Group 3: Sanjay John, Debosmit Majumder, Unnikrishnan Sivaprasad, Tianyi Yang

ECE 143: Project Proposal

Visualisation & Analysis Of FIFA Player Attribute Dataset

Github link(https://github.com/Deb1993/ECE_143_Group_3)

Overview

Sport analytics has become a burgeoning field in the world of professional sports. FIFA, the world governing body in soccer, has been accumulating data on soccer players since the inception of the videogame series created by EA Sports. SoFIFA.com (https://sofifa.com) has accumulated this dataset onto its website; this data contains widespread information on player attributes from the qualitative, like relative agility and shot power, to the quantitative, such as height and weight. Our analysis will serve to extract valuable information for coaches, team management and scouts. Such insights include determining what attributes are found to be most successful for a particular position on the field, what factors determine the long term trend of a player's success rate, and what kinds of players work best together in a team. With our analysis, we will also create several visualizations to compare players and determine effectiveness of formations, therefore allowing soccer clubs to make better decisions regarding playing strategies and transfers.

Real World Applications

Ever since the release of the movie 'Moneyball', analytics in sports has become a hot topic among sports pundits. Many modern soccer teams have already opened up their analytics wing to improve their performance on the field. In this project, we will be using the FIFA player attributes to provide insights to managers and scouts. Our analysis techniques and visualisation methods can be used on similar datasets belonging to any other team sport and even some solo ones. The following are some of the possible real world applications:

1. Suggesting the ideal position of developing player based on his current skills
2. Find possible role changes for a player based on his skills
3. Provide suggestions on player contract extensions and future wages

Dataset

We will be primarily using the open source SOFIFA dataset (https://sofifa.com) collected from 25,000+ matches in world-wide leagues, comprising player's attributes, and team attributes including their formations in the matches:

- Physical attributes such as age, height, weight, preferred foot
- Attacking and movement records
- Assessment of skills, mentality, defending, and goalkeeping
- Evaluation of overall ratings and potential
- Records of contracts and teams, as well as their wages and values

Additionally, we will be supplementing this dataset with additional data from the official FIFA website (http://www.fifa.com/fifa-tournaments/players-coaches/index.html) to get the number of goals each player made.

Pipeline

Our system primarily consists of the following three steps:

1) Data Scraping and Data Cleaning - This will involve parsing of data from the website and extracting them using open source python libraries like 'Scrapy', etc. After the extraction process, the raw data can be arranged and organized in a suitable format for the subsequent steps.
2) Data Visualisation - This is the heart of our project which will involve creation of visual representation of the obtained data. This will include plotting bar graphs, trend charts, radar plots using matplotlib and bokeh. Also, we will map players and their attributes to the different regions of the world they belong to.
3) Analysis of Trends, Correlation between attributes and Player Profiling - We plan to perform long-term trend analysis with attributes like rating, ranking, age and various skill-sets. We plan to suggest ideal position for players based on their skills. We also plan to provide suggestions for contracts and future wages. In addition, we will use long-term trends to indicate whether or not players have reached their peak performance, create possible transfer scenarios, etc.

Schedule

Task	**Duration**	**People**
Data Scraping and Data Cleaning	May 7-14	Sanjay, Tianyi, Unnikrishnan
Data Visualization	May 14-28	Sanjay, Debosmit, Tianyi
Analysis & Correlation	May 16-31	Debosmit, Unnikrishnan
Presentation	June 1-3	Everyone