

Project Proposal

My proposal for a project is to use NCAA basketball stats to predict how far a team will make it during March Madness. It makes sense that a team's stats would correlate with how well they will do in the tournament. The NCAA keeps the offensive and defensive stats of (at least) the top 50 teams available online. I would use a team's stats up until the tournament to predict their success in the tournament in the same year. I would use the stats for the top 50 teams from past years. I would make the output classes linear values from 0 to 7. Zero means that the team didn't make it into the tournament, one means they were eliminated in the first round, and so on to seven. A seven would mean that the team won the tournament.

I would use the following features in the project: win-loss percentage, turnovers per game, total rebounds per game, three-point FGs per game, FT percentage, assists per game, blocked shots per game, steals per game, scoring defense, and scoring offense. All of these are real, continuous values. These data would be gathered from the NCAA's basketball website. It has stats for the top 50 teams all the way back to 2000, although I am not totally sure how complete the data are. Labeling might take some effort, since we would have to manually look up the NCAA bracket for each year, find out which teams won/lost in which rounds, and add those labels to the dataset.

A sample data instance follows:

Team: Wisconsin

Win-Loss Percentage	Turnovers/Game	Total Rebounds/Game	3pt FG/Game	FT%
91.2	7.4	38.1	9.3	76.3

Assists/Game	Blocked shots/Game	Steals/Game	Scoring Defense (PPG)	Scoring Offense (PPG)	Label
15.1	4.8	7.8	56.1	79.1	6

I am interested in being a part of this project, if it is selected.