Scouting Players with FIFA19

Applying Data Mining to Scouting

Data Driven Approach to Scouting

In the era of eight-figure salaries and nine figure signing fees, player recruitment is a high-stakes game. In the past, soccer scouts have relied on rudimentary data and intuition to evaluate the performance and value of soccer players. With the recent rise in data analytics that can capture many aspects of a player's performance, statistics and data science are beginning to play a more prominent role in identifying rising stars and overvalued / undervalued players.

For this project, we are positioning ourselves as a scouting agency that uses analytics to, among other things, enhance the discovery of talents and help soccer clubs better understand the dynamics (features) that come into play when determining the value, overall and future potential of a player. Our agency will be focusing on solving these fundamental scouting problems:

1. Finding undervalued players for a given club to acquire,
2. Analyzing a team's current roster for over-payed and/or underperforming players that could be traded or sold,
3. Developing a database of similar players for clubs looking for a specific player type,
4. Build a predictive model to evaluate the future potential of young players.

We will be utilizing the FIFA 19 Player dataset available on Kaggle and apply various Data Mining techniques to achieve our objectives.

Project Objectives

- Cluster players based various features to identify different player types for our similarity database.
- Identify under-valued and over-valued players based on ability measures relative to their value, salary, and/or release clause.
- Building predictive models for future value and potential of players.

Dataset

- **Source:** Kaggle
- **Description:** Detailed attributes for every player registered in the latest edition of FIFA 2019 database.
- **Size:** 9.1MB (18.2k observations x 89 features)
- **Features:**

• ID • Name • Age • Photo • Nationality • Overall • Potential • Club • Position	• Value • Wage • Special • Preferred Foot • International Reputation • Weak Foot • Skill Moves • Work Rate • Jersey Number	• Joined • Loaned From • Contract Valid Until • Height • Weight • Ability by positions (26 features) • Ability by skills (34 features) • Release Clause

Team & Roles

- **Markus Wehr:** Finding undervalued players.
- **Nazih Kalo**: Analyzing current roster of players.
- **Stephen Stark:** Developing similarity database.
- **Tam Nguyen:** Predictive model for future potential/value.
- **Woo Jong Choi**: Predictive model for future potential/value.

Data Mining Steps:

Data pre-processing	• Missing value, data type • Features distribution • Feature engineering
Analysis Stages	1. Pre-processing and EDA 2. Clustering 3. Build predictive models 4. Analyze performance & make final predictions 5. Visualize Output
Potential Methods	• PCA • t-SNE • K-means • DBSCAN • SVD • Regression: linear/ logit • Hierarchical Clustering • Latent Class Clustering • Discriminant Analysis • Regression Trees • Random forest • Decision trees • Association rules
Tools	1. Microsoft Teams 2. Python – Jupyter Notebook, Google Collab

	– Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Scipy 3. Tableau