Machine Learning with March Madness

by: Brett Thompson, Grant Wooldridge, Kris Wasemiller & Scott Frazier

The ides of March Madness.

Underdogs vs Giants, last second buzzer beaters, and legends that form from each tournament. And with every tournament, comes the tournament bracket challenge! No one has ever created the perfect bracket. So we took on the challenge to create a machine learning model to help us better predict the outcome of each matchup for the men's tournament.

We discovered a large dataset composed of college basketball teams from 1985 - 2021. For a more digestible application we only used data for 2016 - 2021. Additionally, in our research we discovered the Ken Pomeroy aka "KenPom" data which includes current and past season ratings and statistics.

We joined our regular season data and KenPom data together centered on the team names. Our predictive model was centered on how many points any given team will score and how many they will allow. This effort in joining gives us our outcome with some surprising accuracy.

Results:

(i) Who will win the 2021 NCAA Tournament and how many points will they Score?

- Iowa wins against Illinois: 80 - 76

(ii) Who would have won the 2020 NCAA Tournament had it not been canceled due to COVID 19?

- Michigan State wins against BYU: 76 - 74

(iii) Model Results

- The model accurately predicted 23 out of 32 (72%) first round matchups correctly for the 2021 NCAA Tournament
- Since the model was finalized after the first and second round of the tournament had already been completed, we entered the future predictions into ESPN's 2nd chance tournament challenge. The 2nd chance tournament challenge includes the 16 remaining teams in the NCAA tournament and their matchups. As of 3/28, the model's "predicted" brackets sits in the 91th percentile and has correctly predicted all but 2 outcomes.

How can we use this predictive model?

There are various yearly competitions in the coding and Machine Learning community. Some awarding prize money for the most accurate model. We could donate these winnings to a charity in the company's or your name. This would generate positive media feedback on our company's charitable efforts. Of course at the very least we could guarantee you a high rate of success every year for the company wide bracket challenge. We look forward to discussing with you other options to utilize Machine Learning as a pipeline for this company's success.