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)

THE BRACKET


First Round March 16-17
Second Round March 18-19
Sweet 16 March 23-24
Elite 8 March 25-26
Final Four April 1
Championship April 3
Final Four April 1
Elite 8 March 25-26
Sweet 16 March 23-24
Second Round March 18-19
First Round March 16-17
March Madness 2017
FINAL FOUR PHOENIX 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
Villanova
Wisconsin
Virginia
Florida
USC
Baylor
South Carolina
Duke
Wisconsin
Florida
Baylor
South Carolina
Florida
South Carolina
South Carolina
West
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
Gonzaga
Northwestern
Notre Dame
West Virginia
Xavier
Florida St.
Saint Mary's
Arizona
Gonzaga
West Virginia
Xavier
Arizona
Gonzaga
Xavier
Gonzaga
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
Kansas
Michigan St.
Iowa St.
Purdue
Rhode Island
Oregon
Michigan
Louisville
Kansas
Purdue
Oregon
Michigan
Kansas
Oregon
Oregon
South
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
North Carolina
Arkansas
Mid. Tennessee
Butler
Cincinnati
UCLA
Wichita St.
Kentucky
North Carolina
Butler
UCLA
Kentucky
North Carolina
Kentucky
North Carolina
Gonzaga
North Carolina
North Carolina
Champions
First Four - March 14-15
Mt. Saint Mary's (19-15) 16
New Orleans (20-11) 16
East
Providence (20-12) 11
USC (24-9) 11
East
16 NC Central (25-8)
16 UC Davis (22-12)
Midwest
11 Kansas St. (20-13)
11 Wake Forest (19-13)
South

- 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

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

- Adaptative 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

METHOD OF EVALUATION

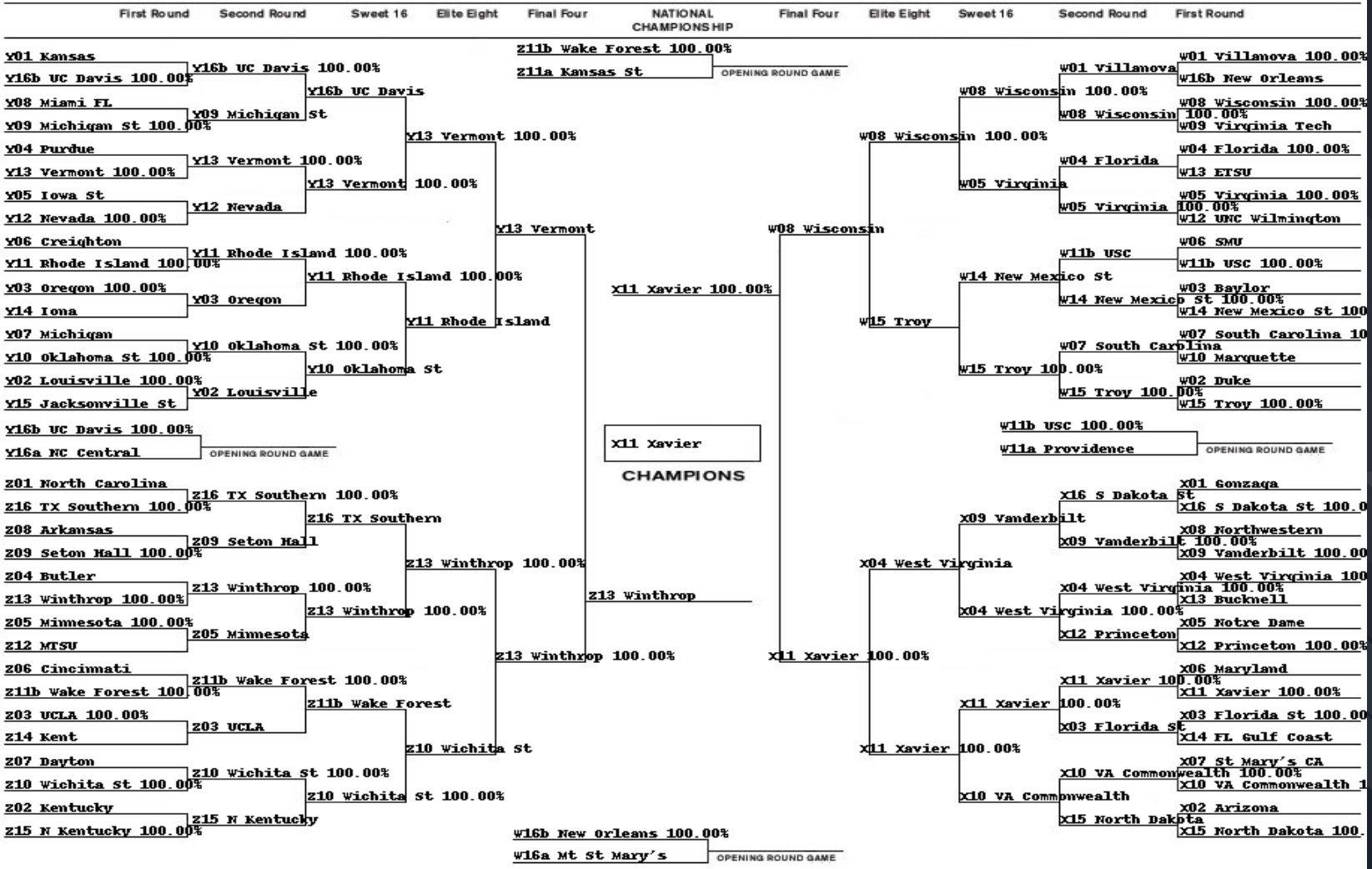
- Accuracy scores
- Create Brackets
- Bracket scores

