

College Coaching Legends

Any sports ranking is going to be contentious. Players, coaches, and teams are often judged subjectively, not so much by how many games they win so much as how inspiring their story is. When we were given the challenge of deciding the greatest college coaches of all time, we knew that we'd have to be as objective as possible. Rather than choose any one way of picking the greatest coaches, we started with simple models and then extended to models that captured win-loss records between coaches, which we then compared to existing opinion polls and awards.

1 Simple Models

We considered how well the most common metrics for rating coaches work in this situation. Win ratio (wins over total games) is often used. We found that this doesn't work because it will put a coach with a 1-0 record above one with a 200-20 record. The reason why win ratio works when ranking good coaches is usually because every coach surveyed has played a similar number of games. We could have filtered our data (only consider coaches with a bowl win, etc.), but we wanted to remove arbitrary factors from our process. Another metric we tested was the win-loss difference. We found that this simply favored coaches with long careers.

To make up for the flaws of win ratio and net wins, we gave more weight to coaches that had a high win ratio but also have won more games by multiply win ratio by wins, which when we tested produced great results without any filtering. This effectively dismisses coaches who have played too few games, but doesn't allow one coach to dominate simply because of a high number of games.

2 Connecting the Dots

One problem with this model is that it doesn't consider the connections between coaches. Surely career stats don't matter when one coach consistently beats another. We explored the interactions between coaches with what is known as a graph model. The idea is like this: I get points from every coach that I beat, and I evenly distribute those points to coaches who beat me. We give every coach the same amount of points to start, and see where they stabilize. This adds transitivity to our model: if I beat another coach and he beats 100 other coaches, then I get a good share of all those points. This also has the effect of eliminating small outliers. A coach with a 15-0 record gets a lower score than one with a 200-30 record simply because the better coach has a better inflow of points.

For Basketball, we collected score data from every NCAA tournament game, and for Football we collected the results of every postseason bowl game. The original plan was to collect data on every regular season game ever as well, but that would have violated the data use policies of our sources. We were also unable to find any more data than tournament final scores for Baseball (the data exists but not in a usable format), and only considering two coaches every year would not have been enough data.

With the data that we could get, we did some math relating the score differences and game importance. For instance, a first-round game is not as important as a Final Four game, and an easy victory weighs more heavily than a close win. We used this data to say whether one coach beats another over the course of their games together. With this, we used a variant of Google’s Pagerank algorithm (originally designed to rank website results) to rank coaches. We call this system CoachRank.

3 Results

Based on our models, our top five coaches in football, basketball, and baseball are:

Rank	Football	Basketball	Baseball
1	Joe Paterno	Mike Krzyzewski	Ed Cheff
2	Mack Brown	Dean Smith	Gene Stephenson
3	Bear Bryant	Roy Williams	Mike Martin
4	Lloyd Carr	John Wooden	Augie Garrido
5	Pete Carroll	Rick Pitino	Gordie Gillespie

4 Best Coach Attributes

Okay cool, so we have this model that spits out the top 5 coaches for a sport by looking at interactions between the coaches. Wouldn’t it be interesting to find out what attributes of a coach made them show up in the top 5? It’s really difficult to tell this from the graphical model because there’s a lot of thousands of coach interactions going on. So to combat this, we took the CoachRank values that the graph model spit out for each coach and tried to put a line through the points based on attributes we thought might be important, such as when they started coaching, how many games they won, how many times they ended up winning the championship, etc. Once we ran our new model, we found that the attributes that gave the coach a high rank were number of times they appeared in the final round of a sport and how many games they had won. This also confirmed our graph model because although we didn’t explicitly use those attributes in making the model, they still showed up as indicators of a good coach.

We believe these models have a minimum amount of subjectivity, which sets them apart from many other rankings and opinion polls. These models make the best use of the data available to us, but that doesn’t mean they are infallible. The final say on who’s the greatest coach can only truly be tested on the field, and we think we do well in capturing that.