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

***Abstract*—In this paper, we introduce a novel Football Player Recommendation System, harnessing the extensive dataset from Football Manager 2023. This system is designed to revolutionize football scouting by integrating data analytics and machine learning techniques. The core methodology involves meticulous data cleaning and preprocessing of a dataset comprising 8,452 players with 98 different attributes. The Apriori algorithm is employed for the strategic selection of player features, focusing on optimizing player positions and roles. Preliminary results from the system demonstrate its effectiveness in pinpointing key attributes for distinct playing positions, alongside the capability to filter player recommendations based on specific parameters such as age, market value, and salary. These findings underscore the system's potential to transform traditional scouting methodologies by offering a more data-centric and precise approach. Looking ahead, the project aims to refine the predictive accuracy of the model and expand its utility across diverse football leagues, setting a new standard for player scouting in football.**

I. Introduction

In recent years, the integration of data analytics and machine learning in sports has profoundly transformed various aspects of gameplay and team management, especially in football. The application of these technologies in player scouting and performance analysis has become a pivotal component in modern football management. This report introduces the development of a Football Player Recommendation System utilizing the comprehensive dataset from Football Manager 2023. The primary objective of this project is to innovate the traditional approach to football scouting by applying sophisticated data analytics and machine learning techniques.

Football scouting, traditionally reliant on the subjective assessment of scouts, faces challenges in consistency and breadth of player evaluation. The advent of big data in sports has opened new avenues for objective and comprehensive player analysis. Data-driven approaches have been increasingly adopted in sports analytics for performance analysis, injury prediction, and player scouting, contributing significantly to strategic decision-making in sports (Lucey et al., 2014; Rein & Memmert, 2016). However, the specific area of player recommendation remains relatively unexplored and holds immense potential for exploration and application.

Our project seeks to bridge this gap by developing a system that can analyze extensive player data and provide targeted player recommendations based on a range of parameters. This system aims not only to enhance the efficiency and accuracy of the scouting process but also to provide a more nuanced understanding of player capabilities and potential.

The methodology of the project involves a detailed analysis of the Football Manager 2023 dataset, which encompasses a wide array of player attributes and statistics. Through meticulous data cleaning and preprocessing, followed by the application of the Apriori algorithm, the system identifies key attributes that are crucial for different player positions. The preliminary findings from this project indicate a promising direction towards a more refined and data-backed approach in football scouting.

In summary, this project represents a significant step forward in the application of data analytics in football, with the potential to revolutionize player scouting and recruitment strategies. The integration of machine learning techniques in evaluating and recommending players based on comprehensive data analysis is expected to offer substantial advantages over traditional scouting methods.

II. Related Work

The field of sports analytics, particularly in football, has witnessed significant advancements over the past decade, with data analytics and machine learning playing pivotal roles. The application of these technologies in player performance and scouting is a burgeoning area of research and development.

One of the early adopters of data analytics in football is the Moneyball approach, which emphasized the use of statistical data for player recruitment and game strategy in baseball (Lewis, 2003). This concept, though originating in baseball, has permeated football, influencing how clubs evaluate and recruit players based on data-driven insights (Pappalardo et al., 2019).

In the realm of player performance analysis, studies have focused on various aspects like physical, technical, and tactical evaluations using data analytics (Sarmento et al., 2018). For instance, Rein and Memmert (2016) explored the use of big data in tactical analysis in football, showcasing how data can inform game strategies. However, these studies primarily emphasize performance analysis during matches, rather than scouting and recruitment.

Player scouting, traditionally reliant on scouts' subjective evaluations, has seen a gradual shift towards data-supported decision-making. Research by Decroos et al. (2019) introduced a machine learning approach to predict the future performance

of soccer players, but it did not extend to a comprehensive recommendation system for scouting.

