# THESIS

**Zsolt Lajos Kontra**

**2023**

**Kodolányi János University**
**Business Administration & Management Department**

# Artificial Intelligence-Based Sports Economics Analyses Project Starlight: Robot-Eye for (Turkish) Football Player Talent Identification

**Consultant: Dr. László Pitlik**

**By: Zsolt Lajos Kontra**
**Bachelor of Science in Business Administration & Management**

**Budapest, 2023**

# Special Thanks

I would like to express my sincere gratitude to Dr. László Pitlik, my esteemed consultant, for his invaluable guidance, unwavering support, and exceptional insights throughout the course of my research.

Without Dr. Pitlik's expert knowledge, vast experience, and relentless dedication, this thesis would not have been possible. He has been more than a consultant to me - he has been a mentor, a guide, always ready to offer his wisdom and encouragement whenever I needed it.

Dr. Pitlik's unique perspective and innovative approach have been a true inspiration to me, opening my mind to new ways of thinking and challenging me to push the boundaries of my research. His guidance has been instrumental in shaping my understanding of digital economics, and his passion for the subject has been infectious. Thanks to his help, my data analysis skills have improved so much that I'm now able to find meaningful patterns in a random number generator. His ability to process and interpret complex information is truly mind-boggling. Sometimes, I wonder, if he is an advanced AI-based robot in disguise!

I am deeply grateful for the opportunity to work with Dr. Pitlik, and I look forward to continuing our collaboration in the future. Thank you for being the light in a dark space, illuminating the path towards endless knowledge and discovery.

# Table of Contents

## List of Figures



## Keywords

sport-economics, talent-management, artificial intelligence, optimization, forecasting, closed-loop feedback system, expected return

# 1. Introduction

In the world of sports, teams and clubs are always on the lookout for the best and most valuable players to add to their roster. But how do you determine which players are and will be in the future the most valuable, and which attributes contribute the most to a player's worth? This is a question that has fascinated analysts and fans alike since the game exists, and one that this study examines through a pool of Turkish football players.

In this thesis, the aim is to establish a new benchmark for player valuation in the sports industry. Previous research in this field has analysed various attributes and selection methods to predict player value and earnings. The study builds upon this foundation by introducing a novel approach that utilizes correlation calculations to determine the most significant attributes for player valuation.

Establishing the benchmark involved analysing a dataset of 129 players and 46 professional attributes over a five-year period from 2017 to 2022. A subset of 28 players was identified as the most valuable through a rigorous filtering process. The correlation-based approach was then applied to determine the most significant attributes for each player, and subsequently predict their earnings potential.

The accuracy of the player selection and predicted earnings for each player will be used to establish the benchmark. It is expected that this approach will surpass previous benchmarks, resulting in more precise player selection and higher earnings potential for each player. The goal of creating this new benchmark is to contribute to the ongoing development of player valuation methods within the sports industry. Additionally, it provides valuable insights into the factors that contribute to a player's value in football and how teams can use data-driven approaches to identify and select the best players for their team. It highlights the importance of specific attributes, such as composure, passing, anticipation, decisions, and stamina, and the impact they can have on a player's worth.

Overall, this study is just one example of the potential that data analysis and machine learning, especially the chained similarity analyses can bring to the world of sports. As technology and techniques continue to advance, there is much more that can be done to improve the selection and management of players, ultimately leading to better team performance and success on the field.

I am pleased to announce that the base of this thesis was successfully presented at the International Congress of Finance and Tax, held on March 10-11, 2023, in Konya, Turkey. The congress was organized by Iksad Institute, in partnership with Kodolányi János University, and was attended by a diverse range of experts and scholars in the field. In honour of the of the partner, the selected players in this study are Turkish. Also, I am delighted to report that the study has been published and is available at www.iksadinstitute.org/books .

This thesis introduces a new benchmark for player valuation in the sports industry, using a correlation- and artificial intelligence-based approach to determine the most significant attributes for player valuation and there is even more. Not only that, but it also introduces a brand-new metric, expected return (xR), to predict the earnings potential of a player based on their relevant attributes.

## 1.1. Goals

The goal of this project/thesis is to identify the professional attributes that contribute the most to the value of football players and to develop a methodology for determining the future value development of players based on these attributes. This study focuses on a pool of Turkish football players and uses various statistical methods and models to analyse their performance data and attribute values in order to identify the most valuable players in the future, it means to predict their potential and the earnings it can bring. The findings of this study have implications for player selection, team building, and financial decision-making in the world of football. (The final goal is to establish a working global scouting model.)

## 1.2. Tasks

The study's primary objective was to identify the top attributes that contribute to a football player's value, and to determine the most effective selection process for identifying valuable players. The thesis also explores the potential benefits of using AI in quality control and assurance, highlighting the advantages of creating a closed-loop feedback system to improve processes and product quality. Additionally, the thesis provides valuable insights into the steps that can turn this thesis into a successfully established foundation program in the sports industry, including discussions on HR, funding programs, and marketing strategies. The following tasks were carried out to achieve this objective:

1. Data Collection: The study began with the collection of data concerning Turkish football players in 2017. This included information about the players' professional attributes, such as Aggression, Jumping Reach, Tendency to Punch , Natural Fitness, Vision, Long Throws, Long Shots, Off the Ball, Tackling, Technique, Teamwork, Composure, Free Kick Taking, Reflexes, Positioning, Penalty Taking, Passing, Flair, Anticipation, Crossing, Marking, Leadership, Corners, Concentration, Determination, Decisions, Heading, First Touch, Communication, Acceleration, Pace, Aerial Ability, Strength, Throwing, Handling, Eccentricity, Dribbling, Balance, Kicking, Stamina, Agility, Work Rate, Bravery, Command of Area, Finishing, One on Ones, Tendency to Rush Out, ...

2. Attribute Ranking: The next step involved analysing the data to rank the attributes that have the most significant impact on a player's value. Through the use of an object-attribute-matrix (OAM), the 46 professional attributes were analysed, with a focus on the 5 attributes that had the most significant influence on player value: Composure, Passing, Anticipation, Decisions, and Stamina.

3. Player Selection: After ranking the attributes, a selection process was used to identify valuable players. The study used a positive curve for the player values between 2017-2022, with further filtering under the age of 23, resulting in 28 desired players out of 129.

4. Attribute Valuation: The study then set the values of the top 5 attributes (Composure, Passing, Anticipation, Decisions, and Pace) to 9, 10, 11, and 12, respectively, to evaluate the resulting player values. This was done using a simple filtering method where all the attributes were set to equal or more than 9.

5. Alternative Approaches: Two alternative approaches were also used in the study to evaluate the results. In the first approach, each of the top 5 attributes' modes was calculated, and the total was set as the marking selection line. This resulted in 36 players earning a total of 472 479,32 euros/year. In the second approach, Stamina was replaced by a random positive collating attribute (Pace - rank 15), resulting in 20 players earning 1 276 149,24 euros/year, 14 players earning 1 957 090,37 euros/year, and 3 players earning 1 252 899,28 euros/year.

6. Regression and Robot-eye Methods (similarity analysis): Finally, the study used a regression method and a robot-eye method to compare the results of the attribute valuations. The regression method resulted in total earnings of 1 957 744,74 euros/year, while the robot-eye method resulted in total earnings of 2 393 055,26 euros/year.

7. Total Quality Management (TQM) is a customer-focused management approach that emphasizes continuous improvement, employee involvement, and evidence-based decision making. TQM is a comprehensive system that involves all employees in the pursuit of quality and has spread to organizations worldwide since its origin in Japan in the 1950s. The benefits of TQM include improved customer satisfaction, increased employee engagement, and better financial performance, but it requires continuous improvement and adaptation to changing customer needs and market conditions. This thesis explores the potential benefits of using AI in quality control and assurance, highlighting the advantages of creating a closed-loop feedback system to continually improve processes and ultimately improve product quality, customer satisfaction, and overall business performance.

8. In the thesis, there is a detailed description of the path to becoming a foundation program or startup. This includes discussions on human resources (HR), funding programs with a win-win scenario for both investors and players, and marketing strategies. By addressing these key areas, the thesis offers valuable insights into the steps that organizations can take to successfully establish themselves as foundation programs in the sports industry.

## 1.3. Motivations

As a sports enthusiast and amateur football player the possible development of the world's most well-known and popular sport using the knowledge gained from this study is a motivation for me. This includes helping young players to reach their full potential and to gain a slice of the indescribably large multimillion-dollar industry through program development. Furthermore, the study can assist teams and clubs in identifying the most valuable players and the attributes that contribute the most to their worth. The use of an object-attribute-matrix approach can provide a more systematic and comprehensive analysis, enabling a better understanding of the relationship between a player's attributes and their value. Ultimately, the

findings of this study can inform decision-making processes in the football industry, making it a valuable contribution to the field.

## 1.4. Target audience

The target audience for this project includes a range of individuals and organizations involved in the football industry. Football clubs looking to build a competitive team and maximize player value could benefit from the insights and findings of this study. Player scouting organizations and individual scouts could use the results to inform their player evaluations and identify top prospects.

Television channels and broadcasting platforms could also find this study valuable in understanding the factors that contribute to player value and building their coverage and commentary around these insights. Journalists covering football could use the findings to inform their reporting and provide deeper analysis of player performance.

In addition, startups and program developers in the sports analytics space could use the OAM methodology and the results of this study as a framework for developing their own player valuation models and algorithms. Overall, the target audience for this project includes anyone interested in gaining a deeper understanding of the factors that contribute to the value of football players and developing strategies for building competitive teams and maximizing player value.

## 1.5. Utility

The utility of this study lies in its potential to shape player scouting perspectives and provide opportunities for young and often overlooked players to prove themselves through data analysis. With the use of OAM methodology, players who may have been initially rejected can now be given a chance to demonstrate their worth on the field, as quiet players who produce less shots but achieve higher accuracy can now be identified as potentially valuable players for their team. This study opens new possibilities for both scouts and young players, ultimately contributing to the development of the football industry and the growth of competitive teams.

## 1.6. Structure of the Thesis

The idea behind this thesis began as a research project for a congress, but it has since evolved into the idea of a start-up project. As such, it was important to consciously consider the

formatting and structure of the thesis in order to ensure that it effectively communicates the ideas and goals of the project to the reader.

One of the primary goals of the thesis was to provide a clear and concise explanation of the research, including the motivations behind it and the methods used to carry it out. To this end, the thesis was organized into distinct sections that follow a logical progression, with each section building upon the ideas presented in the previous one.

In some cases, certain information may appear more than once in the thesis. This is done for various reasons such as emphasis, reminder, clarification, and recap to facilitate easier understanding and comprehension by the reader. It is important to note that repeating information is not meant to be redundant, but rather to reinforce key concepts and ensure that the reader is able to fully grasp the ideas presented in the thesis.

