**Football Players Market Value Prediction: A Literature Review**

**Abstract**

Analysis of the data has been done in various fields in today’s world be it business, health etc. But in recent years there has been its impact lately in the Sports which statisticians have taken a liking to. Football, most popular and watched sport in the world, has been affected too. But it was difficult to analyze the football data until recently because of difficulty to collect data and organize as it is the sport which is played all around the world.

Sure, football is mostly about the goals and cups but it also dependent on hugely on the quality of the players. When its quality of the players, their value in the market also plays a vital role. Huge transfers occur every year where top players are always involved. Recently, Neymar was transferred from FC Barcelona (Spanish Team) to PSG (French Team) for a record-breaking transfer fee of the 222 million euros.

This project mainly focuses on predicting market value of the players using modelling techniques. Various steps are involved to get this job done. Data has to be collected and organized into the features and target value, model this data on various models, test and evaluate the models, and finally make predictions based on the best model and check its accuracy.

Various papers have been published on the similar lines. Few of them have been reviewed to compare what is the current state of the problem and what we approach to do.

**Linear Models**

In the paper by Yuan He, he has predicted the market value of the soccer(football) players using linear modeling techniques. He has collected the data from online resources like transfermarkt.de and Wikipedia. He consolidated the data to contain the personal information of the player and performance data of the player. Here he has tried to fit a linear model to the data which gives the best accuracy. Initially he did some summary statistics between the predictors and dependent variables to know the relationship between them. Then ANOVA was used to get the features which had acceptable p values.

Now he fitted the data with the selected features for various linear models such as OLS(Ordinary Least Squares), KNN (K-Nearest Neighbors),Ridge Regression for various lambdas and PCR( Principal Component Regression). He obtained the results which gave him the least RMS error for the PCR model with 16 predictor variables.

**Neural Networks**

Another approach for this task used by Sourya Dey is Multilayer deeply connected Neural Network. He has used the FIFA 2017 dataset which contains roughly about 15000 players data. He has created a 3 layered dense Neural Network. He has cross-validated the number of neurons for each layer and has chosen the combination which gives the highest accuracy. He tried using different combinations of activation functions like Sigmoid + Softmax etc. for different number of neurons in the hidden layer. Also the accuracy for various learning rate was also calculated. All the optimum results were used for various parameters for modelling. He used the momentum for the gradient descent to avoid the local minimum.

**Recommendation System**

A recommendation system was created by Jaka, Karsten and Kostas for recommending the starting lineup 11 players. They have used Collaborative Filtering Recommendations which uses the preferences or tastes of many users to recommend the new user. It first determines the worst player available based on age, overall and position. And it gets players which are at least better than worst player and selects the top 11 players.

**Players’ Selection**

A model for selecting the players in the football team was developed by Onwuachu Uzohukwu C and P Enyindah in their paper. They have used multiple Neural Network Models for Player’s Resistance, Speed, Physical and Technique. They have used the PES stats dataset. The mean of the output of each Network is taken which is used as a criterion for the selection of the player.

**Our Approach**

In various paper which we have surveyed above majority of the models compare players with different positions which seldom is case in real world. So, we will try to make different use different models for corresponding positions.

We are taking into account of players with high value which may seem like outliers but are integral part. We split the data on the overall performance value to separate them out.

Other models try to predict the transfer value of the player based on only the single year data which doesn’t give much insight of the caliber of the player. But analysis of his performance over the years tells a lot about the player which has high effect in the value of the player.

From dataset of 2014-15 to 2018-19 the market value of the goalkeepers was undervalued. But in the present year, market value of the goalkeepers has risen. None of these above models consider this, but we aim to rectify it by introducing the bias.

Models are predicting the value on the new set of the attribute values which are given by the user which may or may not be real values. But we are trying to predict the value of the existing player taking into the considerations of his performance over the years.