**NCAA Bracket Predictor**

# Sprint 2

Team Members:

Kevin Brosam, Nate Lang, John Hattas, Alex Berkhout, Matt Petter

**User Stories:**

Nate is looking for a way to compare multiple tournaments at a time and find a way to predict the winner based on previous tournaments. He wants to select several indicators and have the program find the ideal combination for the best possible bracket. He also wants to be able to use advanced statistics such as SRS or RPI.

Alex is interested in seeing how different indicators affect the predicted bracket. He wants to be able to select many different combinations of indicators and weights to vary the output bracket, so he can compare different possibilities. He would also like to be able to compare these outputted brackets against past seasons to get a general idea of how accurate it might be. Most importantly, he needs many different options for statistics that he would be able to use as indicators.

Charles is a fan of NBA basketball, but is not very familiar with college basketball. He understands how brackets work, so all he needs to know is how each team in a matchup compares to each other. The website gives him all the important metrics for evaluating how good each team is, which Charles really appreciates. Since he is familiar with how basketball works, he also wants to see each team’s strength of schedule, their record, and how well they did in previous years of the tournament, which the website also provides.

Kevin knows a fair amount about college basketball and compares stats regularly. He knows that there are other factors which affect games that do not appear on the stat sheet. Some of these include injuries, location of game, and recent play. He wants a way to easily compare these new stats with each team without combing through data for hours.

**Task Cards:**

Get Kaggle data and adapt code to use it

Create SRS, SOS, RPI, methods

Refine code and optimize for better run time

Implement a basic machine learning algorithm

Run basic machine learning algorithm for multiple years

Create an extra stats column

Have multiple working indicators

Have a weighting system for indicators

Brain storm about “intangible stats”

**Sprint Backlog:**

|  |  |  |
| --- | --- | --- |
| Task | Priority [1-10 (1 being lowest)] | Completed(Y/N) |
| Create extra stats column | 7 | N |
| Weighting system for indicators | 5 | N |
| Basic Machine Learning algorithm | 4 | N |
| Multiple working indicators | 10 | Y |
| Create SRS, SOS, RPI methods | 7 | Y (Partly) |
| Optimize code | 7 | Y |
| Import Kaggle Data, adapt code | 10 | Y |
| Run against multiple years | 3 | Y |
| Brain storm about “intangible stats” | 4 | N |

**Product Backlog:**

|  |  |  |
| --- | --- | --- |
| Task | Priority [1-10 (1 being lowest)] | Completed(Y/N) |
| Develop an algorithm that predicts previous tournaments results | 1 | N |
| Integrate more advanced statistics | 7 | Y |
| Create picture of the bracket with appropriate teams | 7 | N |
| Collect Data | 10 | Y |
| Have a basic working model | 10 | Y |
| Back test for better prediction results | 4 | N |
| Potentially display through HTML | 1 | N |
| Update for 2018 tournament | 3 | Y |
| Compare different basic algorithms to find the easiest while not losing accuracy | 8 | N |
| Display data in charts and tables, potentially using R | 5 | N |
| Add location as one of the indicator | 2 (if reasonably possible) | N |
| Create User Interface | 4 | N |

## Sprint Retrospective

Overall this Sprint was a success and we made good progress on the project. It may not look like we did because the backlog has uncompleted tasks but the main parts of the project we needed to accomplish were completed. The main inhibiter to this sprint was Spring Break. Very little was done during this time, everyone was on vacation. The main portion of the progress this sprint was converting the Kaggle data to our project. Essentially, we are now using a whole new data set which so far has been more accurate. The biggest leap with this new data set is that there was a large portion of missing data depending on if a team went winless or undefeated. This stemmed from the fact that there were no winning or winning opponent stats, so we had to devise a method to produce these stats accurately.

On the agile side of things, we were more consistent then Sprint 1 on our daily scrums. They were not “daily” but they occurred more often then they did for Sprint 1. We paired programmed more and we adapted to what we needed to do for the project. When we switched data sets we were able to reuse most of our code which saved us time and we spent a fair amount of time this Sprint optimizing it. One of the biggest inhibitions to our project is the amount of time each back test takes, and by splitting up the code into chunks and trying to reduce some of the computational time we are trying to limit the amount of time each back test takes.

The results of this Sprint are technically the same as last, we do not have a perfect bracket but, we are getting closer to getting an algorithm which could theoretically work. The new data set helps because our last data set had improper data in it which artificially gave us better results. Going forward as a group we can make our daily scrums daily, paired program again, spend time refining the code to make it quicker to run and simply spend time back testing it for better results. We are optimistic for the next sprint because this years NCAA bracket will be complete.