Decision-Support System for Managers:
Prescribing Substitutions for Professional Soccer Games

Atharva Navaratne[1] and Malcolm Gaynor[1]

[1]Sloan School of Management, Massachusetts Institute of Technology
15.095: Machine Learning Under a Modern Optimization Lens

December 9, 2024

MANAGEMENT
SLOAN SCHOOL

Abstract

In professional soccer, substitutions are a critical tactical element that can significantly influence match outcomes. Traditionally, these decisions have relied heavily on a coach's intuition and limited real-time data. This study develops a machine learning-based framework to optimize substitution strategies, leveraging player-specific metrics, match context, and historical substitution patterns. The approach incorporates clustering techniques to categorize players based on their roles and game impact and uses predictive models to suggest optimal substitution timings and candidates. A decision tree-based methodology ensures interpretability and adaptability, enabling practical applications in real-time scenarios. The model is calibrated to align with intuitive decision-making logic, such as emphasizing offensive substitutions when trailing or defensive changes when leading, providing a structured and data-informed approach that complements a coach's expertise.

The proposed system was tested on Manchester United's 2024-2025 season opener, where it successfully recommended substitutions that aligned with the actual game decisions, contributing to a favorable outcome. By identifying substitution clusters and ranking players within those clusters, the framework adapts to varying match scenarios, such as game state and opponent strength. This project bridges the gap between traditional heuristics and advanced data-driven strategies, demonstrating the potential to enhance tactical decisions in soccer. Beyond its application to the English Premier League, the methodology offers insights for broader use in sports analytics, showcasing the transformative role of machine learning in optimizing team performance.

Keywords: Clustering, Soccer, Machine Learning, Premier League, Manchester United

Introduction

The substitution rule in soccer allows coaches to replace 3 (or 5, depending on the competition) outfield players with substitutes from the bench. Originally introduced to combat fatigue and bring on "fresh legs," substitutions have evolved into a critical strategic tool over the decades. Coaches now leverage substitutions not only to manage player fatigue but also to alter game dynamics, disrupt opponents' strategies, and address in-game challenges. This strategic emphasis has grown significantly in recent years, largely due to the increasing number of games in professional soccer (see Figure A.1 in the appendix). With players facing packed schedules, the demand on their physical endurance has risen dramatically. Detailed analyses of player performance data reveal that the time players spend in their "critical zone"—the period of peak physical and mental exertion—has also increased, further underscoring the importance of timely substitutions [1].

In modern soccer, substitutions are pivotal to a team's tactical approach. Deciding which player to bring on is still largely based on the coach's judgment, informed by experience and intuition. However, the growing availability of real-time game data and advancements in analytics provide an exciting opportunity to enhance this decision-making process [4]. By integrating dynamic game data into a decision-support system, it is possible to identify the players most likely to positively impact the game at a given moment. Such systems can analyze various factors, including player fatigue, performance metrics, and opposition strategies, to provide prescriptive analytics that supplement the coach's decisions. We believe the potential for prescriptive analytics in optimizing substitution strategies is immense, offering a data-driven approach to improve outcomes in this increasingly demanding and strategically nuanced sport [1].

Dataset Description

We leveraged three datasets for different parts of the project: two sourced from Kaggle and one from the Fantasy Premier League (FPL). The Kaggle datasets provided match event data and FIFA player ratings, while the FPL dataset offered up-to-date statistics for the ongoing 2024-25

Premier League season. Together, these datasets enabled a comprehensive analysis of substitutions and player performance.

Football Events & FIFA Ratings

The first dataset includes 941,009 notable events such as shots, fouls, and substitutions, derived from 9,074 matches across the five biggest European football leagues—England, Spain, Germany, Italy, and France—spanning the years 2011 to 2017 [4]. The data was preprocessed to focus on substitution events, with each observation capturing match-specific metrics before and after the substitution. These metrics include goals, shot attempts, fouls, corner kicks, and penalty kicks conceded and attempted by the team.

To enrich this dataset, we merged it with a second Kaggle dataset containing FIFA player ratings. These ratings provide qualitative descriptions of each player's skillset across various attributes like passing, shooting, and defending [3]. This integration allowed us to include detailed features for the players subbed in and out during each substitution event. The final dataset, therefore, combines match-level data (pre- and post-substitution metrics) with player-specific qualitative features, providing a holistic view of substitution impacts.

