina 67.21%</b> |                                  | <b>W02 Duke</b>                  | <b>W02 Duke</b>                | <b>W02 Duke</b>               |
| <b>Y15 Jacksonville St</b>       | <b>Y02 Louisville 76.98%</b> |                              |             |                                |                       |                                  |                                  | <b>W15 Troy</b>                  | <b>W15 Troy 95.00%</b>         | <b>W15 Troy 95.00%</b>        |
| <b>Y16b UC Davis 80.90%</b>      |                              |                              |             |                                |                       |                                  |                                  |                                  |                                |                               |
| <b>Y16a NC Central</b>           | OPENING ROUND GAME           |                              |             | <b>Z14 Kent</b>                |                       |                                  |                                  | <b>W11b USC</b>                  |                                |                               |
|                                  |                              |                              |             | <b>CHAMPIONS</b>               |                       |                                  |                                  | <b>W11a Providence 92.19%</b>    | OPENING ROUND GAME             |                               |
| <b>Z01 North Carolina 55.13%</b> |                              |                              |             |                                |                       |                                  |                                  |                                  |                                | <b>X01 Gonzaga</b>            |
| <b>Z16 TX Southern</b>           | <b>Z01 North Carolina</b>    |                              |             |                                |                       |                                  |                                  | <b>X16 S Dakota St 95.00%</b>    | <b>X16 S Dakota St 69.02%</b>  | <b>X16 S Dakota St 69.02%</b> |
| <b>Z08 Arkansas</b>              | <b>Z09 Seton Hall</b>        |                              |             |                                |                       |                                  |                                  | <b>X08 Northwestern</b>          | <b>X08 Northwestern 72.3</b>   | <b>X08 Northwestern 72.3</b>  |
| <b>Z09 Seton Hall 63.69%</b>     | <b>Z09 Seton Hall 83.80%</b> |                              |             |                                |                       |                                  |                                  | <b>X09 Vanderbilt</b>            | <b>X09 Vanderbilt</b>          | <b>X09 Vanderbilt</b>         |
| <b>Z04 Butler</b>                |                              | <b>Z12 MTSU</b>              |             |                                |                       |                                  |                                  | <b>X04 West Virginia 95</b>      | <b>X04 West Virginia 95</b>    | <b>X04 West Virginia 95</b>   |
| <b>Z13 Winthrop 50.37%</b>       | <b>Z13 Winthrop</b>          |                              |             | <b>Z14 Kent 89.04%</b>         |                       |                                  |                                  | <b>X13 Bucknell</b>              | <b>X13 Bucknell</b>            | <b>X13 Bucknell</b>           |
| <b>Z05 Minnesota</b>             | <b>Z12 MTSU 70.04%</b>       |                              |             |                                |                       |                                  |                                  | <b>X05 Notre Dame</b>            | <b>X05 Notre Dame</b>          | <b>X05 Notre Dame</b>         |
| <b>Z12 MTSU 59.90%</b>           | <b>Z12 MTSU 76.16%</b>       |                              |             |                                |                       |                                  |                                  | <b>X12 Princeton 75.49%</b>      | <b>X12 Princeton 75.49%</b>    | <b>X12 Princeton 75.49%</b>   |
| <b>Z06 Cincinnati</b>            |                              | <b>Z14 Kent 90.66%</b>       |             |                                |                       | <b>X15 North Dakota 82.38%</b>   |                                  | <b>X06 Maryland</b>              | <b>X06 Maryland 95.00%</b>     | <b>X06 Maryland 95.00%</b>    |
| <b>Z11a Kansas St 95.00%</b>     | <b>Z11a Kansas St</b>        |                              |             |                                |                       |                                  |                                  | <b>X11 Xavier</b>                | <b>X11 Xavier</b>              | <b>X11 Xavier</b>             |
| <b>Z03 UCLA</b>                  | <b>Z14 Kent 95.00%</b>       |                              |             |                                |                       |                                  |                                  | <b>X14 FL Gulf Coast</b>         | <b>X03 Florida St</b>          | <b>X03 Florida St</b>         |
| <b>Z14 Kent 95.00%</b>           | <b>Z14 Kent 89.63%</b>       |                              |             |                                |                       |                                  |                                  | <b>X14 FL Gulf Coast 51.43%</b>  | <b>X14 FL Gulf Coast 89</b>    | <b>X14 FL Gulf Coast 89</b>   |
| <b>Z07 Dayton 95.00%</b>         |                              | <b>Z14 Kent 90.38%</b>       |             |                                |                       |                                  |                                  | <b>X15 North Dakota 62.56%</b>   | <b>X07 St Mary's CA</b>        | <b>X07 St Mary's CA</b>       |
| <b>Z10 Wichita St</b>            | <b>Z07 Dayton 92.18%</b>     |                              |             |                                |                       |                                  |                                  |                                  | <b>X10 VA Commonwealth 7</b>   | <b>X10 VA Commonwealth 7</b>  |
| <b>Z02 Kentucky 56.37%</b>       | <b>Z07 Dayton</b>            |                              |             |                                |                       |                                  |                                  | <b>X15 North Dakota 75.59%</b>   | <b>X02 Arizona</b>             | <b>X02 Arizona</b>            |
| <b>Z15 N Kentucky</b>            | <b>Z02 Kentucky</b>          |                              |             |                                |                       |                                  |                                  | <b>X15 North Dakota 90.18%</b>   | <b>X15 North Dakota 95.0</b>   | <b>X15 North Dakota 95.0</b>  |
|                                  |                              |                              |             | <b>W16b New Orleans 85.88%</b> |                       |                                  |                                  |                                  |                                |                               |
|                                  |                              |                              |             | <b>W16a Mt St Mary's</b>       | OPENING ROUND GAME    |                                  |                                  |                                  |                                |                               |



# Conclusion

- Can we use NCAA historical data to predict the games outcome?
  - From my analyses, we can still attempt to predict the outcome
- Can we predict what team will move forward?
  - Yes, we can predict the likelihood that a team will move forward
- How accurate can these results be used to create brackets?
  - These results can be used to create brackets. But, as we all know, March Madness is essentially madness, and we can never be confident in our selections



# Obstacles and Future Work

- Poor model performance due to poor feature selection
- Other successful work have used rolling averages
- Time constraint



# Reference

- Andrew Lavandoski and Jonathan Lobo, “Predicting the Men’s Basketball Tournament with Machine Learning”, CS 2750: Machine Learning, April 2017
- Z. Shi, S. Moorthy, A. Zimmermann, “Predicting NCAAB match outcomes using ML techniques – some results and lessons learned”, October 2017
- Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction., 2nd edition, Springer, 2009