What factors influence the overall rating of soccer players in the FIFA video game*

Yizhuo Liu, Leo Cheng, Haobo Ren

October-02-2024

Contributions

Liu Yizhuo: Data cleaning, introduction
Leo Cheng: Preliminary result
Haobo Ren: Data description

Introduction

In the long history of soccer, there have been many ways of evaluating the performance of soccer players. Whether by the number of goals they score or the number of trophies they earn, but these are very unilateral methods that only favor a certain kind of player, not everyone, especially when it comes to unknown players. Certainly, the video game franchise FIFA comes to mind. Every year FIFA evaluates professional soccer player's previous year's performance and gives them a rating from 0 to 100. From 1993 to now, there has been a new FIFA game every single year, yet no one knows what factors influence the rating from FIFA that the majority of the people agree on. Therefore, the main objective of this report and the research question is going to be "What factors influence the overall rating of soccer players in the FIFA video game". If we can accomplish this objective, then we can use it to quantify players' performance. Therefore, use it to predict further performance and other stuff related to performance. With our research question in mind, we proposed a hypothesis: factors such as shooting, wage, international reputation, passing, and physics will increase with the response variable - FIFA rating. To back up our hypothesis, the article "Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques" (Al-Asadi and Tasdemır, 2022), which is on using FIFA to get a data-driven approach to player valuation, also uses very similar predictors that we have chosen, and the base model they end up using is also linear regression. In addition, in "PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach" (Pappalardo et al., 2019) they use real-world data to quantify players' performance with a reasonable success. Last but not least, the article "Predicting the Future Performance of Soccer Players" (Arndt and Brefeld, 2016), uses a combination of linear regression and multitask regression, to evaluate current player performance and predict the future to predict the outcome of a soccer game. Two out of the three articles have chosen linear regression, therefore it shows that linear regression is commonly used in this area. Also, by looking at the scatter plot for our response variables and predictions, there is a linear relationship. By using linear regression method we are looking to get an accuracy prediction of FIFA rating.

*Code and data are available at: https://github.com/HaoboRrrr/FIFA20_Player_Potential_Rating_Analysis/tree/main

Data Description

PLOTS & DESCRIPTIONS

Methods

We outline the our methods as following: First, we check the linear assumption for each of our predictors. Second, we fit our preliminary model and plot the corresponding diagrams to address the potential problem of our preliminary model. Third, we use the diagrams plotted to decide whether to perform transformations and add interaction terms to the preliminary model. Fourth, on the newly developed model we will perform hypothesis test to test the significance of the coefficients. Finally, we will interpret the model and address our research question. The following sections give a detailed explanation.

Linear Assumption Check

Variable Selection

Interpretations and Validations

Results

Conclusions