For example, a particular message or concept may be presented in one section of the thesis but is then repeated in a later section to remind the reader of its importance or to clarify its meaning in a new context. Similarly, important findings or conclusions may be recapped at the end of a section or chapter to help the reader retain the most important information.

By carefully considering the structure and formatting of the thesis, including the deliberate use of repetition and emphasis, it is our hope that readers will come away with a clear and thorough understanding of the research and ideas presented within it.

In terms of formatting, the thesis uses a consistent style throughout, with clear headings and subheadings to help guide the reader through the text. Additionally, careful consideration was given to the use of visuals, including color-coded charts and graphs, to help illustrate complex ideas and data in a more accessible and engaging way. It should be noted that due to the level of detail required in some of the figures and tables, the digital version of this thesis is ideal for reading. The ability to zoom in and out enables the reader to fully appreciate the level of detail and intricacy of the data presented, which is essential for reproducing the findings of the study.

Overall, the structure and formatting of the thesis were chosen deliberately in order to best communicate the research and ideas presented within it, and to make it as accessible and engaging as possible for the reader.

# 2. Relevant literature

This chapter presents following subchapters:

- History of the problem/phenomenon
- The dataset of problem/phenomenon
- Methodology for the problem/phenomenon
- Potential solution alternatives

By addressing these four subchapters above, you can provide a comprehensive review of the relevant literature on the problem or phenomenon of interest. This will help to establish the context for the own research and demonstrate the importance of the study in advancing the understanding of the topic.

## 2.1. History of the problem/phenomenon

Following relevant components will be handled from the quasi unlimited set of data analysis projects:

*"The Moneyball" phenomenon*

*"The Numbers Game" phenomenon*

*"The Robot-Coach" phenomenon*

In the late 1990s, the Oakland Athletics baseball team faced significant challenges in competing against more financially endowed teams in the league. To address this, the team's general manager, Billy Beane, began exploring alternative player evaluation methods that would enable them to build a winning team despite their limited financial resources. He turned to sabermetrics, a field that uses statistical analysis to measure and evaluate player performance and leveraged data on player performance to identify undervalued players. Through this approach, Beane was able to build a highly competitive team with only a fraction of the budget of other teams. As Brad Pitt, who portrayed Beane in the film adaptation of the team's story, noted, *"It's an unfair game. The rich win, the poor lose. And we've got to change the game."* (Lewis, 2003)

The following chapters will delve more deeply into the Moneyball phenomenon and its impact on the world of baseball.

*"The Numbers Game"* by Chris Anderson and David Sally, a book published in 2013, aimed to disrupt traditional thinking about soccer and the metrics used to evaluate player and team performance. The authors posited that many commonly held beliefs about soccer were rooted in outdated thinking and anecdotal evidence, rather than rigorous analysis. To support their claims, Anderson and Sally drew upon a range of data-driven insights and case studies, challenging readers to think critically about the game they thought they knew.

In her 2021 study, *Miléna Gergics* aims to present in a clear and understandable way what performance data can be analysed by a robot coach and what conclusions can be drawn from them, using football as a basis. The robot coach is able to determine the extent to which one attribute (such as market value, matches, goals, own goals, yellow cards, second yellow cards, red cards, substitutions) has an effect on any other attribute (such as market value) based on the knowledge of various player attributes. This allows for a better understanding of where individual players need to improve and how they can increase their value. For instance, the chapter titled *"The Effects of Attributes on Player Value"* provides a detailed analysis of how different attributes affect the value of a player. Additionally, based on factors such as position, yellow/red cards, goals, starting/reserve/substitute status, the robot coach can estimate how successful or unsuccessful a player is and identify areas where they need to improve in order to earn a spot in the starting line-up. The study also explores the relationship between factors such as the number of penalties and goals, which the robot coach can uncover to provide further insight into player performance, as shown in the chapter titled *"Factors Affecting Goal-Scoring Ability."*

## 2.2. The dataset of problem/phenomenon

According to Michael Lewis (2003), the dataset used in *"Moneyball"* was primarily focused on baseball statistics, such as batting average, on-base percentage, slugging percentage, and other metrics that were used to evaluate player performance. Beane and his team relied heavily on historical data to identify patterns and trends in player performance, which they used to make decisions about which players to acquire and retain. The success of the Oakland Athletics under Beane's leadership has since inspired other teams to adopt similar approaches, and sabermetrics has become a widely accepted practice in the baseball industry.

The authors of *"The Numbers Game"* relied on a diverse range of data sources to support their arguments about soccer metrics. In addition to using detailed match statistics, player performance data, and team records, the authors also conducted interviews with soccer experts and coaches. By drawing on a broad range of data sources and expertise, the authors were able to provide a comprehensive and nuanced analysis of soccer performance metrics that challenged traditional thinking in the field.

The following subsections outline the different data assets utilized in the *"Robot-coach*" study. The personalized diagnosis and therapy section focuses on the analysis of the 10 best soccer players' performance data, including market value, number of matches, goals, own goals, yellow cards, second yellow cards, red cards, and substitutions. These data were obtained from the Transfermarkt website. The section on goal-scoring ability and its influencing factors examines players' age, number of goals and penalty cards, and their starting, substitution, and bench data, collected from the Hungarian Football Federation's (MLSZ) website, as well as personal sources for players' positions. The data sources for this study include the MLSZ website and the Transfermarkt website, with the latter downloaded on April 6, 2021. Additionally, some of the data were obtained from personal sources. The URLs used for data download were retrieved from the MLSZ website on November 23, 2020, and from the Transfermarkt website on April 6, 2021.

## 2.3. Methodology for the problem/phenomenon

According to a source cited in *"The Guardian", (Ryan Baldi, 2022)"* the methodology used in *"Moneyball"* was primarily based on statistical analysis and data mining techniques. Beane and his team used advanced statistical models to identify undervalued players and predict their future performance. They also used machine learning algorithms to identify patterns in player performance data and predict which players would be the most valuable in the future. The article further notes that the focus of the methodology used in *"Moneyball"* was on leveraging data to gain a competitive advantage in player evaluation and team building.

*"The Numbers Game"* by Chris Anderson and David Sally, published in 2013, aimed to challenge traditional thinking about soccer and the metrics used to evaluate player and team performance. Drawing on a wide range of soccer data, including detailed match statistics, player performance data, and team records, the authors used a range of statistical techniques, such as regression analysis and Monte Carlo simulations, to analyse the data. They also developed new

metrics for evaluating player and team performance, such as expected goals and expected points. Through their analysis, the authors sought to challenge many commonly held beliefs about soccer and highlight the importance of data-driven analysis in the sport. Additionally, the book features interviews with soccer experts and coaches to gain additional insights into the sport and its metrics.

The methodology used by *Miléna Gergics* in her study is comparative research (CBR = cases-based reasoning), which includes a novel online element of similarity analysis. This choice was made due to the availability of online support, which is fast, easy to use, optimized, and automatable. Similarity analysis is a self-driving process of producing human intuition by a computer, and it can generate results much faster than manually calculated ones. By providing various indicators of players, the automated online similarity analysis produced results that promote the development of players, teams, and coaches. The study's basic concepts of similarity analysis include COCO (= Component based comparison for objectivity), an algorithmic family that performs similarity analysis, (including object-) attribute orientation, a measure of an indicator with a unit of measurement, benchmarking, a comparison that primarily represents a mindset, COCO-STD (STD=standard model/production function generator), which shows a real Y variable (in this case, the market price) that is built as a step function of X variables (e.g., the better the performance-X, the higher the market price-Y), COCO_Y0 (Y0=ideal searcher and/or optimized antidiscrimination model), which searches for the most deviating object from the average for each X with the direction (direct/inverse) that leads to the ideal state, and similarity analysis, which converts influencing factors into rankings (steps, serial numbers) in a totally context-free way.

## 2.4. Potential solution alternatives

There are several potential solution alternatives that could be explored to address the problem of accurately assessing a football player's value and potential. Here are a few examples:

While player profiling has not been fully automated, according to a report by Forbes (Zak Garner-Purkis, 2020), Liverpool FC has implemented a data-driven approach to player recruitment, using analytics to identify undervalued players and build a team that fits their specific style of play. The club has also partnered with the technology firm Apptio to enhance their data analytics capabilities and gain a competitive edge in player evaluation.

Similarly, Manchester City FC (Training Ground Guru, 2022) has invested heavily in data analytics and machine learning algorithms to track player performance and evaluate potential transfers. The club has reportedly built a complex database of player attributes and performance metrics, allowing them to make more informed decisions about player recruitment and optimize team selection.

In terms of scouts, one example is Vahe Tanielian (Scisports, 2021) who is Director of Data Analytics at Major League Soccer's Real Salt Lake City. Tanielian has a background in data analytics and has implemented a data-driven approach to player evaluation, utilizing video analysis software and other tools to identify potential targets for the club.

Despite not being fully automated, several companies are using computer vision and machine learning algorithms to collect data and track the location and speed of every movement during a match, in order to identify subtle patterns and tendencies that are not easily detectable. One such company is Opta Sports, which has been providing data to professional football clubs since 1996. Opta Sports collects a wide range of data points, including the number of passes, shots, tackles, and interceptions made by each player, as well as more advanced metrics such as expected goals and expected assists.

Another company utilizing data analytics in football is Prozone Sports, which provides performance analysis tools to clubs and coaches. Their software allows coaches to track player movements on the pitch and analyse individual and team performance metrics. The company has been working with football clubs since the early 2000s, and has since been acquired by Stats Llc., a leading sports data and technology company.

Other notable companies include Football Radar (2010), Wyscout (2004), and StatsBomb (2016), all of which provide data analytics services to football clubs and organizations. These companies have contributed to a growing trend in the use of data analytics in football, as clubs seek to gain a competitive edge in player recruitment and performance analysis. Despite the increasing reliance on data analytics, it is worth noting that human scouting and evaluation remains an important aspect of player recruitment and development in football.

Although collaboration between analysts and coaches has not been fully automated, some sports analytics companies are providing coaches with data insights that aid in their training plans, tactics, and player selection. Additionally, analysts are continuously refining their models by incorporating feedback from coaches to improve accuracy.

- *"How Data and Analytics Are Changing the Sports Industry"* by Forbes (Abhas Ricky, 2019): This article highlights how coaches are increasingly relying on data insights to make decisions and plan training sessions. It also mentions how some analytics companies, such as Catapult Sports and Second Spectrum, are providing coaches with data insights.

- *"How Big Data is Changing the Way Football is Played"* by Information Age (Nick Ismail, 2017): This article discusses how football clubs are using data analytics to improve their tactics, training, and player selection. It cites examples of clubs, such as Chelsea and Leicester City, using analytics to gain a competitive edge.

While these sources don't provide specific examples of collaborations between analysts and coaches, they do suggest that such collaborations are happening in the sports industry.

