**PROJECT STARLIGHT: ROBOT-EYE FOR (TURKISH) FOOTBALL PLAYER-TALENT-IDENTIFICATION**

Zsolt Lajos Kontra (<https://orcid.org/0000-0002-9280-0560>) - László Pitlik (<https://orcid.org/0000-0001-5819-0319>)

e-Mails: [zskontra00@gmail.com](mailto:zskontra00@gmail.com), [pitlik@my-x.hu](mailto:pitlik@my-x.hu) / Kodolányi János University and MY-X research team Hungary

Mobile phones: Zsolt Lajos Kontra: +36709478339 / László Pitlik: +36309816001

Field of study: economics, finance, computer science

Keywords: sport-economics, talent-management, artificial intelligence, optimization, forecasting

# Abstract

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 Pace) 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. 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, it will be crucial to not only evaluate a player's individual attributes but also their contribution to the team's overall capabilities. This calls for a more comprehensive approach to data analysis that takes into account not just the player's individual attributes, but also their ability to fit within the team dynamic. This shift from solely analysing a player's attributes as an individual to considering their impact on the team will be necessary for effective talent management in football.

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

# Keywords

sport-economics, talent-management, artificial intelligence, optimization, forecasting

# 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, we aim 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. Our study builds upon this foundation by introducing a novel approach that utilizes correlation calculations to determine the most significant attributes for player valuation.

To establish our benchmark, we analysed a dataset of 129 players and 46 professional attributes over a five-year period from 2017 to 2022. Through a rigorous filtering process, we identified a subset of 28 players who were deemed to be the most valuable. By applying our correlation-based approach, we were able to determine the most significant attributes for each of these players, and from there, predict their earnings potential.

Our benchmark will be based on the accuracy of our player selection and the earnings predicted for each player. We anticipate that our approach will outperform previous benchmarks, with more accurate player selection and higher earnings potential for each player. By establishing this new benchmark, we hope to contribute to the ongoing development of player valuation methods in 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.

## 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.)

## 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 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.

In conclusion, this study aimed to provide a comprehensive analysis of the top attributes that contribute to a football player's value and to determine the most effective selection process for identifying valuable players. Through the use of an OAM and a variety of selection and valuation methods, the study was able to identify a set of relevant attributes and evaluate their impact on player value, but the case with the random selected attributes show, the correlation level as filter is less effective than expected.

## 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.

## 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.

## 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.

# 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 three criteria above, you can provide a comprehensive review of the relevant literature on your problem or phenomenon of interest. This will help to establish the context for your own research and demonstrate the importance of your study in advancing our understanding of the topic.

## 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) In the following chapters, we 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, Gergics Miléna 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."

## 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.

## 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, (incl. (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.

## 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.

# Data and methods

## Own dataset

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

Our 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. We also included subjective attributes such as market value.

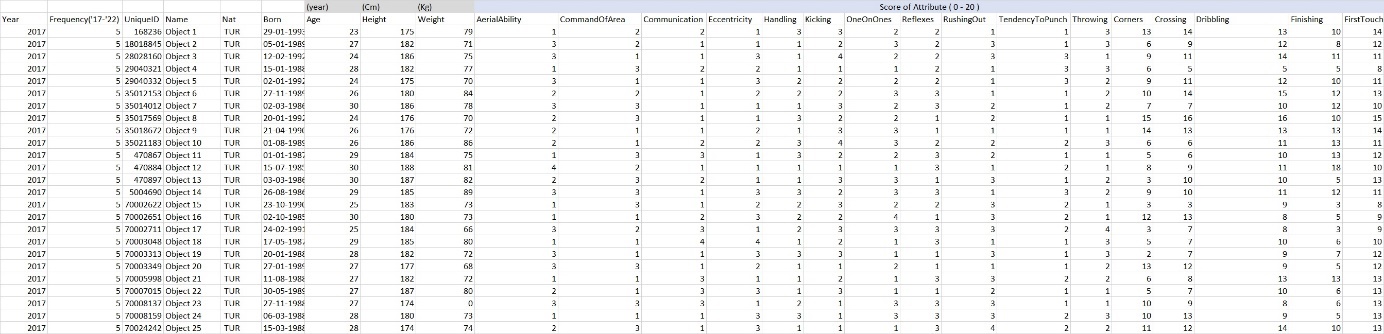
To ensure the accuracy/quality and quasi completeness of our dataset, we employed a rigorous data cleaning and validation process which also involved cross-checking the data (like filtering the whole dataset of the year 2017 [50.000+ players] for nationality, age etc. until we got acceptable pool of players (circa 130), then using MS Excel VLOOKUP function to check the players unique ID for occurrence year by year).

