Case Study: Using ML to Simulate World Cup Matches

I used a Monte Carlo Simulation to simulate the 2018 World Cup and the probable winner. Attempting to simulate a 16-team bracket can quickly reveal the possible permutations and combinations of outcomes, so I knew it was necessary to enlist the help of a machine to hasten and improve the accuracy of this simulation.

Monte Carlo simulations are useful when the system in question is complex, has several inputs, and the possible outcomes are numerous. Outside of sports, it is popular in finance, engineering, insurance, and healthcare industries.

After scraping the necessary data from internet sources, and cleaning and manipulating to improve ease of interaction with the computer, I ran different classifier algorithms to determine the best one.

The three classifier algorithms were logistic regression, random forest, and a linear support vector classifier. Logistic regression is used for binary classification tasks; it is simple and interpretable and is fast to train. Random forest is a collective approach that uses multiple decision trees in combination to improve performance and overfitting. It can work with many features but can be slower to train and less interpretable. Linear SVC works well with high-dimensional data and is most effective when the margin separating classes is clear (winners and losers).

With linear SVC boasting the highest accuracy score, it was then used in conjunction with a Monte Carlo Simulation to determine the most likely winner of the World Cup. Using a Monte Carlo Simulation in tandem with linear SVC allows me to simulate multiple datasets and account for the uncertainties in this real-world data.

While humans are brilliant in many ways, using them to project a winner in a World Cup bracket is not appropriate, since most people fail to not only view the world in terms of probabilities, but they cannot process large amounts of data, simulate several different scenarios, identify hidden patterns, and provide objective, data-driven predictions (everything necessary in this situation).