# 3. Data and methods

The next chapter presents the "Data and Methods" section, which showcases the data and methodology used in the research. In the "Own dataset" subsection, a detailed overview of the dataset used in the study is provided. In the "Own methodology" section, the methodology used to evaluate player values based on the dataset is described. The "Similarity Analysis" subsection demonstrates the similarity analysis, which was used to compare player performances in the dataset. In the "Benchmark results" subsection, the findings are presented on the basis of which the best players were selected. The "Experiments and Hypotheses" section introduces the experiments conducted and the hypotheses tested. Finally, in the "Outcome" subsection, the results of the research are summarized.

## 3.1. Own dataset

In this study, the own dataset was created by collecting information on Turkish football players from the dataset acquired from the SI Games' Football Manager series.

The dataset consists of both objective and subjective player attributes, including performance statistics, physical attributes, and positional information. Performance statistics such as goals, assists, and minutes played were collected for each player for the 2017-2022 seasons. Physical attributes, such as height and weight, were also included to provide insight into the players' physical capabilities. Subjective attributes such as market value were also included in the dataset.

o ensure the accuracy and completeness of the dataset, a rigorous data cleaning and validation process was employed, which involved cross-checking the data. For instance, the whole dataset of the year 2017 [50,000+ players] was filtered for nationality, age, etc. until an acceptable pool of players (circa 130) was obtained. MS Excel VLOOKUP function was also used to check the players' unique ID for occurrence year by year.

The resulting dataset and methodology provide a comprehensive and robust approach to analysing player value in Turkish football. The findings are expected to be valuable for a range of stakeholders in the football industry.

## 3.2. Similarity Analysis

Similarity analysis (SA) is a powerful tool for transforming biological intuition into source codes, enabling a mathematical interpretation of sustainability at the deepest operationalization level possible. As a result of artificial intelligence research focused on biological objects within individual and emergence-oriented frameworks, SA delivers a self-monitoring control and consistency-based simulation. SA lays the foundation for a new generation of robots that prioritize optimum energy consumption for individuals or robot teams, with information technology support.

My-X system has already undergone several partial automation and scaling steps, and the next phase of development is the creation of sophisticated control processes around available modules, potentially within the frame of grid-systems. For example, a targeted robot-

vet/breeder could automatically explore disease-suspicions based on individual-specific data of the supervised biological systems, simulate expected effects of therapy variants, forecast expected yields, simulate inter-individual interaction, strategically and operationally define breeding goals, identify paired individuals, and evaluate keeping technology alternatives.

The self-monitoring learning process of SA is sustainability and consistency-oriented, ensuring that it does not lead to arbitrary decisions. In the event that the robot can only decide by drawing among the recognized alternatives, the involvement of a human expert is possible. The example of a robot breeder outlined above can be extended to decisions in crop production, horticulture, plantation or fishing, as well as wildlife management and forestry.

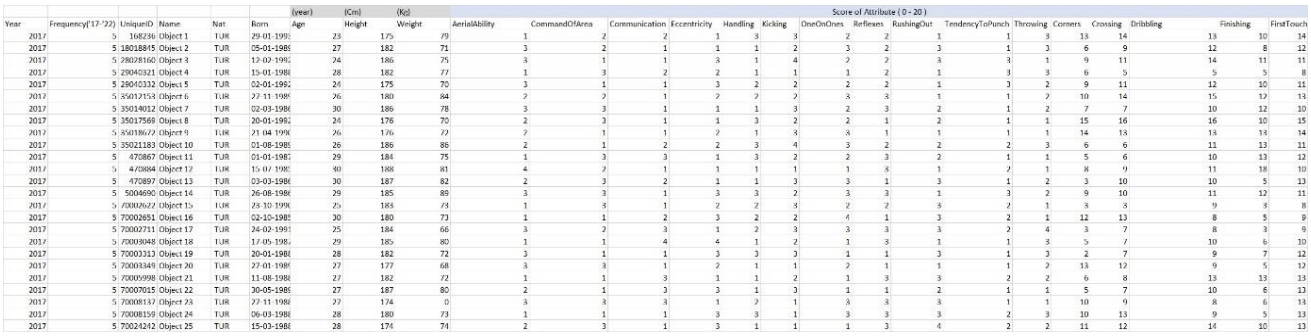
In addition to its applications in biology and agriculture, similarity analysis is also being utilized in various other fields. For instance, it is capable of scouting footballer talents by analysing player performance data and identifying potential future stars. This highlights the versatility and potential of similarity analysis to be applied in a range of industries and domains.

source: "https://miau.my-x.hu/miau/196/My-X%20Team_A5%20fuzet_EN_jav.pdf" (My-X research team, 2004)

## 3.3. Own metholodgy

In this study, the Object-Attribute-Matrix (OAM) methodology has been utilized to analyse the performance of Turkish football players. The OAM is a widely used analytical tool in sports analytics, allowing for the comparison of individual player performances based on a set of attributes relevant to the sport. Similar method has been used in the already explained Robot-Coach method by *Miléna Gergics*.

The OAM model was developed by collecting data on a large sample of Turkish football players, including their game statistics, personal characteristics, and team performance metrics. A set of relevant attributes was identified based on the knowledge of the sport, previous research, and consultation with experts in the field. (Fig. 1.)



| | | | | | | (year) | (Cm) | (Kg) | Score of Attribute ( 0 - 20 ) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | Frequency('17-'22) | UniqueID | Name | Nat | Born | Age | Height | Weight | AerialAbility | CommandOfArea | Communication | Eccentricity | Handling | Kicking | OneOnOnes | Reflexes | RushingOut | TendencyToPunch | Throwing | Corners | Crossing | Dribbling | Finishing | FirstTouch |
| 2017 | 5 | 168236 | Object 1 | TUR | 29-01-199 | 23 | 175 | 79 | 1 | 2 | 2 | 1 | 3 | 3 | 2 | 2 | 1 | 1 | 3 | 13 | 14 | 13 | 10 | 14 |
| 2017 | 5 | 18018815 | Object 2 | TUR | 05-01-198 | 27 | 182 | 71 | 3 | 2 | 1 | 1 | 1 | 2 | 3 | 2 | 3 | 1 | 3 | 6 | 9 | 12 | 8 | 12 |
| 2017 | 5 | 28028160 | Object 3 | TUR | 13-02-199 | 24 | 186 | 75 | 3 | 1 | 1 | 3 | 1 | 4 | 2 | 2 | 3 | 3 | 1 | 9 | 11 | 14 | 11 | 11 |
| 2017 | 5 | 29040321 | Object 4 | TUR | 15-01-198 | 28 | 182 | 77 | 1 | 3 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 3 | 3 | 6 | 5 | 5 | 5 | 8 |
| 2017 | 5 | 29040332 | Object 5 | TUR | 02-01-199 | 24 | 175 | 70 | 3 | 1 | 1 | 3 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 9 | 11 | 12 | 10 | 11 |
| 2017 | 5 | 35012153 | Object 6 | TUR | 27-11-198 | 26 | 180 | 84 | 2 | 2 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 | 2 | 10 | 14 | 15 | 12 | 13 |
| 2017 | 5 | 35014012 | Object 7 | TUR | 02-03-198 | 30 | 186 | 78 | 3 | 3 | 1 | 1 | 1 | 3 | 2 | 3 | 2 | 1 | 2 | 7 | 7 | 10 | 12 | 10 |
| 2017 | 5 | 35017569 | Object 8 | TUR | 20-01-199 | 24 | 176 | 70 | 2 | 3 | 1 | 1 | 3 | 2 | 2 | 1 | 2 | 1 | 1 | 15 | 16 | 16 | 10 | 15 |
| 2017 | 5 | 35018677 | Object 9 | TUR | 21-04-199 | 26 | 176 | 77 | 2 | 1 | 1 | 2 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 14 | 13 | 13 | 13 | 14 |
| 2017 | 5 | 35021183 | Object 10 | TUR | 01-08-198 | 26 | 186 | 86 | 2 | 1 | 2 | 2 | 3 | 4 | 3 | 2 | 2 | 2 | 3 | 6 | 6 | 11 | 13 | 11 |
| 2017 | 5 | 470867 | Object 11 | TUR | 01-01-198 | 29 | 184 | 75 | 1 | 3 | 3 | 1 | 3 | 2 | 2 | 3 | 2 | 1 | 1 | 5 | 6 | 10 | 13 | 12 |
| 2017 | 5 | 470884 | Object 12 | TUR | 15-07-198 | 30 | 188 | 81 | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 2 | 1 | 8 | 9 | 11 | 18 | 10 |
| 2017 | 5 | 470897 | Object 13 | TUR | 03-03-198 | 30 | 187 | 82 | 2 | 3 | 2 | 1 | 1 | 3 | 3 | 1 | 3 | 1 | 2 | 3 | 10 | 10 | 5 | 13 |
| 2017 | 5 | 5004690 | Object 14 | TUR | 26-08-198 | 29 | 185 | 89 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 3 | 1 | 3 | 2 | 9 | 10 | 11 | 12 | 11 |
| 2017 | 5 | 70002622 | Object 15 | TUR | 28-10-199 | 25 | 183 | 73 | 1 | 3 | 1 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 1 | 3 | 3 | 9 | 3 | 8 |
| 2017 | 5 | 70002651 | Object 16 | TUR | 02-10-198 | 30 | 180 | 73 | 1 | 1 | 2 | 3 | 2 | 2 | 1 | 1 | 3 | 2 | 1 | 12 | 13 | 8 | 5 | 9 |
| 2017 | 5 | 70002711 | Object 17 | TUR | 24-02-199 | 25 | 184 | 66 | 3 | 2 | 3 | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 4 | 3 | 7 | 8 | 3 | 9 |
| 2017 | 5 | 70003048 | Object 18 | TUR | 17-05-198 | 29 | 185 | 80 | 1 | 1 | 4 | 4 | 1 | 2 | 1 | 3 | 1 | 1 | 3 | 5 | 7 | 10 | 6 | 10 |
| 2017 | 5 | 70003313 | Object 19 | TUR | 20-01-198 | 28 | 182 | 72 | 3 | 1 | 1 | 3 | 3 | 3 | 1 | 1 | 3 | 1 | 3 | 2 | 7 | 9 | 7 | 12 |
| 2017 | 5 | 70003349 | Object 20 | TUR | 27-01-198 | 27 | 177 | 68 | 3 | 3 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 13 | 12 | 9 | 5 | 12 |
| 2017 | 5 | 70005998 | Object 21 | TUR | 11-08-198 | 27 | 182 | 72 | 1 | 1 | 3 | 1 | 1 | 2 | 1 | 3 | 3 | 2 | 2 | 6 | 8 | 13 | 13 | 13 |
| 2017 | 5 | 70007015 | Object 22 | TUR | 30-05-198 | 27 | 187 | 80 | 2 | 1 | 3 | 3 | 1 | 3 | 1 | 1 | 2 | 1 | 1 | 5 | 7 | 10 | 6 | 13 |
| 2017 | 5 | 70008137 | Object 23 | TUR | 27-11-198 | 27 | 174 | 0 | 3 | 3 | 3 | 1 | 2 | 1 | 3 | 3 | 3 | 1 | 1 | 10 | 9 | 8 | 6 | 13 |
| 2017 | 5 | 70008159 | Object 24 | TUR | 06-03-198 | 28 | 180 | 73 | 1 | 1 | 1 | 3 | 3 | 1 | 3 | 3 | 3 | 2 | 3 | 10 | 13 | 9 | 5 | 13 |
| 2017 | 5 | 70024242 | Object 25 | TUR | 15-03-198 | 28 | 174 | 74 | 2 | 3 | 1 | 3 | 1 | 1 | 1 | 3 | 4 | 2 | 2 | 11 | 12 | 14 | 10 | 13 |

1. Figure: Pool of Turkish players

source: own presentation / https://miau.my-x.hu/miau/297/starlight_v1.xls (Sheet = Sheet 1 Range = A89:Y115)

A combination of statistical analysis and expert judgment was used to assign numerical values to each player for each attribute. The values were organized into a matrix format, where each row represented a player object and each column represented an attribute. This matrix enabled easy player comparisons based on their attribute scores and overall performance using similarity analyses.