Overall, our dataset and methodology provide a comprehensive and robust approach to analysing player value in Turkish football, and we believe our findings will be valuable for a range of stakeholders in the football industry.

## Own methodology

In this study, we utilized the Object-Attribute-Matrix (OAM) methodology 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.

To develop our OAM model, we began by collecting data on a large sample of Turkish football players, including their game statistics, personal characteristics, and team performance metrics. We then identified a set of relevant attributes based on our knowledge of the sport, previous research, and consultation with experts in the field.



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)

Using a combination of statistical analysis and expert judgment, we assigned numerical values to each player for each attribute. These values were then organized into a matrix format, with each row representing a player object and each column representing an attribute. This matrix allowed us to easily compare players based on their individual attribute scores and overall performance by using similarity analyses.

Using correlation analysis, we filtered out the attributes considering the values of the players, and then we applied regression analysis using the players' market value and attributes. These calculations allowed us to determine the 5 attributes out of 46 that have the greatest influence on a player's market value each year. We examined this for each year separately and then ranked the 5 key attributes based on this filtering to try to demonstrate the correlation between attributes and player value. We then used this as a benchmark since our goal was to find players whose value only increases over time, making them a promising investment. In one experiment, the top 5 attributes were replaced with a random one, and this resulted in higher hit rates than the 5 highest value attributes, demonstrating that the benchmark could be surpassed.



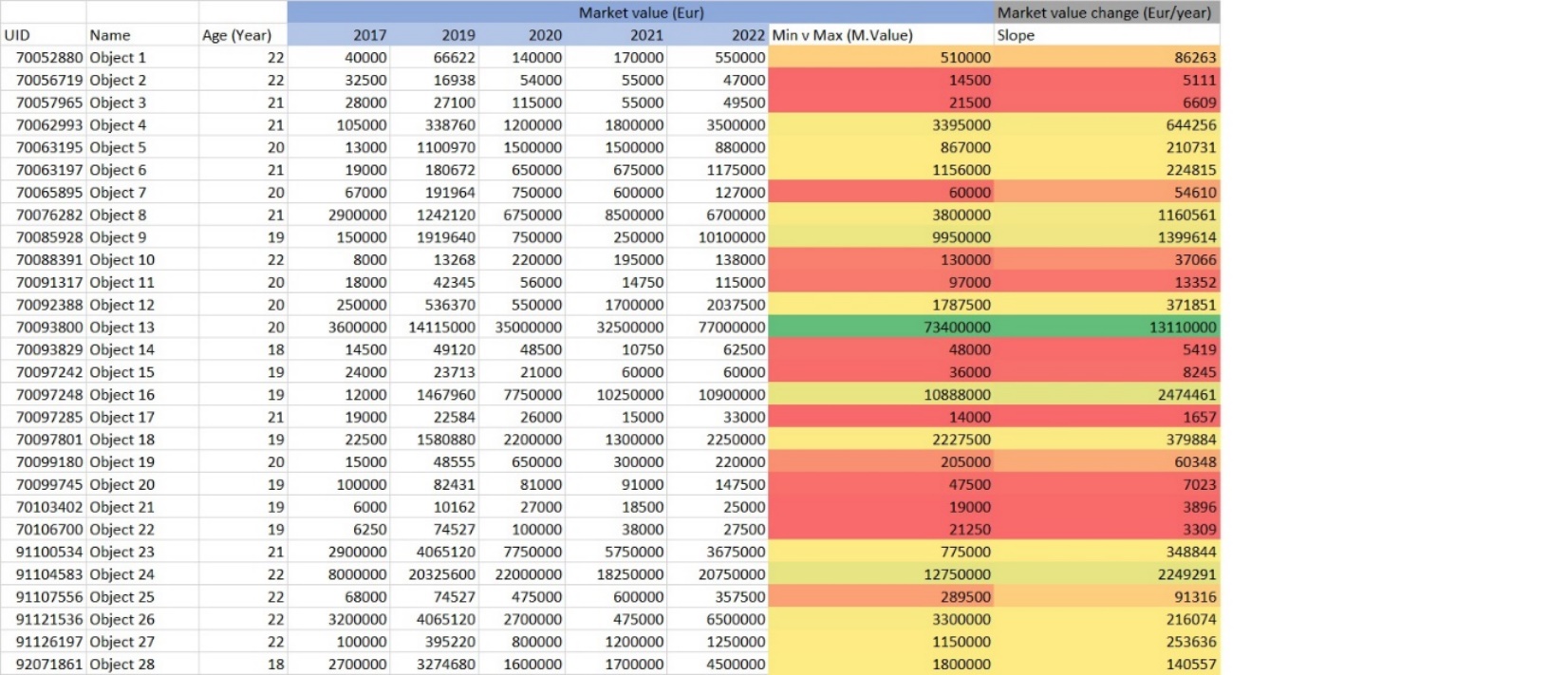
2. Figure: Correlation between attributes and market value