Fantasy Premier League

The Fantasy Premier League (FPL) is an online game where participants manage virtual teams composed of real-life Premier League players, earning points based on their actual match performances. Beyond its entertainment value, FPL serves as a rich source of granular, up-to-date data for analyzing player performance, team dynamics, and game strategies. The platform provides comprehensive statistics for each player and team, including:

- Player-specific metrics including but not limited to goals, assists, minutes played, clean sheets, saves (for Goalkeepers), yellow and red cards, and bonus points.
- Team metrics such as overall performance, fixture difficulty ratings, and transfer trends.
- Advanced metrics including expected goals and assists (xG, xA), influence, creativity, and threat that quantify a player's contribution to match outcomes.

Additionally, the FPL dataset includes contextual data such as player price changes, ownership percentages, and transfer trends, offering insight into player popularity and market dynamics. This data is updated weekly, reflecting changes in player form, fitness, and availability. For our analysis, we utilized a third-party repository to access FPL data and filtered it by team [3,4]. This allowed us to conduct a detailed exploration of individual player performances and team strategies. By integrating these datasets, we created a robust foundation for analyzing substitution patterns and optimizing decision-making in professional soccer.

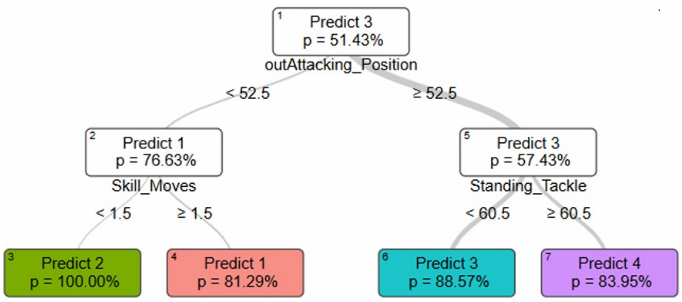
Methods

Clustering Substitution Types

The first step in the analysis involves clustering the types of substitutions. For this step, only the FIFA ratings of the player being subbed in and the player being subbed out are considered. The goal is to classify each substitution as if each cluster represents a distinct treatment, akin to treatment groups in a medical context.

Consulting the scree plot, and keeping in mind that fewer clusters improve interpretability, we choose $k = 4$ clusters. Although the "elbow method" (see Figure A.2 in the appendix) suggests that 5 clusters may be appropriate—indicated by a sharp decrease in the within-cluster sum of squares when moving from 4 to 5 clusters—we opted for 4 clusters to prioritize interpretability.

Next, to interpret the identified clusters, we fit an Optimal Classification Tree (OCT). With a tree depth of just 2, we classify substitutions with an out-of-sample accuracy of approximately 86%, compared to a baseline classification accuracy of under 51%. The key variables influencing the clustering include:



- **outAttacking_Position**: A rating (out of 100) describing the positional attacking skill of the player being subbed out.
- **Skill_Moves**: A rating (out of 5) capturing the offensive technique in one-on-one scenarios for the player being subbed in.
- **Standing_Tackle**: A rating (out of 100) reflecting the defensive technique for dispossessing an attacker for the player being subbed in.

Based on these variables, the four clusters can be intuitively interpreted as follows:

- **Cluster 1**: Offensively skilled player subbed in, defensively skilled player subbed out.
- **Cluster 2**: Defensively skilled player subbed in, defensively skilled player subbed out.
- **Cluster 3**: Offensively skilled player subbed in, offensively skilled player subbed out.
- **Cluster 4**: Defensively skilled player subbed in, offensively skilled player subbed out.

It is essential to note that these clusters do not directly reflect the positional roles of players. Instead, they describe the players' skillsets. For instance, a Cluster 1 substitution could involve swapping a midfielder prioritizing defense for one with a more offensive skillset.

Optimal Prescriptive Tree (OPT)

Once the substitution data is replaced with the corresponding cluster labels, we fit an Optimal Prescriptive Tree (OPT) using pre-substitution match metrics (e.g., goals scored, goals conceded, shot attempts, shot attempts conceded, etc.). Here, the treatment corresponds to the substitution cluster (1–4).