Attributes were filtered out based on the correlation analysis of player values (Fig. 2.), and regression analysis was applied using the players' market value and attributes. This allowed the determination of the 5 attributes out of 46 that have the greatest influence on a player's market value each year. The 5 key attributes were ranked based on this filtering to demonstrate the correlation between attributes and player value (Fig. 3.). This was used as a benchmark to find players whose value increases over time, making them a promising talent (Fig. 4.). In one experiment, the top 5 attributes were replaced with a random one, resulting in higher hit rates than the 5 highest value attributes, demonstrating that the benchmark could be surpassed (Fig. 5., 6.)

| Units | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation | Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Aggression | Jumping Reach | Tendency to Punch | Natural Fitness | Vision | Long Throws | Long Shots | Off the Ball | Tackling | Technique | Teamwork | Composure | Free Kick Taking | Reflexes | Positioning | Penalty Taking | Passing |
| 2019 | 0.13 | 0.20 | -0.07 | 0.12 | 0.31 | 0.15 | 0.25 | 0.18 | 0.12 | 0.18 | 0.26 | 0.38 | 0.25 | -0.02 | 0.13 | 0.33 | 0.30 |
| 2020 | 0.09 | 0.21 | -0.09 | 0.12 | 0.22 | 0.16 | 0.18 | 0.06 | 0.22 | 0.17 | 0.23 | 0.38 | 0.16 | 0.00 | 0.22 | 0.18 | 0.31 |
| 2021 | 0.07 | 0.17 | 0.00 | 0.12 | 0.18 | 0.15 | 0.16 | 0.04 | 0.20 | 0.14 | 0.19 | 0.37 | 0.11 | -0.02 | 0.18 | 0.16 | 0.28 |
| 2022 | 0.06 | 0.21 | 0.01 | 0.11 | 0.16 | 0.06 | 0.01 | -0.05 | 0.18 | 0.12 | 0.15 | 0.18 | 0.03 | 0.04 | 0.23 | 0.02 | 0.18 |
| | | | | | | | | | | | | | | | | | |
| Average | 0.09 | 0.20 | -0.04 | 0.12 | 0.22 | 0.13 | 0.15 | 0.06 | 0.18 | 0.15 | 0.21 | 0.33 | 0.14 | 0.00 | 0.19 | 0.17 | 0.27 |
| Rank | 35 | 16 | 46 | 30 | 12 | 28 | 25 | 36 | 21 | 24 | 13 | 1 | 26 | 38 | 20 | 22 | 5 |

2. Figure: Correlation between attributes and market value

source: own presentation / https://miau.my-x.hu/miau/297/starlight_v1.xls (Sheet = Sheet 1 Range = A1:R9)

| Units | Correlation | Correlation | Correlation | Correlation | Correlation |
|---|---|---|---|---|---|
| Year | Composure | Stamina | Anticipation | Decisions | Passing |
| 2019 | 0.38 | 0.39 | 0.39 | 0.41 | 0.30 |
| 2020 | 0.38 | 0.33 | 0.36 | 0.37 | 0.31 |
| 2021 | 0.37 | 0.34 | 0.30 | 0.36 | 0.28 |
| 2022 | 0.18 | 0.26 | 0.22 | 0.12 | 0.18 |
| | | | | | |
| Average | 0.33 | 0.33 | 0.32 | 0.32 | 0.27 |
| Rank | 1 | 2 | 3 | 4 | 5 |

3. Figure: Top 5 Attributes

source: own presentation / https://miau.my-x.hu/miau/297/starlight_v1.xls (Sheet = Sheet 1 Range = A12:F20)

| | | | Market value (Eur) | | | | | | Market value change (Eur/year) |
|---|---|---|---|---|---|---|---|---|---|
| UID | Name | Age (Year) | 2017 | 2019 | 2020 | 2021 | 2022 | Min v Max (M.Value) | Slope |
| 70052880 | Object 1 | 22 | 40000 | 66622 | 140000 | 170000 | 550000 | 510000 | 86263 |
| 70056719 | Object 2 | 22 | 32500 | 16938 | 54000 | 55000 | 47000 | 14500 | 5111 |
| 70057965 | Object 3 | 21 | 28000 | 27100 | 115000 | 55000 | 49500 | 21500 | 6609 |
| 70062993 | Object 4 | 21 | 105000 | 338760 | 1200000 | 1800000 | 3500000 | 3395000 | 644256 |
| 70063195 | Object 5 | 20 | 13000 | 1100970 | 1500000 | 1500000 | 880000 | 867000 | 210731 |
| 70063197 | Object 6 | 21 | 19000 | 180672 | 650000 | 675000 | 1175000 | 1156000 | 224815 |
| 70065895 | Object 7 | 20 | 67000 | 191964 | 750000 | 600000 | 127000 | 60000 | 54610 |
| 70076282 | Object 8 | 21 | 2900000 | 1242120 | 6750000 | 8500000 | 6700000 | 3800000 | 1160561 |
| 70085928 | Object 9 | 19 | 150000 | 1919640 | 750000 | 250000 | 10100000 | 9950000 | 1399614 |
| 70088391 | Object 10 | 22 | 8000 | 13268 | 220000 | 195000 | 138000 | 130000 | 37066 |
| 70091317 | Object 11 | 20 | 18000 | 42345 | 56000 | 14750 | 115000 | 97000 | 13352 |
| 70092388 | Object 12 | 20 | 250000 | 536370 | 550000 | 1700000 | 2037500 | 1787500 | 371851 |
| 70093800 | Object 13 | 20 | 3600000 | 14115000 | 35000000 | 32500000 | 77000000 | 73400000 | 13110000 |
| 70093829 | Object 14 | 18 | 14500 | 49120 | 48500 | 10750 | 62500 | 48000 | 5419 |
| 70097242 | Object 15 | 19 | 24000 | 23713 | 21000 | 60000 | 60000 | 36000 | 8245 |
| 70097248 | Object 16 | 19 | 12000 | 1467960 | 7750000 | 10250000 | 10900000 | 10888000 | 2474461 |
| 70097285 | Object 17 | 21 | 19000 | 22584 | 26000 | 15000 | 33000 | 14000 | 1657 |
| 70097801 | Object 18 | 19 | 22500 | 1580880 | 2200000 | 1300000 | 2250000 | 2227500 | 379884 |
| 70099180 | Object 19 | 20 | 15000 | 48555 | 650000 | 300000 | 220000 | 205000 | 60348 |
| 70099745 | Object 20 | 19 | 100000 | 82431 | 81000 | 91000 | 147500 | 47500 | 7023 |
| 70103402 | Object 21 | 19 | 6000 | 10162 | 27000 | 18500 | 25000 | 19000 | 3896 |
| 70106700 | Object 22 | 19 | 6250 | 74527 | 100000 | 38000 | 27500 | 21250 | 3309 |
| 91100534 | Object 23 | 21 | 2900000 | 4065120 | 7750000 | 5750000 | 3675000 | 775000 | 348844 |
| 91104583 | Object 24 | 22 | 8000000 | 20325600 | 22000000 | 18250000 | 20750000 | 12750000 | 2249291 |
| 91107556 | Object 25 | 22 | 68000 | 74527 | 475000 | 600000 | 357500 | 289500 | 91316 |
| 91121536 | Object 26 | 22 | 3200000 | 4065120 | 2700000 | 475000 | 6500000 | 3300000 | 216074 |
| 91126197 | Object 27 | 22 | 100000 | 395220 | 800000 | 1200000 | 1250000 | 1150000 | 253636 |
| 92071861 | Object 28 | 18 | 2700000 | 3274680 | 1600000 | 1700000 | 4500000 | 1800000 | 140557 |

4. Figure: Desired players

source: own presentation / https://miau.my-x.hu/miau/297/starlight_v1.xls (Sheet = Sheet 1 Range = A58:J87)

| Benchmarks | | 1a: top5'_score9 | | | 1b: top5'_score10 | | | 1c: top5'_score11 |
|---|---|---|---|---|---|---|---|---|
| Accuracy rate (%): | | 80% | | | 71% | | | 100% |
| Slope of Market values avg. (eur): | | 1 276 149 | | | 1 957 090 | | | 1 252 899 |
| | UID | Slope (Eur) | | UID | Slope (Eur) | | UID | Slope (Eur) |
| | 70056719 | 5111 | | 70062993 | 644256 | | 70076282 | 1160561 |
| | 70062993 | 644256 | | 70063141 | | | 91100534 | 348844 |
| | 70063141 | | | 70065895 | 54610 | | 91104583 | 2249291 |
| | 70063197 | 224815 | | 70076282 | 1160561 | | | |
| | 70065895 | 54610 | | 70085928 | 1399614 | | | |
| | 70076282 | 1160561 | | 70087488 | | | | |
| | 70085928 | 1399614 | | 70092189 | | | | |
| | 70087488 | | | 70092388 | 371851 | | | |
| | 70091317 | 13352 | | 70093800 | 13110000 | | | |
| | 70092189 | | | 91100534 | 348844 | | | |
| | 70092388 | 371851 | | 91104583 | 2249291 | | | |
| | 70093800 | 13110000 | | 91104803 | | | | |
| | 70097242 | 8245 | | 91107556 | 91316 | | | |
| | 70097801 | 379884 | | 92071861 | 140557 | | | |
| | 91100534 | 348844 | | | | | | |
| | 91104583 | 2249291 | | | | | | |
| | 91104803 | | | | | | | |
| | 91107556 | 91316 | | | | | | |
| | 91121536 | 216074 | | | | | | |
| | 92071861 | 140557 | | | | | | |

