

Project Luther Proposal

What will be the final score of an EPL match?
All the EPL matches remainder of the season?
Who will be crowned Champion of the EPL?

Michael Weber

Scope

More than 4 out of 10 people in the world consider themselves soccer fans according to Nielsen data(Bloomberg.com) Soccer or football differs from most other sports in that one outcome can determine a match, a single GOALLLLLLL!!!!!!

The world's most popular football league is the English Premier League (EPL). My favorite team is Liverpool FC currently on top of the EPL table.
Can you build a prediction model using linear regression to predict the final scores for EPL games for the rest of this season and determine who will end the season with the most points aka who will be crowned CHAMPION of the premier league?
(something Liverpool hasn't done in almost 30 yrs)

Methodology

- 1) Get EPL results data for all matches last n years
- 2) Determine by team probability of goals per match
- 3) Build linear regression model relating match statistics to goals per game

Data

DataHub.io / premierleague.com
Using match EPL data from DataHub.io
Web scrape for additional feature data

Features

Date the match was played
Home Team
Away Team
Full time Goals for each
Full Time Result
Referee

Team Shots
Team Shots on Target
Fouls committed
Corner kicks
Yellow cards
Red cards

If I can also use through web scrape:

Passing statistics
Possession %
Total Team Wages
Kick off Time
Temperature at match start
How much it rained that day
Match attendance
Day of the week
Manager of each team at time of game
Ground the match was played on