The first outcome metric considered is the **goal differential** (goals scored minus goals conceded) after the substitution. To improve this metric, we introduce a "regularization-inspired" framework defined as follows, for λ between 0 and 1:

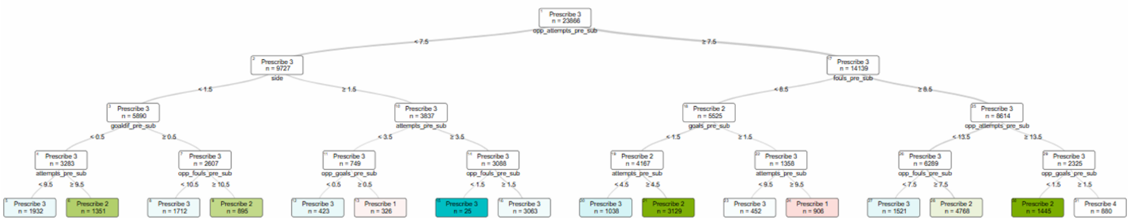
- **When Losing:** Goal Differential $+\lambda\times$ (Shots Attempted).
- **When Winning:** Goal Differential $-\lambda\times$ (Shots Conceded).
- **When Drawing:** Goal Differential $+2\lambda\times$ (Shots Attempted $-$ Shots Conceded).

This new metric accounts for the unpredictability of goal differential by incorporating shot attempts and concessions. For instance, if a team is losing and a substitution generates numerous shot attempts that are saved, the substitution should not be deemed entirely unsuccessful. The framework also reflects different managerial strategies depending on the game state:

- **When Losing:** The focus is on creating offensive opportunities, so only Shots Attempted are included.
- **When Winning:** The emphasis shifts to defensive stability, hence only Shots Conceded are considered.

- **When Drawing:** Both Shots Attempted and Shots Conceded are relevant, but their difference is scaled by 2λ to adjust for smaller variations in this metric.

Using this adjusted outcome metric, we fit an OPT to maximize the prescribed substitutions' effectiveness. The resulting tree has a depth of 4 and provides actionable insights:



- **Proportion of Cluster Prescriptions:**

  1. 5% Cluster 1 (Offensive Player In, Defensive Player Out)
  2. 48% Cluster 2 (Defensive Player In, Defensive Player Out)
  3. 43% Cluster 3 (Offensive Player In, Defensive Player Out)
  4. 4% Cluster 4 (Defensive Player In, Offensive player Out)

  We see from the distribution that we are much more likely to prescribed a balanced substitution. This intuitively makes sense, as only specific drastic scenarios would require such a shift in strategy. Two (of the three) examples of such scenarios from the OPT are given below:

- **Cluster 4 (Defensive Player In, Offensive player Out):** Too many shots attempted by the opposition, too many fouls committed, and opponent has scored at least once. Intuitively, in this situation, we can see that the defense has been insufficient and needs reinforcement, even at a cost to the attack.

- **Cluster 1 (Offensive Player In, Defensive player Out):** Both teams have few attempts, there are too many fouls, and opponent has scored at least once. In this situation, the defenses are playing well, but you need to score and the crowd is hostile, leading to the aggressive decision of adding a more offensively focused substitute.

- **Prescription Breakdown:** Due to the nature of treatment prescriptions, it is impossible to be certain about exact counterfactual calculations. However, it is interesting to note that the OPT prescribed a different treatment (style of substitute) than the treatment observed in the testing set 76% of the time. Of this 76% of the testing set, the OPT predicted a better objective after its prescription than was observed 56% of the time.

Finally, the use of a **doubly robust** approach ensures reliable estimates of treatment effects, combining outcome modeling and propensity scoring to reduce bias and variance in the results.

Clustering Individual Players

Clustering based on attacking and defensive impact is a critical tool for segmenting players into meaningful groups based on their contributions to the offensive and defensive phases of play. This segmentation enables coaches to strategically identify players available on the bench who align with specific game scenarios.

Attacking Impact Score

The attacking impact score is a composite metric designed to quantify a player's contributions to offensive play. This score is derived from a weighted average of various FIFA statistics that measure a player's ability to create scoring opportunities, contribute directly to goals, and progress the ball effectively. The attacking impact score incorporates the following key metrics:

- **Goals and Assists:** These are direct contributions to scoring, with goals reflecting the player's ability to finish chances and assists indicating their ability to create opportunities for teammates.
- **Key Passes:** This metric measures a player's ability to make passes that lead directly to a shot on goal. It highlights a player's creativity and vision in setting up scoring opportunities.
- **Progressive Passes/Carries:** These statistics reflect how effectively a player advances the ball into dangerous areas. Progressive passes measure forward passes that move the ball at least 10 yards, while progressive carries track how far a player moves the ball while dribbling. Both metrics are crucial for breaking defensive lines and creating attacking opportunities.