5. Figure: Example for human benchmarks

source: own presentation / https://miau.my-x.hu/miau/297/starlight_v1.xls (Sheet = Sheet 1 Range = W22:AE45)

| Accuracy rate (%): | benchmark1a | 80% | | benchmark1b | 71% | | benchmark1c | 100% | | anti-discrimination-optimizing | 53% | | linear regression | 77% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Slope of Market values avg. (eur): | | 1 276 149 | | | 1 957 090 | | | 1 252 899 | | | 2 393 055 | | | 1 957 745 |
| | UID | Slope (Eur) | | UID | Slope (Eur) | | UID | Slope (Eur) | | UID | Slope (Eur) | | UID | Slope (Eur) |
| | 70056719 | 5111 | | 70062993 | 644256 | | 70076282 | 1160561 | | 70062993 | 644256 | | 70062993 | 644256 |
| | 70062993 | 644256 | | 70063141 | | | 91100534 | 348844 | | 70063128 | | | 70065895 | 54610 |
| | 70063141 | | | 70065895 | 54610 | | 91104583 | 2249291 | | 70063141 | | | 70076279 | |
| | 70063197 | 224815 | | 70076282 | 1160561 | | | | | 70076279 | | | 70076282 | 1160561 |
| | 70065895 | 54610 | | 70085928 | 1399614 | | | | | 70076282 | 1160561 | | 70080329 | |
| | 70076282 | 1160561 | | 70087488 | | | | | | 70080329 | | | 70085928 | 1399614 |
| | 70085928 | 1399614 | | 70092189 | | | | | | 70085928 | 1399614 | | 70087488 | |
| | 70087488 | | | 70092388 | 371851 | | | | | 70087488 | | | 70093800 | 13110000 |
| | 70091317 | 13352 | | 70093800 | 13110000 | | | | | 70092189 | | | 91100534 | 348845 |
| | 70092189 | | | 91100534 | 348844 | | | | | 70093800 | 13110000 | | 91104583 | 2249292 |
| | 70092388 | 371851 | | 91104583 | 2249291 | | | | | 70096907 | | | 91121536 | 216075 |
| | 70093800 | 13110000 | | 91104803 | | | | | | 91100534 | 348845 | | 91126197 | 253637 |
| | 70097242 | 8245 | | 91107556 | 91316 | | | | | 91104583 | 2249292 | | 92071861 | 140558 |
| | 70097801 | 379884 | | 92071861 | 140557 | | | | | 91107556 | 91316 | | | |
| | 91100534 | 348844 | | | | | | | | 92071861 | 140558 | | | |
| | 91104583 | 2249291 | | | | | | | | | | | | |
| | 91104803 | | | | | | | | | | | | | |
| | 91107556 | 91316 | | | | | | | | | | | | |
| | 91121536 | 216074 | | | | | | | | | | | | |
| | 92071861 | 140557 | | | | | | | | | | | | |

6. Figure: Human benchmarks vs Robot-eye method

source: own presentation / https://miau.my-x.hu/miau/297/starlight_v1.xls (Sheet = Sheet 1 Range = W50:AK72)

Overall, the OAM-based methodology provides a powerful and flexible tool for analysing football player performance and identifying the key attributes that contribute to success in the sport. By developing and applying this methodology, a better understanding of football

performance can be achieved, and more effective strategies can be developed for building competitive teams and maximizing player value.

## 3.4. Benchmarking Results

The research findings indicate that a team-based approach to player valuation can provide valuable insights into the factors that contribute to player value and team success in football. Using the OAM methodology, the study identified several key attributes that are associated with higher player value and demonstrated the utility of this approach for identifying top prospects and building competitive teams.

The aim of this study was to develop a method for predicting which young Turkish football players are likely to increase their value in the future. A dataset of 160,000 football players was created and filtered to only include Turkish players. After additional filtering, 129 players were analysed based on 46 professional attributes related to their value.

## 3.5. Experiments and Hypotheses

The goal is to develop a methodology that can accurately identify young football players with high potential for development, with the ultimate aim of maximizing player value and team performance. The following experiments has been proposed to test the efficacy of the methodology:

Dataset Creation: The dataset creation process involved filtering through 160 000 players, narrowing it down to only Turkish players and further filtering until 129 were left. Numerous statistical data had to be accessed about these players and focused on their 46 attributes.

Data Analysis: Statistical analysis and machine learning techniques will be used to identify patterns and trends in the data, with a focus on identifying the key factors that contribute to player development and value.

Model Development: Based on the data analysis, a model for predicting player development and future value will be developed, incorporating key performance and physical indicators as well as personal characteristics.

Model Testing: The model will be tested on a separate dataset of young players to evaluate its accuracy in predicting their future development and value.

Implementation: Working with football clubs and scouting organizations can implement the methodology in real-world player evaluation and scouting processes, with the aim of maximizing player value and team performance.

The hypothesis of accurately identifying young football players with high potential for development, investing in and developing them, and increasing their value and generating profit for clubs and organizations can be achieved by using the methodology. The model's comprehensiveness and accuracy in player evaluation and development can also be improved by incorporating personal characteristics and other non-performance factors into the model. The goal is to track players' development and identify future talents using sufficient data. The research also aims to determine if there are five crucial attributes common among outstanding football players and if any trends can be observed in the dataset.

## 4. Outcome

Several experiments were conducted with different methods, including setting the values of the top attributes, calculating the mode of each of the top five attributes, and replacing the top attribute with another positively correlated attribute. The outcomes of the experiments are as follows:

Setting the top attributes to a certain value:

- 66 players with 21.53% accuracy and 313 083,46 euros/year in earnings.
- 39 players with 20.51% and 490 883,1293 euros/year in earnings.
- 15 players with 20% and 250 579,85 euros/year in earnings.
- 4 players with 50% and 377 351,48 euros/year in earnings.

Calculating the mode of each of the top five attributes:

- 36 players with 13.88% and 472 479,32 euros/year in earnings.

Replacing the top attribute with another positively correlated attribute:

- 20 players with 80% and 1 276 149,24 euros/year in earnings.

- 14 players with 71.42% and 1 957 090,37 euros/year in earnings.
- 3 players with 100% and 1 252 899,28 euros/year in earnings.

Experiments were conducted using the robot-eye method and regression method, which resulted in 2 393 055,26 euros/year and 1 957 744,74 euros/year in earnings, respectively.

## 5. Discussion

After filtering for players under the age of 23, 28 players were selected from the original 129, and an experiment was conducted to determine if the values of these players could be predicted based on their attributes. The top 5 attributes were set to values ranging from 9 to 12, and the outcomes were recorded. The experiment showed that by setting the attributes to these values, 66 players had a 21.54% accuracy and earned 313 083,46 euros/year, while 39 players had a 20.51% accuracy and earned 490 883,12 euros/year. Additionally, 15 players had a 20% accuracy and earned 250 579,85 euros/year, and 4 players had a 50% accuracy and earned 377 351,48 euros/year.

Another approach involved calculating the mode of each of the top 5 attributes and adding them up, resulting in a total of 52 points out of 100. This approach resulted in 36 players with 13.89% accuracy and 472 479,32 euros/year in earnings.

A third approach replaced the top attribute (Stamina) with another positively correlated attribute, Pace, resulting in 20 players with 80% accuracy and 1 276 149,24 euros/year in earnings, 14 players with 71.43% accuracy and 1 957 090,37 euros/year in earnings, and 3 players with 100% accuracy and 1 252 899,28 euros/year in earnings.

Finally, the robot-eye method and regression method were used to predict player values, resulting in 2 393 055.26 euros/year and 1 957 744,74 euros/year in earnings, respectively.

The findings suggest that a player's value can indeed be predicted based on certain professional attributes. However, the experiment showed that different approaches can result in varying levels of accuracy, indicating that further research is necessary to determine the most effective method for predicting player values. Future research could involve expanding the

dataset to include players from other countries or conducting a more in-depth analysis of specific attributes to determine their impact on player values.

# 6. Conclusions

The experiments show that it is possible to develop a method to select football players who are likely to increase in value in the future. The analysis showed that the top five attributes that had the most influence on a player's value were Composure, Passing, Anticipation, Decisions, and Stamina. However, the accuracy of the method and the potential earnings varied depending on the specific filtering and selection methods used. Further studies could explore other factors that may influence a player's value and expand the dataset to include players from other countries.

## 6.1. Total Quality Management

Total Quality Management (TQM, source: https://en.wikipedia.org/wiki/Total_quality_management) is a management approach that has been embraced by organizations around the world. TQM is based on a comprehensive system that encompasses all aspects of the organization, from suppliers to customers, and involves all employees in the pursuit of quality.

It is a customer-focused approach that emphasizes continuous improvement, employee involvement, and evidence-based decision making. TQM originated in Japan in the 1950s and has since spread to organizations worldwide. The philosophy of TQM is that quality is not just the responsibility of a quality control department, but a shared responsibility of everyone in the organization. By focusing on improving processes, products, and services to meet or exceed customer expectations, TQM can lead to several benefits for organizations, including improved customer satisfaction, increased employee engagement, and better financial performance. However, TQM is a dynamic and ongoing process that requires continuous improvement and adaptation to changing customer needs and market conditions.

### 6.1.1. Closed-loop feedback system (Robot inspecting its own robot-eye method)

One particularly intriguing aspect of this thesis is the fact that the AI method being used will be checking its own results. (Fig. 7.) This highlights the potential benefits of using AI in quality control and assurance, as it can not only provide more accurate and consistent results, but it can also verify those results independently.

By using an AI system to check its own results, a closed-loop feedback system is essentially created (source: "https://learn.microsoft.com/en-us/archive/msdn-magazine/2019/april/machine-learning-closed-loop-intelligence-a-design-pattern-for-machine-learning" ) that can continually improve itself over time. This could lead to more efficient and effective quality control processes, resulting in improved product quality, customer satisfaction, and overall business performance.

| Direction | 0 | 0 | 0 | 1 | 0 | 0 | |
|---|---|---|---|---|---|---|---|
| | Upper Quartile | Min. | Max. | Spread | Accuracy rate (%): | Slope of Market values avg. (eur): | Time Consumption |
| benchmark1a | 773332 | 5111 | 13110000 | 3215879.981 | 80% | 1276149 | |
| benchmark1b | 1339851 | 54610 | 13110000 | 3981084.815 | 71% | 1957090 | |
| benchmark1c | 1704926 | 348844 | 2249291 | 953582 | 100% | 1252899 | |
| anti-discrimination-optimizing | 1612034 | 91316 | 13110000 | 4391361.355 | 53% | 2393055 | |
| linear regression | 1339851 | 54610 | 13110000 | 3980230.43 | 77% | 1957745 | |
| | | | | | | | |
| | | | | | | | |
| | Ranking (1-5) | Ranking (1-5) | Ranking (1-5) | Ranking (1-5) | Ranking (1-5) | Ranking (1-5) | Ranking (1-5) |
| benchmark1a | 5 | 5 | 1 | 2 | 2 | 4 | 3 |
| benchmark1b | 4 | 4 | 1 | 4 | 4 | 3 | 3 |
| benchmark1c | 1 | 1 | 5 | 1 | 1 | 5 | 3 |
| anti-discrimination-optimizing | 2 | 2 | 1 | 5 | 5 | 1 | 1 |
| linear regression | 3 | 3 | 1 | 3 | 3 | 2 | 1 |

7. Figure Closed-loop feedback system

source: own presentation /https://miau.my-x.hu/miau/299/xR/ (Sheet = Sheet 2 Range = A1:H15)

Overall, this thesis presents a promising exploration of the potential benefits of using AI in quality control and assurance, and the use of a closed-loop feedback system to continually improve these processes is an interesting and innovative approach.