A notable advancement in player evaluation using data analytics is the development of metrics like Player Efficiency Rating (PER) in basketball, which encompasses a player's statistical performance into a single number (Hollinger, 2004). While such metrics offer a condensed view of a player's abilities, they are yet to be widely adapted and refined for football.

Despite these advancements, there remains a gap in the development of a holistic player recommendation system in football that combines a wide array of attributes, from technical skills to physical and mental traits. The current project addresses this gap by leveraging the extensive Football Manager 2023 dataset to develop a comprehensive system for player scouting and recruitment. This system aims to go beyond performance analysis, offering a data-driven approach to identify and recommend players tailored to specific team needs and strategies.

III. Proposed Work

The proposed Football Player Recommendation System aims to transform the conventional approach to player scouting in football by leveraging the comprehensive dataset from Football Manager 2023. This section outlines the methodology, tools, theoretical framework, and expected outcomes of the project.

Methodology

A. Data Cleaning and Preprocessing

The initial phase involves processing the dataset containing 8,452 players and 98 attributes. This step is crucial for ensuring data quality and relevance. Attributes that do not contribute significantly to player evaluation, such as 'Race', 'Skin Colour', and 'UID', are removed. Additionally, missing values in key columns like 'Club', 'Salary', and 'Rental Club' are addressed through imputation strategies, ensuring a complete dataset for analysis.

B. Feature Selection and Analysis

The core of the project revolves around the strategic selection of player features using the Apriori algorithm, a popular method in data mining for discovering frequent itemsets. This algorithm helps identify which attributes are most relevant for different playing positions, thereby tailoring player recommendations based on positional needs.

C. Player Position and Role Optimization

By analyzing the 'Position' attribute, the system categorizes players into specific roles (e.g., Goalkeeper, Defender, Midfielder, Striker) and further into sub-roles based on their playing side (e.g., Left, Right, Central). This categorization allows for more precise player recommendations.

Fig. 1. Example player profile and important features

Tools and Techniques

The project primarily uses Python for data analysis, given its robust libraries and tools for data manipulation and machine learning. Key libraries include Pandas for data processing, NumPy for numerical operations, and MLxtend for implementing the Apriori algorithm.

Hypotheses and Expected Outcomes

The project hypothesizes that a data-driven approach to player scouting can significantly enhance the accuracy and efficiency of identifying suitable players for specific team needs. We expect the system to provide actionable insights and recommendations, leading to more informed scouting decisions.

Design and Implementation Strategies

The system design focuses on user-friendly interfaces for querying and retrieving player recommendations. Implementation involves iteratively refining the data analysis and feature selection processes, ensuring the system remains adaptable to evolving data patterns and user requirements.

In summary, this project represents a comprehensive approach to applying data analytics in football scouting. By integrating advanced data analysis techniques and machine learning algorithms, it aims to provide a novel tool for clubs and scouts in their player recruitment and evaluation processes.

IV. Experimental Results

The preliminary phase of the Football Player Recommendation System yielded insightful results, underscoring the potential of data-driven approaches in transforming football scouting. This section presents the key findings from the initial experiments and analyses.

Data Analysis Outcomes

A. Feature Relevance

The application of the Apriori algorithm to the Football Manager 2023 dataset revealed significant patterns in player attributes related to specific positions. For instance, for defenders, attributes like 'Tackling', 'Marking', and 'Positioning' were frequently associated, aligning with the conventional understanding of defensive roles.

B. Positional Analysis

Analysis of the 'Position' attribute demonstrated a clear distinction in attribute relevance across different positions. Forwards exhibited a higher frequency of attributes like 'Finishing' and 'Off The Ball', whereas midfielders showed a strong association with 'Passing' and 'Vision'.

Interpretation of Results

The preliminary results validate the hypothesis that player attributes vary significantly based on their playing position and role. The findings align with traditional football theories, indicating the model's accuracy in capturing the nuances of player performance and suitability.