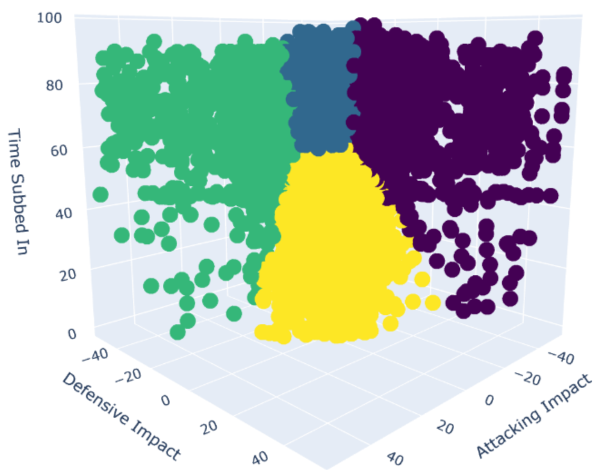
These statistics are weighted according to their relative importance in contributing to offensive success, and the resulting score provides an aggregate measure of a player's attacking impact during a match.

Defensive Impact Score

The defensive impact score is similarly calculated as a weighted average of key FIFA metrics that capture a player's contributions to defensive play. This score reflects a player's ability to prevent the opposition from scoring, disrupt the opponent's attacks, and maintain defensive solidity. The defensive impact score includes the following metrics:

- **Tackles and Interceptions:** Tackles measure a player's ability to regain possession by challenging for the ball, while interceptions track their ability to read the game and intercept passes. Both are critical for stopping opposition attacks and winning the ball back.
- **Pressure Success Rate:** This metric measures how often a player successfully wins possession or forces an error by pressuring an opponent. It reflects a player's effectiveness in pressing the ball and disrupting the opposition's rhythm.
- **Blocks and Clearances:** Blocks refer to a player's ability to stop shots or passes with their body, while clearances indicate their ability to remove the ball from dangerous areas, typically under pressure. Both are crucial for maintaining defensive stability, particularly in the final third of the field.

These metrics were aggregated as weighted averages based on prior studies, with penalties for disciplinary actions (e.g., yellow or red cards) and adjustments for athleticism (e.g., stamina, vision, composure). Clustering was then performed using these impact scores and the time of substitution, enabling insights into duration-dependent player impacts.



Interpreting Individual Player Clusters

We obtain 4 clusters from this approach. The skree plot (see Figure A.3 in the appendix) seems to suggest k of more than 5 as optimal, but we pick 4 for interpretability. To interpret the clusters, we fit an **Optimal Classification Tree (OCT)** with parallel splits (see figure A.5 in the appendix). The clusters obtained can be summarized as follows (as also visible in the figure above):

- **Cluster 1: Early Substitutes**
  Substitutions before the 50th minute, typically involving players with neither high attacking nor high defensive impact scores. This is consistent with logic - coaches hedge their bets earlier on in the game and avoid risky substitutions.
- **Cluster 2: Late Attack-Minded Substitutes**
  Substitutions after the 50th minute with a high attacking impact score but a low defensive impact score. This could indicate the team needs a goal in the final stages of the game.
- **Cluster 3: Late Defense-Minded Substitutes**
  Substitutions after the 50th minute with a high defensive impact score but a low attacking impact score. This could indicate a defensive reinforcement to avoid the opposition from scoring.
- **Cluster 4: Late Balanced Substitutes**
  Substitutions after the 50th minute with balanced attacking and defensive impact scores. Rarer than Cluster 2 and 3; according to our EDA these substitutes occur on an average once every 3 games, and never more than once in a game.

This clustering aligns with intuition: early substitutions often aim for balance, while later substitutions tend to be more drastic, either prioritizing attack or defense. Late balanced substitutions (Cluster 4) are relatively rare compared to specialized attacking (Cluster 2) or defensive (Cluster 3) substitutions.

From Cluster to Player Selection

Once the optimal substitution cluster is determined, we must identify the specific player to substitute in. To achieve this, we analyze the probability of improving the goal differential for each player.