### 6.1.2. Possible Integration of Machine Learning Algorithms

Machine learning algorithms are a type of artificial intelligence that can analyse large volumes of data and identify patterns and trends that can be used to improve quality control processes. By integrating machine learning algorithms with closed-loop feedback systems, organizations can create more intelligent quality control systems that can adapt to changing circumstances and improve over time.

One of the main benefits of integrating machine learning algorithms with closed-loop feedback systems is that it can enable identifying areas of weakness in quality control processes. For example, machine learning algorithms can be used to analyse data from quality control inspections to identify patterns of defects that may indicate problems in the production process. By identifying these patterns, it can take a quick corrective action to improve the production process and reduce the risk of defects in the final product.

Another benefit of integrating machine learning algorithms with closed-loop feedback systems is that it can help organizations to continually refine and improve their quality control processes. By analysing data from quality control inspections and other sources, machine learning algorithms can identify potential opportunities for improvement and provide recommendations for changes to the production process.

Overall, the integration of machine learning algorithms with closed-loop feedback systems has the potential to transform the way organizations approach quality control. By leveraging machine learning technologies, organizations can improve the efficiency and effectiveness of their quality control efforts, leading to improved product consistency, results and customer satisfaction.

## 6.2. Further development of the research

"In the next chapter, "Further Development of the Research," the ongoing progress of the study will be explored. The subsections include "Robot vs. Human – AI vs Human Methods," where the differences between robotic and human approaches to the research are analysed, and "AI-Human Collaboration," where the potential for collaboration between artificial intelligence and human researchers to enhance the quality of the findings is examined. Additionally, in subsection "Advanced Quality Control Techniques Based on Closed-Loop Feedback Systems," advanced quality control techniques will be discussed, highlighting the importance of using closed-loop feedback systems to ensure the accuracy and consistency of the results."

### 6.2.1. Robot vs. Human – AI vs Human Methods

While AI systems have shown great promise in many fields, including quality control and assurance, human judgment and expertise may still be necessary to achieve the highest levels of accuracy and quality control. It is possible that in certain situations, human methods may outperform AI methods in terms of accuracy and efficiency. However, it's worth noting that, AI systems are continually improving, and as they learn from their mistakes and are updated with new data, their performance is likely to improve as well.

Furthermore, while humans may be better at certain tasks, AI systems are often better at handling large amounts of data and performing repetitive tasks without getting tired or making mistakes.

Although AI systems have shown great promise in many fields, including quality control and assurance, they are not infallible. The performance of five different methods was compared in this study: three human benchmarks (1a, 1b, and 1c cf. 3.3.), a classic linear regression model, and an anti-discrimination model. These methods were evaluated based on several metrics, including the upper quartile, minimum, maximum, spread, accuracy rate, and slope of the average market values.

When looking at pure results, the human benchmark 1c achieved the best performance with a 100% accuracy rate, but the lowest earnings of approximately 1 250 000 euros. In contrast, the anti-discrimination model (AI) had a 53% accuracy rate at the time of writing, but nearly 2.4 million euros in earnings, the highest among all methods tested.

When considering time consumption as a factor, it should be noted that the AI model has a significant advantage over human benchmarks. The model was already developed and ready to use, while the benchmarks had to be invented, researched, and tested. This highlights an advantage of using AI in quality control and assurance, as it can potentially save time and resources in the long run.

In addition to comparing the performance of different methods, it is important to note that the use of AI in quality control and assurance has the potential to create a closed-loop feedback system that can continually improve itself over time. This system can not only provide more

accurate and consistent results but also verify those results independently, which is particularly intriguing.

Moreover, the use of AI can reduce the risk of bias and discrimination in the quality control process. For example, the anti-discrimination model used in this study was specifically designed to identify and address potential biases in the data. However, it is important to recognize that the use of AI in quality control and assurance also raises ethical concerns, such as the potential for AI to perpetuate or even amplify existing biases and discrimination. Therefore, it is crucial to ensure that AI systems are designed and implemented in an ethical and responsible manner, with appropriate safeguards and oversight in place.

In summary, the comparison of different methods in this study suggests that while human benchmarks may outperform AI systems in some cases, the use of AI in quality control and assurance has several advantages, including the ability to handle large amounts of data, perform repetitive tasks without getting tired or making mistakes, and save time and resources. The potential for AI to create a closed-loop feedback system that can continually improve itself over time is also a significant advantage. However, it is important to address ethical concerns and ensure that AI systems are designed and implemented in an ethical and responsible manner.

### 6.2.2. AI - Human Collaboration

Furthermore, a collaborative approach can also help address some of the ethical concerns related to the use of AI in quality control and assurance. By involving human workers in the decision-making process, there is a greater likelihood of identifying and addressing potential biases and discrimination.

One example of a successful human-AI collaboration in quality control and assurance is the use of image recognition software in the food industry. AI systems can quickly and accurately identify defects or contaminants in food products, which can help reduce the risk of foodborne illness. However, human workers are still necessary to oversee and make final decisions on quality control and assurance, especially when it comes to more complex or subjective evaluations.

Another example is in the healthcare industry, where AI systems can assist in identifying potential medical errors or discrepancies in patient data. Human workers can then use their clinical expertise to make decisions on how to address these issues.

To ensure successful human-AI collaboration, it is important to establish clear communication and roles between human workers and AI systems. Human workers should be trained on how to effectively use and interpret AI results, while AI systems should be designed with human oversight and feedback in mind.

Moreover, it is important to consider the ethical implications of human-AI collaboration, such as the potential for AI systems to replace human workers or perpetuate existing biases. It is crucial to ensure that AI systems are designed to augment human abilities, not replace them, and that they are developed in an ethical and responsible manner.

In summary, a collaborative approach that leverages the strengths of both human workers and AI systems is likely to result in the most effective quality control and assurance processes. By combining the automation and efficiency of AI with the creativity and decision-making abilities of human workers, more accurate, efficient, and ethical quality control and assurance processes can be created.

### 6.2.3. Advanced Quality Control Techniques Based on Closed-Loop Feedback Systems

Closed-loop feedback systems can also be used to implement advanced quality control techniques that can help organizations to improve the consistency and quality of their products and services. (cf. 6.1.1.) One example of an advanced quality control technique is Statistical Process Control (SPC).

SPC involves using statistical methods to monitor and control the production process in real-time. By continuously monitoring the process and measuring key performance indicators (KPIs), any deviations can be quickly detected from the norm and take corrective action to ensure that the final product meets the desired specifications.

Another advanced quality control technique is Six Sigma, which is a data-driven approach to quality control that focuses on minimizing variation and defects in the production process. Six Sigma uses statistical methods to analyse the production process and identify areas for improvement. This approach can lead to significant improvements in product quality, customer satisfaction, and overall business performance.

Closed-loop feedback systems can also be used to implement Total Quality Management (cf. 6.1.), which is a management approach that seeks to improve the quality of an organization's products and services by involving all employees in a continuous improvement process. By using closed-loop feedback systems to collect data and feedback from employees at all levels of the organization, organizations can identify areas for improvement and implement changes to improve the quality of their products and services.

Overall, closed-loop feedback systems offer a wide range of advanced quality control techniques that can help organizations to improve the consistency and quality of their products and services. By using these techniques, organizations can reduce the risk of errors and defects, improve customer satisfaction, and increase overall business performance.

## 6.3. Path to Foundation Program (Start-up)

In this chapter, the path to establish a foundation program (start-up) for the research is discussed. The "Human Resources (HR)" subsection examines the future of AI from an HR perspective and its impact on the research. The "Funding Program (Win-Win Scenario)" subsection explores how a mutually beneficial arrangement for funders and players can be created during the recruitment process. Finally, the "Marketing" subsection outlines strategies for promoting the program and reaching the target audience.

### 6.3.1. Human Resources (HR)

As AI technology continues to evolve and become more advanced, it is likely that more and more jobs will require workers to collaborate with AI systems.
This may require workers to develop new skills, such as the ability to understand and work with AI systems, as well as the ability to analyse and interpret the data that those systems produce.
In many fields, the use of AI systems may become increasingly commonplace, such as in healthcare (source:"https://nanonets.com/blog/ai-in-healthcare/"), where AI can be used to help diagnose diseases or in manufacturing, where AI can help streamline production processes.

However, it's worth noting that as AI becomes more advanced, there may also be concerns around job displacement and the need for workers to retrain or acquire new skills in order to remain competitive in the job market.

From a scientific perspective, the use of AI in quality control and assurance can help to improve the accuracy and reliability of scientific research and experimentation.

In everyday life, the use of AI in quality control and assurance can lead to higher quality products and services, as well as more efficient and effective processes.

However, there may also be concerns around issues such as privacy, data security, and the potential for bias in AI systems.

As AI technology continues to evolve, it will be important to carefully consider the ethical implications of its use in quality control (cf. 6.3.1.) and assurance, as well as its impact on society as a whole.

The use of AI in quality control and assurance can have significant implications for Human Resources departments.

As AI becomes more advanced and capable, it may be able to take over certain tasks that are currently performed by human workers, such as data analysis or quality control inspections.

This could potentially lead to job displacement, which could have a significant impact on workers and their families.

However, it's important to note that AI can also bring new opportunities and create new jobs, particularly in fields such as data science or AI development.

Human Resources departments may need to adapt their hiring practices and job training programs in order to keep up with the changing demands of the job market.

This may include placing a greater emphasis on skills that are less easily automated, such as creativity, problem-solving, and emotional intelligence.

Additionally, HR departments will need to be mindful of the potential for bias in AI systems, particularly in areas such as hiring and performance evaluation.

It will be important to ensure that AI systems are designed and implemented in a way that is fair and unbiased, and that human workers are trained to use and interpret AI data in a responsible and ethical manner.

Overall, the use of AI in quality control and assurance has the potential to significantly impact the role of Human Resources in organizations, and it will be important for HR departments to stay informed and adapt to these changes to remain competitive and effective.

