

METIS PROJECT 2 PROJECT PROPOSAL;

PREDICTING THE MARKET VALUE OF A SOCCER PLAYER USING REGRESSION.

Goals! Goals! Goals!

A soccer team is composed of 11 players, whose single objective is to send the ball to the back of the opposing net aka score a goal. Now goals scored are frankly just as important as goals kept OUT of your net, but who even cares about that?

It is no wonder therefore that the demand for goal scorers from any part of the football pitch supersedes the need for every other type of player, even though they are all equally necessary.

The Question/Need

Goals scored in real-time has proved to be a woefully inadequate method of gauging a player's value since decisions on how much he might be paid, contract review/renewal, club transfer etc.. are often made well before the first goal of the season is scored.

A quantifiable measure of the expectation beyond the probability to score is required, hence the need to predict the number of goals expected – though not necessarily assured – of any player.

This project attempts to do exactly that, using a previous season's performance statistics. Teams, agents, managers, well....everyone - except the players, of course – needs this.

Data Description

The project will employ web-scraping techniques to obtain performance from stats websites like fbref.com, understat.com etc. which maintain data on every single player in almost every sports tournament on the globe. Focus however, will be on the 5 major European soccer leagues.

Features will include Age, height, sprint speed, dribbling ability, dead-balls taken, dead-balls scored, goal-shots taken (but not scored), goals-shots scored, penalties taken, penalties scored, assists, discipline index (yellow and red cards accumulated), time on the pitch (minutes as a measure of ball exposure), fixtures played, 1st-half goals scored, 90+minute goals scored, player position (defender mid-fielder, winger, striker) and many others.

The target is to make the goals score a function of the rest of the features, train the model to predict goals scored and test it with features from the current season.

The project will use Selenium (if required), Python's BeautifulSoup, sklearn, numpy, pandas for model training and prediction.

MVP:

A model fitted on training data with as near to a normally distributed residuals plot as possible.

Dennis Ssekamaanya