Baseline Comparison

For xG, the baseline model used the number of goals scored at the same time last season, while for xGC, it used goals conceded against the same teams last season. This is a simplistic naive baseline which is predominantly used by sport pundits and media outlets to predict player performance. It fails to account for fixture difficulty ratings (FDR), fixture congestion, player form (which we have defined as the average rating out of 10 over the most recent five games), and advanced metrics such as influence, creativity, and threat, collectively called ICT.

Expected Goals (xG) and Expected Goals Conceded (xGC)

For attackers and midfielders, we predict the expected number of goals scored (**xG**), while for defenders, we predict the expected number of goals conceded (**xGC**). An **XGBoost model** was trained using player performance metrics as features, with goals scored or conceded as the target variables. The model achieved an Out-of-Sample R^2 of 0.94 for xG and 0.79 for xGC.We note here that the lower accuracy for xGC is expected, as conceding goals often depends on the entire team, making individual player contributions harder to isolate.

Regularization-Inspired Adjustments

To reflect real-world considerations, we incorporated penalties for:

- **Age:** Younger players are generally inexperienced, and are therefore less likely to be chosen for high-impact games.
- **Minutes Played:** A player that is generally good enough to be selected but has played less minutes this season (due to injury, illness, or any other reason) is also less likely to be selected for high-stakes games.

These penalties were applied for games classified as **difficult** based on the **Opposition Difficulty Index (ODI)**. ODI was defined as a combination of FDR (fixture difficulty rating, supplied in the FPL data) with a positive ϵ added for away games, since away games are generally harder than home games. We found based on our knowledge of the sport that $\epsilon = 0.25$ gives the best intuitive cutoffs. An ODI above 3 indicates a tough opponent.

An additional regularization-inspired adjustment that was made while predicting xG and xGC was based on the number of goals long-staying players have scored or conceded against specific strategic opponents. For instance, attackers with high pace and acceleration ratings do better against teams that are central-defensive (since they're able to exploit the wings and run past the defense); and defenders with a high strength rating do better against teams that use a central striker more than wingers (since they're able to "battle" that central striker, preventing him from scoring). To model this, we introduced binary variables $c_{di} = 1$ if the team's defense is central (0 if wing-oriented), $c_{ai} = 1$ if the team's attack is central (0 if the team is wing-oriented). Using these binary variables, we adjust the probability of goals scored or conceded accordingly for relevant players against relevant oppositions.

This could only be done for long-staying players since for players who recently joined the team, we simply did not have enough data to make such deductions. It may also be noted that, even though the latter adjustment reflects real-world scenarios more accurately, this did not affect our sorting of the substitutes significantly, because the players who are likely to perform well against a certain team has already been picked in the starting XI, and is therefore just not available in the list of substitutes! All metrics were normalized to be between 0 and 1.

Player Selection Process

Once the xG and xGC predictions are obtained:

- **Attackers and Midfielders:** Ranked in descending order of xG. We found xG to be a better metric for midfielders in general than xGC, especially for teams playing in higher divisions.
- **Defenders:** Ranked in ascending order of xGC. It may be noted here that midfielders who are highly defensively oriented were also characterised by xGC rather than xG.

The top-ranked player within the prescribed cluster is selected as the optimal substitute.

Review of the Process

The following is a succinct methodology for substitution prescription, summarizing the preceding sections:

- **Step 1:** Analyze game data after the 60-minute mark (the typical time for strategic substitutions) and derive the Optimal Prescriptive Tree (OPT) prescription.
- **Step 2:** Identify the cluster corresponding to the OPT prescription, narrowing the analysis to the substitutes that belong to this cluster.

- **Step 3:** Within the identified cluster, sort players based on their expected performance metrics:
  - Attackers and midfielders by descending order of xG (expected goals).
  - Defenders by ascending order of xGC (expected goals conceded).
- **Step 4:** Perform a sanity check—validate the substitution's logic against the state of the game. A simplified decision logic is provided below in the appendix (see Figure A.4 in the appendix). (Inspired by [6])

Manchester United vs Fulham

Manchester United faced Fulham in the opening game of the 2024–2025 English Premier League season at Old Trafford on August 16, 2024. We used Manchester United player data to prescribe substitutes. Since most substitutions occur after the 60th minute, we performed our analysis at the 61st minute.

Based on goals, attempts, and other game-specific information from the game report, our OPT prescribed a Class 3 substitution: *Offensive In, Offensive Out*. This meant replacing an attacking-minded player with another attacking-minded player. Based on this prescription, we focused on Cluster 2 players (Late Attack-Minded Substitutes). At that point in the match, Manchester United had four Cluster 2 substitutes on the bench.