In the context of scouting and selecting footballers, AI can be used to analyse player data, such as performance statistics and physical attributes, in order to identify high-potential players. This can help HR departments to make more informed decisions about player recruitment and development.

The use of AI in scouting and selecting footballers has the potential to improve the effectiveness and efficiency.

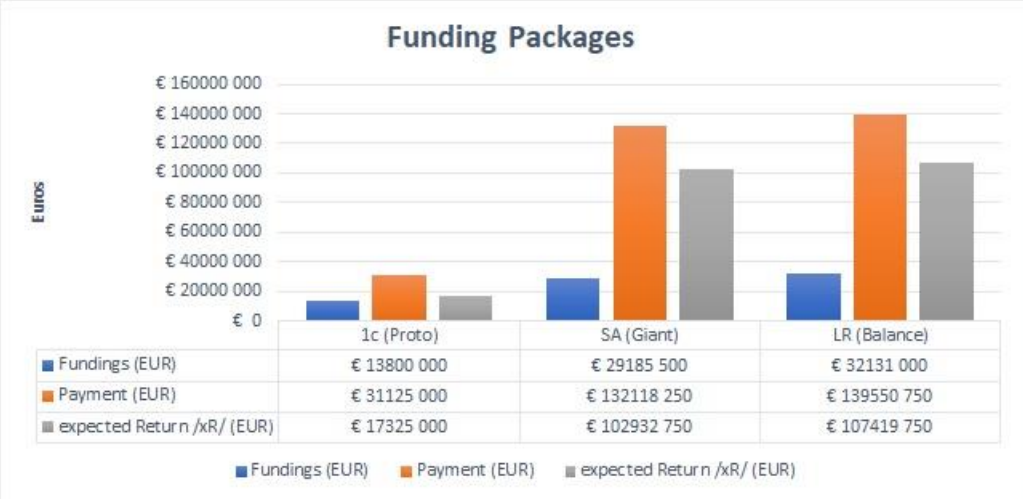
### 6.3.2. Funding Program (Win-win scenario)

The first step in creating funding packages is to carefully analyse the data and results that have been already created. This may involve using economic calculations such as net present value (NPV, Fig 8.), internal rate of return (IRR, Fig. 12.), return of investment (ROI, Fig. 10.), profitability index (PI, Fig 11.) and the brand-new metric: **expected Return (xR)** (Fig.8., 9.)

| Packages | Fundings (EUR) | Payment (EUR) | expected Return /xR/ (EUR) | Return on Investment (ROI) % | Profit/Net Income (EUR) | Net Present Value (NPV) | Profitability Index (PI) | Internal Rate of Return (IRR) |
|---|---|---|---|---|---|---|---|---|
| 1c (Proto) | € 13 800 000 | € 31 125 000 | € 17 325 000 | 126 | 17325000 | 9192815 | 1.67 | 45.00% |
| SA (Giant) | € 29 185 500 | € 132 118 250 | € 102 932 750 | 353 | 102932750 | 1255660 | 1.04 | 10.77% |
| LR (Balanc | € 32 131 000 | € 139 550 750 | € 107 419 750 | 334 | 107419750 | 4558569 | 1.14 | 12.41% |

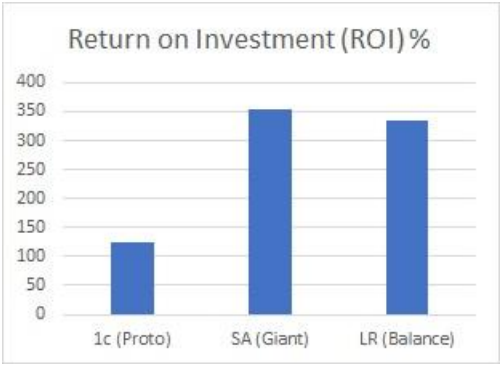
8. Figure: Funding package calculations

source: own presentation / https://miau.my-x.hu/miau/299/xR/ (Sheet = Fundings Range = A1:L4)




9. Figure: Funding packages

source: own presentation / https://miau.my-x.hu/miau/299/xR/ (Sheet = Fundings)




10. Figure: Return on Investment

source: own presentation / https://miau.my-x.hu/miau/299/xR/ (Sheet = Fundings)


11. Figure: Profitability Index

source: own presentation / https://miau.my-x.hu/miau/299/xR/ (Sheet = Fundings)


12. Figure: Internal Rate of Return

source: own presentation / https://miau.my-x.hu/miau/299/xR/ (Sheet = Fundings)

Might creating a conservative package that is focused on low-risk, lower-return fundings, a moderate package that offers a balance of risk and return, and a more aggressive package that is focused on high-risk, extremely high-return fundings.

Each funding package should include detailed information about the fundings that are included, as well as the expected returns and risks associated with each funding.

It's important to keep in mind that investing always involves risk, and that past performance is not a guarantee of future success. As such, it's important to be transparent with investors about the risks involved in each funding package, and to provide them with the information they need to make informed decisions.

Additionally, it's important to ensure that the funding packages are aligned with the values and goals of your organization. For example, if your organization has a focus on sustainability, you may want to include fundings.

Finally, it's important to be flexible and adaptable in your approach to funding packages. As market conditions change and new data becomes available, you may need to adjust your funding packages to ensure that they remain relevant and effective.

This situation can be described as a win-win scenario where both the club and the player benefit. It can be seen as a smart funding by the club to buy a young player with high potential for a low value and develop their skills, which can increase their value in the future. At the same time, the player benefits by having the opportunity to develop their skills in a professional setting and potentially increase their earnings and career prospects. Overall, this is a mutually beneficial situation for both parties involved.

### 6.3.3. Marketing

To create a strong branding identity for Starlight's funding packages, it's crucial to choose names that accurately reflect the content and risk level of each package. For instance, the low-risk funding package can be named "Starlight Proto," while the high-risk funding package can be called "Starlight Giant." These names should be distinctive and memorable to differentiate Starlight's funding packages from other organizations. (cf. 6.3.2.)

In addition to Starlight Proto and Starlight Giant, there is also the "Starlight Balance" funding package, which is designed for investors seeking a moderate level of risk and return. This package offers a balanced mix of investments across different leagues and player positions, with a focus on maximizing returns while minimizing risk.

Each of these funding packages is designed to meet the needs of different types of investors, from those seeking a low-risk, low-return investment to those looking for a high-risk, high-return opportunity. By carefully selecting names that accurately reflect the content and risk level of each package, Starlight can help investors make informed decisions about which package is right for them, while also building a strong and distinctive brand identity in the market.

In addition to choosing the appropriate names, Starlight should develop a visual identity or logo that represents the organization and its funding products. This brand identity should resonate with the target audience and reflect the values of Starlight, such as professionalism, passion for football, and commitment to developing young talents. (Fig. 13.)


13. Figure: An example of a logo design

source: own design

Starlight can leverage social media platforms such as Twitter, Instagram, and Facebook to showcase the success stories of young talents scouted and developed by the organization. This will help to build a strong online presence and increase brand awareness among football fans. Starlight can also organize local football tournaments and events to engage with the community and build a loyal fan base. These events can also serve as a platform to identify new talents and provide them with an opportunity to showcase their skills.

Collaborating with football clubs can help to strengthen Starlight's reputation as a reliable talent scout and developer. This can involve providing scouting services, developing youth academies, or even partnering with clubs to co-own players' rights. Sponsorship deals with

football clubs, tournaments, or individual players can be an effective way to increase brand exposure and establish Starlight as a reputable brand in the football industry.

Starlight can create content that provides insights and tips on football scouting and talent development. This can include blog posts, videos, and webinars. By providing value to the football community, Starlight can position itself as a thought leader in the industry.

Overall, Starlight should focus on its core values and strengths and communicate them effectively through various marketing channels to establish itself as a reputable brand in the football industry. By doing so, Starlight can attract more talented footballers to work with the organization and provide investors with a unique funding opportunity.

## 7. Future Advancements

Based on the results of this study, there are several avenues for future research that can help to improve the process of analysing footballers' positions and team tactics, as well as fine-tune the robot-method to make it more effective and attractive to potential investors.

One key area for improvement is the gathering of more data. While the dataset used in this study was sufficient for the purposes, expanding the dataset to include players from other regions and leagues would provide a more diverse set of data to work with. This could enable us to identify new patterns and relationships between player attributes and value, and test whether the same set of attributes and methods are effective in other contexts. Additionally, gathering more data on external events and other factors that can impact player value, such as injuries and player behaviour, could help to make the predictions more accurate and reliable.

Another area for improvement is the development of more sophisticated analysis techniques. While the methods used in this study were effective in generating profits, they were relatively simple compared to some of the more advanced statistical and machine learning methods that are available. Future research could explore the use of more advanced techniques, such as neural networks or deep learning algorithms, to identify even more subtle patterns and relationships in the data.

Additionally, the accuracy of the predictions could be further improved by incorporating more detailed information about team tactics and player positions. Positional data was used in the analysis, but there is still much to be learned about the nuances of different positions and how they impact player value. In the future, more granular approaches to positional analysis could be explored, such as examining specific attributes that are important for different positions or using advanced techniques such as spatial analysis to understand the relationship between player movement and value.

Finally, the insights gained from this study could be used to adjust the funding packages and make them more attractive to potential investors. By demonstrating the effectiveness of the robot-method in generating profits, the funding packages could be tailored to meet the needs and preferences of different types of investors. For example, packages could be offered that focus on specific leagues or player positions or offer different risk levels depending on the funders' appetite for risk.

In summary, much can be gained from further research in this field, including the expansion of the dataset, the development of more sophisticated analysis techniques, and the fine-tuning of the robot-method to make it even more effective and attractive to potential investors. By continuing to explore these avenues, a better understanding of the complex factors that influence player value can be achieved, and stable positive financial indexes for the funders can be created.

# 8. Summary

Present status:

Moneyball (Miller, 2011/ Lewis, 2003), the American sports drama film can be interpreted as a massive initialization towards data-driven decision making in the sport-economics. Parallel, (inter-)national rankings in all sport disciplines should support the decision making e.g., in talent management (of footballers) and/or in case of sport bets. On the other hand: the talent-management is and will also be a kind of forecast-oriented activity.

Goal/Task:

The goal is simple and trivial: creating a robot being better than the human benchmarks concerning talent-management (c.f. identification of football players here and now having a massive potential in the future, where this potential can be interpreted as the trend of the monetary value development of a football player). If the goal is simple, so the task is also trivial: searching for new knowledge representation forms (models) being capable of better filtering young talents than the human experts do it.

