

# Predicting Market Value of English Premier League Players

Munwon Jung

## Abstract

Being able to predict the market value of players is essential for soccer clubs. Clubs pay an enormous amount of money for players and they will have to accurately know whether each player is worth the money so they can take careful consideration when buying/selling players. The goal of this project is to predict market value of English Premier League Players using different regression modeling.

## Design

This project was designed to investigate and predict the market value of all EPL players in the current season. Market value of a soccer player is crucial as it is directly related with potential transfer cost and salary. Being able to predict the current market value of a player via various regression techniques is essential to be successful in the transfer market and be aware of how valuable each player is in each club.

## Data

The data that was used in this project was scraped using BeautifulSoup and Selenium from 2 sites. One was from the official English Premier League site: <https://www.premierleague.com/> and the other site: <https://www.transfermarkt.com/wettbewerb/startseite/wettbewerb/GB1> was used to scrape market value.

The current ongoing season consists of 850+ players with 30+ features. However, the actual cleaned data was trimmed down to 13 features including 1 categorical variable with 586 observations.

## Algorithms

### Feature Engineering/Models

- Converting categorical features to dummy variables
- Linear Regression
- Ridge Regression
- Lasso Regression

The dataset was split into 80/20 train+validation/test dataset. The candidate models were trained with linear, ridge and lasso regression. Then these candidates were scored using the validation set. Finally, the best model out of 3 was chosen to predict on the remaining 20% train set.

## Tools

- Web scraping tools(BeautifulSoup, Selenium)
- Numpy, Pandas Library for data manipulation
- Seaborn, Matplotlib for data visualization

## Communication

With accompanying PPT slides and visuals, the project was shared with METIS Data Science workgroup and posted on Github.

