Player Recommendation System

1st Oğuz Kağan Pürçek
AI and Data Engineering
Istanbul Technical University
150220759
purcek20@itu.edu.tr

2nd Emirberk Almacı
AI and Data Engineering
Istanbul Technical University
150220751
almaci20@itu.edu.tr

I. INTRODUCTION

Football, known as the world's most popular sport, is characterized by its complexity and the numerous variables that impact team success. In this highly competitive environment, making the right choices regarding player selection, team composition, and tactical strategies is paramount for success. Our project aims to redefine how football talent is scouted, selected, and harnessed, ushering in a new era of data-driven excellence in the world's most beloved sport.

A. Problem Statement

The challenge we seek to address is the need for data-driven player recommendations in football. Conventional scouting methods tend to lack objectivity and efficiency, often relying on human judgments alone, which can lead to missed opportunities and subpar team performance. On the other hand, the football industry is undergoing a data-driven transformation, where insights derived from player statistics, performance data, and other relevant sources can make a substantial difference in team outcomes.

B. Project Objectives

Our Football Player Recommendation Project holds the promise of transforming the football landscape by introducing data-driven player recommendations that consider a multitude of football-related data sources, emerging talents, and evolving game dynamics. This innovation empowers football clubs, coaches, and team managers to make more informed and objective decisions, be it in player acquisitions, team formations, or game plans, ultimately enhancing their chances of success. Moreover, by contributing to the professional development of football players through identifying areas for improvement and providing actionable insights, we assist coaches in optimizing player strategies. Simultaneously, we enhance the football experience for fans by providing them with data-driven insights, statistics, and player recommendations, making football an even more immersive and interactive sport to follow.

C. Team Member

Two people will work on the project. Oğuz Kağan Pürçek and Emirberk Almacı. The division of labor will be balanced and the process will be progressed together. A joint study is planned in the process at stages such as data collection, data processing and model development. Two researchers will also play a role at each stage.

II. DATASET

We will utilize the "Football Manager 2023 Dataset" from Kaggle [1]. This data set contains a lot of information about football players. It contains many features as well as general information such as age, team, nationality. It has various features under headings such as physical, mental and technical. There is detailed information such as the player's current and possible ability. The data also includes how productive the player is in which regions.

The data is available as a .csv file. It consists of 8452 rows (football player) and 98 columns (features). The columns contain generally necessary information and most of them will be used. A few columns that will not be used will be removed from the data. Additionally, if necessary, object columns will be converted to numbers. Empty cells have been detected in some columns. These will be filled with appropriate values. For now, this data set is thought to be sufficient for the model to be created. However, if the project will be developed based on the player's performance, extra data sets will be needed.

Some columns in the data:

- **CA:** Standing for Current Ability, CA is measured on a scale of 1-200 with the higher ratings being reserved for truly world-class talent.
- **PA:** The Potential Ability of a player will determine the maximum possible level a player can reach during their career.
- **Leadership:** Ability to inspire, guide, and effectively manage a team of players, fostering a cohesive and winning environment both on and off the field.
- **Eccentricity:** Unconventional or unique traits, behaviors, or decision-making that deviate from traditional or expected norms within the sport.

III. METHODOLOGY

Our methodology will consist of the following steps:

1) **Data Collection:** As mentioned in the data set section, FM 2023 data will be used.
2) **Data Preprocessing:** Data preprocessing is a crucial step to ensure the quality and suitability of the dataset for modeling. We will perform the following tasks:

- **Handling Missing Values:** Identify and handle missing data using techniques such as imputation or data removal.
- **Data Cleaning:** Columns that will not be used will be removed from the data set.
- **Feature Engineering:** New features will be created that can be informative for the recommendation model, such as player attribute indices or position-specific attributes

3) **Feature Selection:** Selecting relevant features is essential to build an effective recommendation model. We will use techniques like correlation analysis and feature importance ranking to identify the attributes that have the most significant impact on player performance.
4) **Model Selection:** We will try many recommendation algorithms and plan to use the one with the best performance on the model.
5) **Model Training:** The selected models will be trained on the data using different techniques to ensure robust performance.
6) **Optimization:** We will fine-tune the models and optimize player selection strategies based on performance metrics.

IV. EVALUATION METHODS

To measure the success of our project, we will employ the following evaluation methods:

- **Performance Metrics:** Our evaluation methods will rely on the following performance metrics:
  - **Accuracy:** Calculated as the ratio of correctly recommended players to the total recommendations, providing an overall measure of recommendation quality.
  - **Precision:** Precision indicates the fraction of recommended players who were truly successful, emphasizing the relevance of recommendations.
  - **Recall:** Recall measures the proportion of successful players that were correctly recommended, highlighting the completeness of recommendations.
  - **F1-score:** The F1-score balances precision and recall, providing a comprehensive assessment of recommendation quality.
- **Cross-Validation:** To ensure our model's generalization to unseen data, we will implement cross-validation. This technique involves splitting the dataset into multiple subsets, training the model on different subsets, and validating it on the remaining data. This process helps us identify and mitigate overfitting.
- **Comparative Benchmarking:** To assess the effectiveness of our recommendation model, we will establish a baseline model or benchmark. This benchmark could be a simple rule-based system or a basic recommendation algorithm. Comparing our model's performance against the benchmark will provide a clear measure of improvement.

These evaluation methods will enable us to determine the effectiveness of our recommendation model and its potential to enhance player selection strategies.

V. GITHUB REPOSITORY

Github Link of Project: https://github.com/emirberk-almaci/Group_27

VI. CONCLUSION

In conclusion, our project aims to create a football player recommendation model using a comprehensive dataset from Kaggle. By addressing the challenge of optimizing player selection, we intend to have a substantial impact on the football gaming experience, benefiting both teams and individual players. The importance of data-driven decisions in football cannot be overstated, and this project offers an opportunity to leverage the power of data and machine learning for improved performance.

REFERENCES

[1] https://www.kaggle.com/datasets/platinum22/foot-ball-manager-2023-dataset