Sorting these substitutes by their probability of scoring a goal (xG), we prescribed Joshua Zirkzee and Alejandro Garnacho as the optimal substitutes. This recommendation was consistent with the game state—Manchester United were drawing against a relatively weak team, warranting an offensive substitution.

Interestingly, the substitutions made by Manchester United's coach (now former coach, Erik ten Hag) around the 65th minute were exactly the same as our prescription. One of the substitutes, Joshua Zirkzee, went on to score the winning goal in the game.

Limiting Assumptions

In this section, we highlight the assumptions underlying our analysis and justify their validity:

- **FIFA Ratings as Proxies:** FIFA ratings are used as proxies for player quality since they account for multiple aspects of performance, including attacking and defensive play, athleticism, and leadership. [2] [5]
- **FPL Data as Performance Metrics:** Fantasy Premier League (FPL) data serves as a reliable proxy for player performance due to its detailed point system, which includes advanced metrics such as heatmaps, vision, tackles, and movement, in addition to goals and assists.
- **Simplified Sanity Check:** The final sanity check (substitute logic based on game state) is a heuristic often employed by coaches. While simplified, it reflects practical decision-making in domestic league matches. [6]
- **Exclusion of First-Half Substitutions:** Substitutions in the first half are rare and typically injury-driven. Since our analysis focuses on strategic improvements, we exclude such cases.
- **Ignoring timing and number of substitutions** Managers often make multiple substitutions in a match, often at different times. Neither time nor multiple-substitution interactions are considered in our analysis.

Conclusion

This study demonstrates a robust and data-driven methodology for optimizing substitutions in football matches. By integrating clustering techniques, machine learning models, and game-state logic, we present a framework that prescribes substitutions tailored to maximize the team's performance. The application of this framework to Manchester United's game against Fulham showcases its practical validity and effectiveness.

Future improvements could include incorporating real-time match data and refining the xG and xGC models for higher accuracy, particularly for defensive metrics. Overall, this approach highlights the potential of advanced analytics in enhancing decision-making in football.

Distribution of Work

Atharva: Sourcing, pre-processing, and feature engineering FPL API data; clustering substitutes and using OCT to interpret individual player clusters; creating baseline model from 2023-24 data; XGBoost to predict xG/xGC probability and sorting to prescribe the "optimal substitute".

Malcolm: Sourcing, pre-processing, feature engineering, and combining the Kaggle datasets. Clustering substitutions as "treatments", using an OCT to interpret the clusters, and fitting and analyzing the OPT to perscribe "treatments"

Both: Drafting the report and presentation; getting free coffee at MTC while doing so.

References

[1] FIFPRO. At The Limit: Player Workload in Elite Professional Men's Football. 2021. Available at: `https://fifpro.org/en/supporting-players/player-workload/at-the-limit/`. Accessed: December 2, 2024.

[2] Iglesias, B., García-Ceberino, J. M., García-Rubio, J., & Ibáñez, S. J. (2022). How Do Player Substitutions Influence Men's UEFA Champions League Soccer Matches. *Applied Sciences*, 12(22), 11371. doi:10.3390/app122211371.

[3] Kaggle: antoinekrajnc (via Kaggle). Soccer Players Statistics. 2018. Available at: `https://www.kaggle.com/datasets/antoinekrajnc/soccer-players-statistics/data`. Accessed: December 6, 2024.

[4] Kaggle: Secareanu, A. (via Kaggle). Football Events. 2017.

[5] Myers, B. R. (2012). A Proposed Decision Rule for the Timing of Soccer Substitutions. *Journal of Quantitative Analysis in Sports*, 8(1). doi:10.1515/1559-0410.1349.

[6] Unknown. (2017). Decision Support Enhancement for Player Substitution in Football: A Design Science Approach. In *Business Information Systems Workshops* (pp. 357–366). Springer Nature. Access provided by Massachusetts Institute of Technology.

Appendix

Figure A.1 - The rising number of games per year.