ACCURACY SCORE

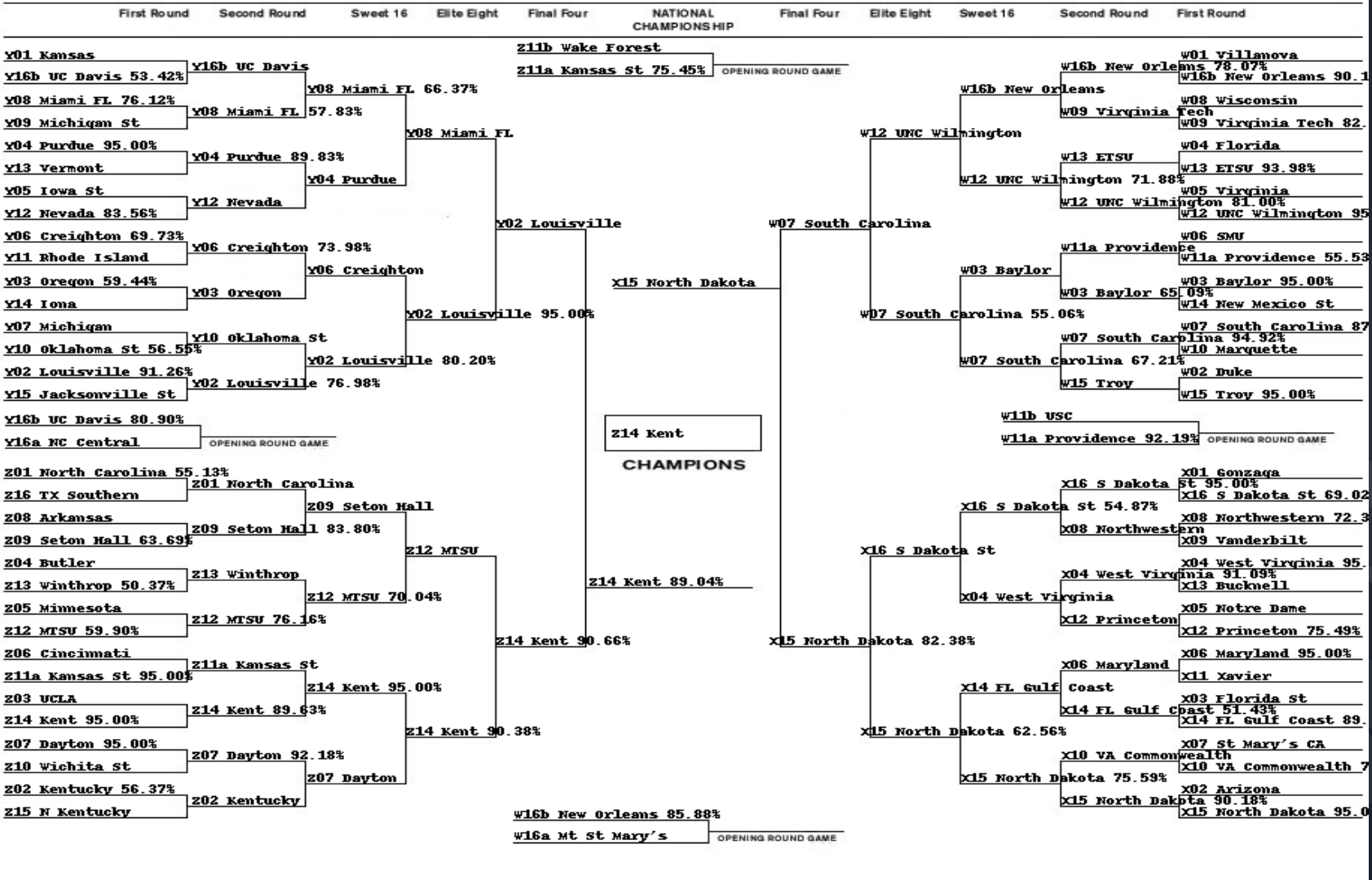
Model	Acc	AUC
Random Forest	0.87	0.94
Neural Network	0.87	0.93
Gradient Boosting	0.87	0.93
Adaptative Boosting	0.85	0.92
SVM	0.79	0.82

BRACKET SCORES

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



Random Forest Bracket



Neural Network Bracket

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