Solution:

The human solution (the benchmark) can be automated - like the searched AI. The human expert can derive correlation values between the economic value of the focused footballer set and each descriptive phenomenon (like Aggression, Corners, Crossing, Dribbling, Finishing, FirstTouch, Freekicks, Heading, LongShots, Longthrows, Marking, Passing, PenaltyTaking, Tackling, Technique, Anticipation, Bravery, Composure, Concentration, Vision, Decisions, Determination, Flair, Leadership, OffTheBall, Positioning, Teamwork, Workrate, Acceleration, Agility, Balance, Jumping, LeftFoot, NaturalFitness, Pace, RightFoot, Stamina, Strength, Consistency, ImportantMatches, Versatility, Adaptability, Ambition, Loyalty, Pressure, Professional, Sportsmanship, etc.). The source of the data is: https://www.sigames.com/ / SI Games. The human expert selects a relatively small number of these attributes (e.g., 5) having high correlation levels in the past (year-by-year). Then, the entire set of footballers will be filtered based on age and the level of the selected attributes.

Thresholds can be constant or even pre-defined through formulas (like MODUS). The human expert (the talent-fisherman) is good, if the ratio of the players with value-increasing-trend in the selection is high. The own AI should optimize the human-like analytical process.The goal of the task is therefore to create a robot that is better than human experts at talent management in football, specifically in identifying young football players with a high potential for future monetary value development. The task involves searching for new knowledge representation forms (models) that can better filter young talent than human experts. The human solution, which serves as the benchmark, involves selecting a relatively small number of attributes that have high correlation levels with the economic value of a set of football players. The human expert selects these attributes based on past data and filters the entire set of players based on age and the level of the selected attributes. The goal is to optimize the human-like analytical process to make the AI even better than the human expert.

Already closed experiments:

Based on the data concerning Turkish footballers in 2017, 129 players were analysed with a focus on 46 professional attributes related to the players' values from 2017 to 2022 each year. The top 5 attributes that had the most influence on a player's value were Composure, Passing, Anticipation, Decisions, and Stamina. After further filtering for players under the age of 23, 28 desired players were selected from the original 129. The experiment then started with setting the values of the top 5 attributes (Composure, Passing, Anticipation, Decisions, and Stamina) to 9, 10, 11, and 12, which resulted in the following outcomes. Results:

66 players with 21.5% accuracy and 313 083 euros/year in earnings. / 39 players with 20.5% and 490 883 euros/year in earnings.

15 players with 20.0% and 250 579 euros/year in earnings. / 4 players with 50.0% and 377 351 euros/year in earnings

Another approach was to calculate the mode of each of the top 5 attributes and add them up, giving a total of 52 points out of 100 (as the attributes rating goes from 1 to 20, with 20 being the best). This resulted in 36 players with 13.9% and 472 479 euros/year.

There was yet another approach where the top attribute (Stamina) was replaced by another positively correlated attribute, Pace (rank 15). The following results were obtained using the same filtering method as before:

20 players with 80.0% and 1 276 149 euros/year in earnings. / 14 players with 71.4% and 1 957 090 euros/year in earnings. / 3 players with 100.0% and 1 252 899 euros/year

Finally, the robot-eye method (similarity analysis - https://miau.my-x.hu/myx-free/) resulted in 2 393 055 euros/year in earnings, while the regression method resulted in 1 957 744 euros/year in earnings.

Discussion:

The goal of this project is to create a robot capable of outperforming human experts in talent management for football, specifically in identifying players with high potential. The human method involves selecting a few attributes, such as anticipation, composure, and stamina, that have a high correlation with the player's monetary value, and then filtering players based on age and the level of these attributes.

However, this approach can result in errors or low success ratios if the number of attributes chosen, or the minimum level of these attributes is too high. The robot-eye, with its ability to see complex patterns in raw data, can increase the success ratio through solver-based processes. The publication focuses on player scouting using data obtained from football and

aims to lay the foundation for a talent recognition system to help young players who are stuck in the grid, based on their position, or playing style.

The research is based on six years of data from over 150 000 real football players and involves determining development patterns and prominent developments through anti-discrimination analysis.

This research also seeks to answer the question of whether it's possible to determine five essential attributes for being an excellent football player.

Additionally, this research has resulted in the development of a new metric called **xR (expected Return)**, which can be used to determine a player's valuation change based on their attributes. The main advantage of this project is the creation of a closed-loop feedback system that surpasses human experts in talent management for football. This is achieved through the robot-eye's ability to identify complex patterns in raw data, increasing the success ratio of player scouting. The system's effectiveness has been demonstrated by the high correlation between **xR** and actual player valuations, as well as the successful identification of players with high potential.

The target audience for this research is vast, as football is the most popular sport worldwide, with billions of people watching it according to FIFA, and the research has applications for a variety of groups, including clubs, scouts, media platforms, journalists, and IT developers and startups.

Future:

In the future of talent management in football, a more comprehensive approach to data analysis will be necessary to effectively evaluate player potential. While individual attributes remain an important consideration, it will also be crucial to evaluate a player's fit within the team dynamic and their contribution to the overall capabilities of the team. This shift towards a more team-focused approach will require a sophisticated data analysis system that takes into account both individual attributes and team dynamics. This research has already laid the foundation for a potential startup, which can even generate interest in the startup community among clubs, scouts, media platforms, journalists, and IT developers.

Demo-URLs: https://miau.my-x.hu/miau/297/starlight_v1.xls

# 9. References


| Author | Year | Content | Publisher | URL | Last Download |
|---|---|---|---|---|---|
| SI Games | 2017-2022 | Football Manager | Sega | https://www.sigames.com/ | 2023.04.14. |
| Michael Lewis | 2003 | Moneyball: The Art of Winning an Unfair Game | W. W. Norton & Company | https://www.amazon.com/Moneyball-Art-Winning-Unfair-Game/dp/0393324818 | 2023.04.14. |
| Bennett Miller | 2011 | Moneyball (film) | Sony Pictures | https://www.sonypictures.com/movies/moneyball | 2023.04.14. |
| Chris Anderson and David Sally | 2013 | The Numbers Game | Penguins Book Limited | https://www.penguinrandomhouse.com/books/314351/the-numbers-game-by-chris-anderson-and-david-sally/ | 2023.04.14. |
| Miléna Gergics | 2021 | "The Robot-coach" | MY-X research team Hungary | https://miau.my-x.hu/miau/281/robotedzo.pdf | 2023.04.14. |
| Zsolt Lajos Kontra, MY-X research team Hungary | 2023 | Demo-Url | MY-X research team Hungary | https://miau.my-x.hu/miau/297/starlight_v1.xls | 2023.04.14. |
| Ryan Baldi | 2022 | Article on "Moneyball" | The Guardian | https://www.theguardian.com/sport/20 | 2023.04.14. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | 22/jun/21/moneyballs-nick-swisher-most-players-were-signing-cards-we-were-signing-books | |
| Zak Garner-Purkis | 2020 | Article on Liverpool FC approach | Forbes | https://www.forbes.com/sites/zakgarnerpurkis/2020/09/28/why-liverpools-scientists-will-decide-the-premier-league-title/ | 2023.04.14. |
| Training Ground Guru | 2022 | Article on Manchester City FC approach | Training Ground Guru | https://trainingground.guru/articles/mistry-promoted-as-man-city-continue-to-grow-data-science-department | 2023.04.14. |
| Scisports | 2021 | Article on Vahe Tanielian | Scisports | https://www.scisports.com/scouting-sessions-how-data-analytics-is-used-in-major-league-soccer/ | 2023.04.14. |
| Opta sports | 1996 | Sports analytics | Stats Llc. | https://statsperform.com/opta | 2023.04.14. |
| Prozone Sports | 1998 | Sports analytics | Prozone Sports Ltd. | http://www.prozonesports.com | 2023.04.14. |
| Football Radar | 2010 | Sports analytics | Private limited Company | https://www.footballradar.com/ | 2023.04.14. |

| | | | | | |
|---|---|---|---|---|---|
| Wyscout | 2004 | Sports analytics | Hudl | https://wyscout.com/ | 2023.04.14. |
| Statsbomb | 2016 | Sports analytics | Private | https://statsbomb.com/ | 2023.04.14. |
| Abhas Ricky | 2019 | How Data and Analytics Are Changing the Sports Industry | Forbes | https://www.forbes.com/sites/forbestechcouncil/2019/01/31/how-data-analysis-in-sports-is-changing-the-game/ | 2023.04.14. |
| Nick Ismail | 2017 | How big data is changing the way football is played | www.information-age.com | https://www.information-age.com/cloud-based-data-analytics-taken-world-professional-football-6533/ | 2023.04.14. |
| GuideToFM | 2018 | Player Attributes | GuidetoFM | https://www.guidetofm.com/players/attributes/ | 2023.04.14. |
| Espen | 2020 | Football Manager Player Attributes & Hidden Attributes Explained | Passion4FM | https://www.passion4fm.com/football-manager-player-attributes/?amp=1 | 2023.04.14. |
| MY-X research team Hungary | 2004 | Guide to Similarity Analysis | : Innoreg Regional Innovation Agency of | https://miau.my-x.hu/miau/196/My-X%20Team_A5%20fuzet_EN_jav.pdf | 2023.04.14. |

| | | | | | |
|---|---|---|---|---|---|
| | | | Central-Hungary Khe. | | |
| Unknown | 2023 | Article on Total Quality Management | Wikipedia | https://en.wikipedia.org/wiki/Total_quality_management | 2023.04.14. |
| Geoff Hulten | 2019 | Article on Closed-loop feedback system | Microsoft | https://learn.microsoft.com/en-us/archive/msdn-magazine/2019/april/machine-learning-closed-loop-intelligence-a-design-pattern-for-machine-learning 04/01/2019 | 2023.04.14. |
| Sathya Jameson | a year ago (2022) | Article on AI in Healthcare | Nanonets | https://nanonets.com/blog/ai-in-healthcare/ | 2023.04.14. |
| Zsolt Lajos Kontra, MY-X research team Hungary | 2023 | own research | MY-X research team Hungary | https://miau.my-x.hu/miau/299/xR/ | 2023.04.14. |

# 10. Attachments/Annexes

## List of abbreviations

| | |
|---|---|
| UID | Unique Identification |
| OAM | Object-Attribute-Matrix |
| COCO | Component-based Object-Comparison for Objectivity |
| COCO_Y0 | Component-based Object-Comparison for Objectivity ideal searcher |
| COCO_STD | Component-based Object-Comparison for Objectivity standard |
| AI | Artificial Intelligence |
| CBR | cases-based reasoning |
| MLSZ | Magyar Labdarúgó Szövetség (HFF – Hungarian Football Federation) |
| FIFA | Fédération Internationale de Football Association |
| IT | Information Technology |
| URL | Uniform Resource Locator |
| FC | Football Club |
| xR | expected return |
| NPV | net present value |
| ROI | return on investment |
| PI | profitability index |
| IRR | internal rate of return |
| HR | Human Resources |
| SA | Similarity Analysis |

## Definition of the attributes

- https://www.guidetofm.com/players/attributes/ - (GuideToFM, 2018)
- https://www.passion4fm.com/football-manager-player-attributes/?amp=1 - (Espen, 2020)