Full Report

1. March Madness
2. Juan-David Dominguez & Michael Gofron
3. [juandominguez2017@u.northwestern.edu](mailto:juandominguez2017@u.northwestern.edu) & michaelgofron2017@u.northwestern.edu
4. EECS 349 Machine Learning, Northwestern University
5. i. We were motivated to find a way to generate a perfect bracket or a better bracket than average based on the idea that it is a mathematically interesting task and economically advantageous. It is mathematically interesting because the predicted odds of getting a perfect bracket are 1 in 1,610,543,269 (for the year 2015). Additionally, it is economically advantageous because we can make money by betting with our friends and others on who has a better bracket.

ii. We used a Decision Tree from the Scikit library. We took in many features about a team’s stats and the regular season game stats and took the differential of these features so we ended with point differential, rebound differential, FG percentage, and other features. We took in 120 features into our decision tree including as seen in Appendix A. We experimented with using the Random Forest decision tree classifier but we couldn’t get consistent results. We also manipulated the data and changed the features to be per game for all statistics to prevent skew from the number of games a team could play in the whole season. Our data set consisted of

iii. We trained our decision tree with the team stats over the season\* as our features and whether a team won a game as the output space and restricting our training set to the regular season. Then we measured our success originally based on the percentage of games that the algorithm guessed correctly.

iv. We found that point differential is the most important feature in our decision tree. Our solution performed well in our first analysis of how many games it was able to correctly predict with 78% accuracy.

**Suggestions for Future Work**

We would like to try using more algorithms to see the validity of other algorithms and their impact. We also want to stack algorithms on top of one other in order to produce a more effective general algorithm for the future. Additionally, there should be more work done in separating the dataset between the playoffs and the regular season for the team statistics to ensure statistics are not skewed based on the number of games a team has played in a season. We would like to make the bracket making process more automatic.