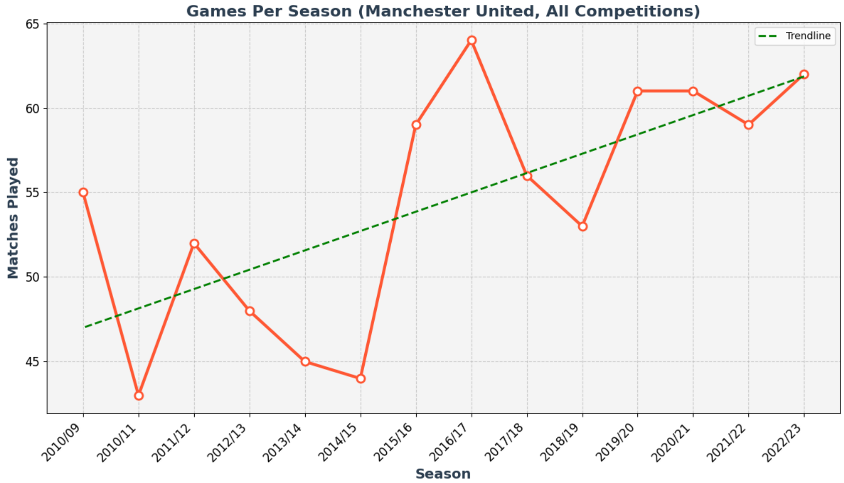


Figure A.2 - Scree plot for clustering of substitution types.

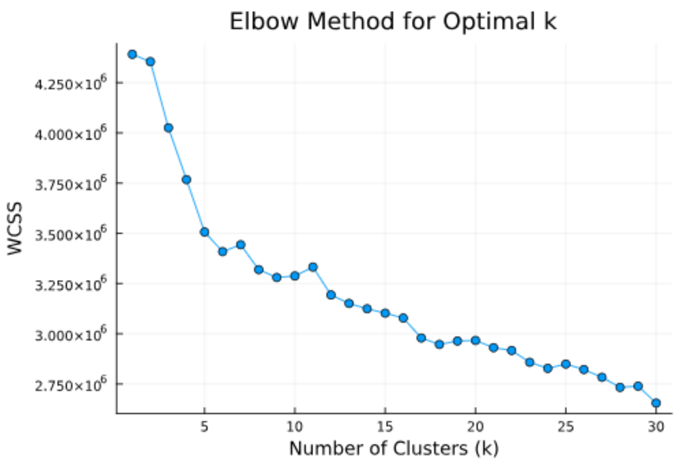


Figure A.3 - Scree plot for clustering of individual players.

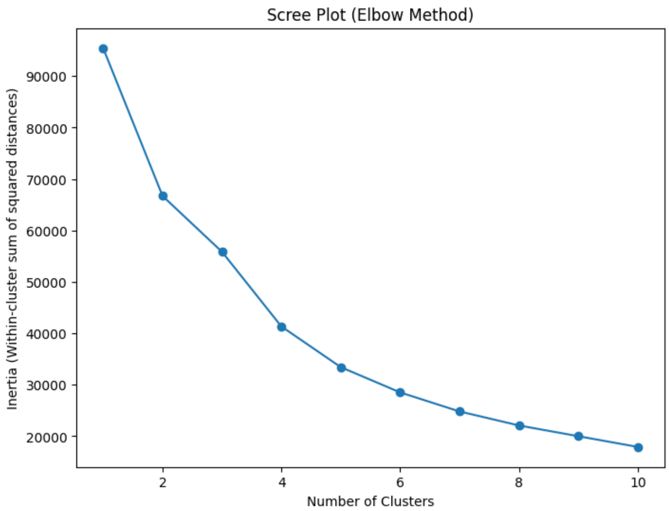


Figure A.4 - A network representation of the decision logic.

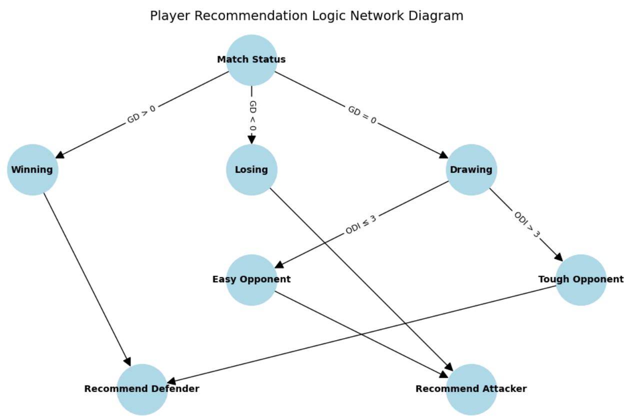


Figure A.5 - OCT used for interpreting individual player clusters.