Implications for Player Scouting

These results have important implications for player scouting:

- **Targeted Player Evaluation:** Scouts can focus on specific attributes that are crucial for a given position, enhancing the efficiency of the scouting process.
- **Identifying Undervalued Players:** By understanding attribute importance, scouts can identify players who excel in key areas but may be undervalued in the transfer market.
- **Customized Recommendations:** The system can tailor player recommendations to fit the specific tactical and strategic needs of a team.

In conclusion, the initial experimental results are promising, offering a glimpse into the system's capability to revolutionize player scouting and recruitment. The next phase of the project will involve further refinement of the recommendation algorithm and validation of the system with real-world scouting scenarios.

V. Conclusion

The development of the Football Player Recommendation System represents a significant step forward in the application of data analytics in football scouting. This report has outlined the project's objectives, methodology, and the promising results obtained from the preliminary analysis. Here, we summarize the key aspects and reflect on the broader implications of this work.

Summary of Findings

- **Objective Achievement:** The project successfully developed a system that applies data-driven techniques to enhance player scouting, aligning with the initial objective of leveraging the comprehensive dataset from Football Manager 2023.
- **Methodological Rigor:** Utilizing Python and its powerful libraries, the project effectively cleaned and processed the dataset, applied the Apriori algorithm for feature selection, and analyzed player attributes by position.
- **Insightful Results:** Preliminary results demonstrated clear patterns in player attributes across different positions, confirming the hypothesis that specific skills are more pertinent to certain roles in football.

Implications and Significance

- **Enhanced Scouting Efficiency:** The system provides a more focused and efficient approach to player evaluation, potentially reducing the time and resources expended in traditional scouting methods.
- **Data-Driven Decision Making:** The system provides a more focused and efficient approach to player evaluation, potentially reducing the time and resources expended in traditional scouting methods.
- **Data-Driven Decision Making:** Clubs can use the system to identify undervalued players who excel in key attributes, thereby gaining a competitive edge in player recruitment.

Limitations and Future Research

- **Dataset Constraints:** The reliance on a single dataset may limit the system's applicability across different leagues and player pools. Future versions could integrate diverse data sources for a more holistic analysis.
- **Algorithmic Refinement:** Ongoing optimization of the feature selection algorithm is necessary to adapt to evolving patterns in player performance data.
- **Real-World Validation:** Future research should involve testing the system with actual scouting professionals and integrating their feedback to enhance its practical utility.

The Football Player Recommendation System showcases the transformative potential of data analytics in sports scouting. By offering a novel tool that harnesses the power of data to inform player recruitment decisions, this project contributes significantly to the field of sports analytics. It opens up avenues for further research and development, paving the way for more sophisticated and nuanced approaches to talent identification in football and beyond.

REFERENCES

[1] P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. Matthews, "The role of data analytics in soccer," *Annals of Applied Sport Science*, vol. 2, no. 2, pp. 1-10, 2014.

[2] R. Rein and D. Memmert, "Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science," *SpringerPlus*, vol. 5, no. 1, p. 1410, 2016.

[3] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*, W. W. Norton & Company, 2003.

[4] L. Pappalardo et al., "A public data set of spatio-temporal match events in soccer competitions," *Scientific Data*, vol. 6, no. 1, pp. 1-15, 2019.

[5] H. Sarmento, R. Marcelino, M. T. Anguera, J. Campaniço, N. Matos, and J. C. Leitão, "Match analysis in football: A systematic review," *Journal of Sports Sciences*, vol. 36, no. 7, pp. 717-732, 2018.

[6] R. Rein and D. Memmert, "Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science," *SpringerPlus*, vol. 5, no. 1, p. 1410, 2016.

[7] T. Decroos, V. Dzyuba, J. Van Haaren, and J. Davis, "Actions Speak Louder Than Goals: Valuing Player Actions in Soccer," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

[8] J. Hollinger, *Pro Basketball Forecast: 2004-2005 Edition*, Brassey's, Inc., 2004.