

Predicting the Outcome of Soccer Games
Joshua Phelps, Riley Ross, Devin Frandsen, Brian Hayes
Interest: 10/10

Description:

We want to see if it's possible to predict the outcome of soccer games across Europe. The European soccer season is heating up which means that most of our weekends are spent watching teams like FC Barcelona, Liverpool, AS Roma, Newcastle United, and any others we can squeeze in. There are tons of analysts, experts, and sports personalities that try to predict outcomes of games. However, people have biases, misunderstandings, and are very flawed when it comes to analysis and predictions. We want to build a machine learning model that can accurately predict the outcome of soccer games.

Features:

There are many features that could be used in the prediction model. On the ESPN website there is plenty of data such as the records of the teams, their recent form, goals scored, goals allowed, and other statistics. In our project we could also include the lineup used, injuries, home or away, average expected goals or shots per game, outcomes of recent matchups between the teams, possession average, and more. These features and more will be used to predict which team will win the game.

Example:

Form	Formation	Injuries	Home	Shots	Prior	Poss	Points	XG	XGA	Value	Outcome
3	2	1	1	4.5	3	55.4	34	3.3	1.7	\$100m	1

Features:

Form (Ws in last 5 matches): nominal

Formation: class

Injuries: nominal

Home or Away: class

Avg. shots per game: continuous

Prior matchups (Ws in last 5 matchups): nominal

Avg. possession: continuous

Current points in table: nominal

Value of Squad: continuous

Outcome: class

Gathering the Data:

Most of this data will come from espn.com, fotmob.com, fbref.com, and other websites that collect data and statistics on soccer. This data is widely available, but compiling it will likely require a web scraper. Once compiled, we can polish and refine the data and feature set.