A Data Science Approach to Football Team Player Selection

Aim of the paper:

This paper presents a data science approach to minimize the time taken in selecting a player for a team by considering the cost and player's skills as constraints. Such an analysis will help an owner to maximize the profit and popularity of an existing club or to create a new club. We present statistical analysis of player performance based on abilities and skills for a new team using powerBI and Python Pandas by minimizing the cost.

Dataset used:

FIFA dataset

Machine learning model used:

QingWang, Hengshu Zhu, Wei Hu etc, proposed unsupervised approach to evaluate typical tactics using Team Tactic Topic Model (T3M) learning technique based on position and passing relations of football players. it also concentrates collaborative team changing actions to score goal. it bounded to features like pass segment, spatial analysis, tactical pattern discovery, player roles.

Mehrsan Javan, Philippe Desaulniers etc, illustrate player performance evaluation system based on locations, actions, players chances to achieve the next goal. It exploits, machine learning techniques, Markov model to identify the game circumstances to predict next goal and to group similar locations of actions with high and low values. It also gives statistical

modelling evaluations among protective players and their corresponding context actions to predict a goal

Statistical measures used:

Luca P, Paolo C etc, provided data driven role aware multi-dimensional evaluation of player ranking based on player versatility, players performance and representation of more customization of detection of players in terms of role, versatility, weight from match to match, no of teams and countries.

Luke Bornn and Javier Fernandez etc, described spatiotemporal data analysis of soccer data to mine meaningful insights using quality, position, frequencies, success of space occupation and goal generation. It also provides performance team cooperative dynamic decisions during off ball soccer activities to pitch control and modelling and bounded to game specific, current players actions, space complexity utilization for goal generation.

Mehrsan Javan, Philippe Desaulniers etc, illustrate player performance evaluation system based on locations, actions, players chances to achieve the next goal.

Most approaches and methods concentrate on dynamic analytics and represent summarized score boards during game time based on situations, context, opponent analysis, goal generation and post analytics, performance issues of different players. In the proposed approach we mainly focus on prior analysis of a players selection based on performance, skill set, forming team, minimize time selection by reducing the cost initially to avoid further consequences risk factors of unproductive team environment.

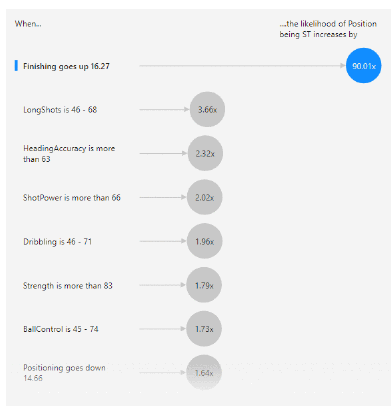
Overall Distribution Value of each player according to overall Rating, performance and age factors.

Number of players distribution from each nationality and their social impacts for investors to select a player as brand ambassadors.

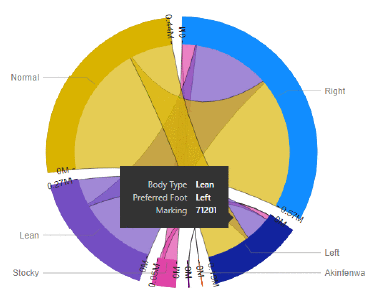
Comparing overall Performance and potentiality of players by nationality

Findings:

Key influencers vary for every position for Stricker's position finishing is the key influencer, in view of all the key influencers above tolerance value using correlation matrix, gives accurate and precise results.



For a new player wage is depends on value, it can be from 1 % to 10% of value but for a player taken from another club expects more, otherwise he may reject to join this may leads to business disappointment. The figure defines the clubs with its average wage amount. Club Real Madrid has the highest wage value.



These analytical results are also helpful for many NGOs to motivate people with physical disabilities. For example, an NGO wants to motivate lean body children or low self-esteemed single dexterous people of different cultures, they can make a report of football players with those abilities can also have higher market value.

Conclusion:

The power of analysis of Data Science makes businesses become more profitable. Business decisions are often tough to make in timely manner for complex datasets.

This paper makes reduction of selection of players risk factors up to higher initially (i.e, 50%), by prominence different players features based on market value, popularity, player's quality, his performance on nature of contest for national teams. we also exploited to minimizing time, cost, inconsistency between approximate and real market values in terms of selection of players for a team. These kinds of approaches principally can be used for managing and commercialization financial profit of sports analytics.

Why this is important:

1. To identify and Compare player's individual performance by their age and nationality
2. To identify and Compare player's value and their overall performance
3. To correlate skills of players to predict player's position
4. To perform Cluster/classification Techniques for positions of players according to their age and overall performance
5. Comparing the qualitative and quantitative properties towards reputation

My critiques:

The primary objective of the paper is to better help football team managers select their players to build a better and stronger team. The approach taken by the author is holistic and provides several key insights.

An additional direction, innovative based AI solutions trying to be incorporated comparative based analytical results for progressing future enhancements for decision making on different data sets by considering other aspects like player injuries, GPS data, spatio-temporal, players video performance extraction and trajectories data.

**Done by – Siddarth S (PES2UG20CS336)**