

To just dive right in, let me begin with the high level approach I took to solving this problem and further on in the document I'll elaborate a bit on some of the items:

1. Create a skeleton. Although this changed a bit over the course of the solution, the basic skeleton for the simulation remained unchanged.
2. Gather data. I felt that more than just the international results from the past four years was necessary for an adequate solution so the first step was to search for and process additional data.
3. Assign base scores to teams. These base scores represent a team's strength independent of the opponent they're facing.
4. Assign weighted scores to teams. The weighted scores represent a team's base score properly adjusted to the opponent they are facing. Therefore, a given team has a distinct weighted score against every other team in the tournament.
5. Assign winning probabilities to teams based on their weighted scores and some other fields. Using the weighted scores from (5) as well as some other fields of a team, each team is given a distinct probability of winning against every other team in the tournament. For instance, the USA might have a winning probability of .5434 against Ghana, .4823 against Brazil, etc.
6. Using these winning probabilities, construct a tree which gives the probable results of the tournament from the round of 16 on. We'll use this tree to find the probable opponent for a given team at a given stage in the tournament.
7. Loop over each team in the tournament and - for each team - iterate over each round in the tournament, find the teams probable opponent in that round using the tree from (7), and calculate the teams probability of being knocked out in that round.

More Details:

2. The additional data I gathered was simply the FIFA rank for teams, easily found on the FIFA site (<http://www.fifa.com/worldranking/rankingtable/>) and a rudimentary listing of players scraped from <http://pesdb.net/pes2014/index.php> into a csv file.

3. A team's base score was derived as a function of the average skill of its players, chemistry score (calculated as the ratio of a team's wins to its players' skill level), number of star players (the top 20 players based on skill level), and FIFA ranking.

4. A team's weighted score against another team was calculated as a function of it's base score, whether or not they "play up" (determined by whether a large portion of their wins/draws came against teams with a higher base score than their own), and past international results against the other team (including second-degree wins - that is, they beat a team which beat the other team).

5. Using the weighted scores found in (4), determining win probabilities was a fairly straightforward matter, simply requiring dividing a team's weighted score against another team by the sum of both team's weighted scores against each other.

Unfortunately - or rather, fortunately - predicting the winner of a football game is far from trivial. My model doesn't account for referee errors, injuries, weather, fans, and a host of other variables. However, I do feel it provides a solid foundation, although only June 12th will tell if my confidence is well-founded.