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



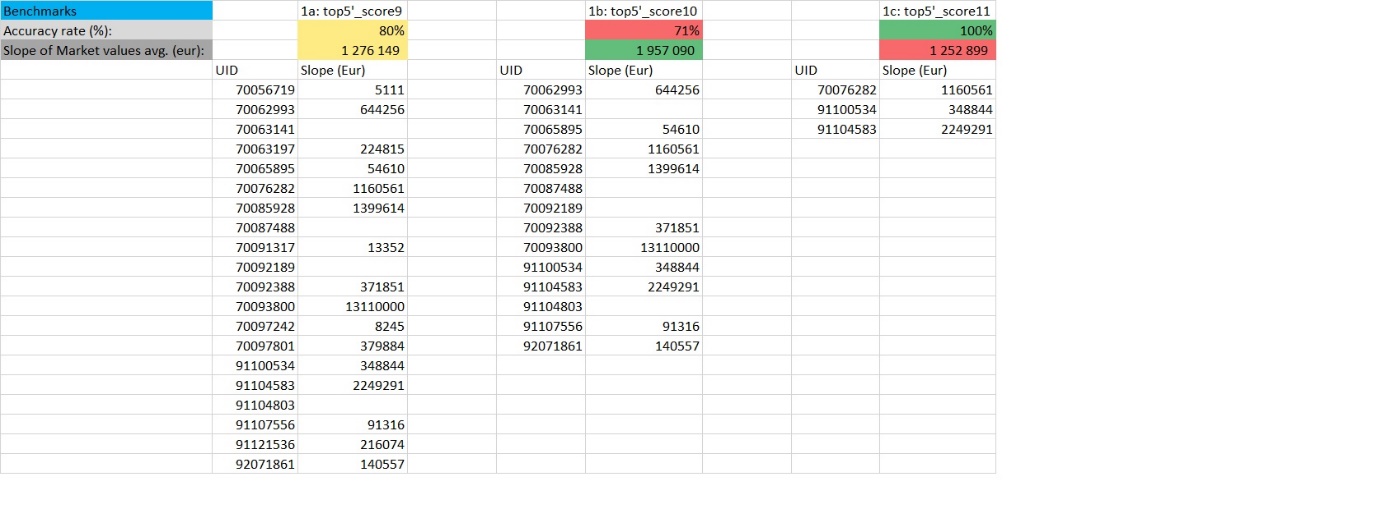
3. Figure: Top 5 Attributes

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



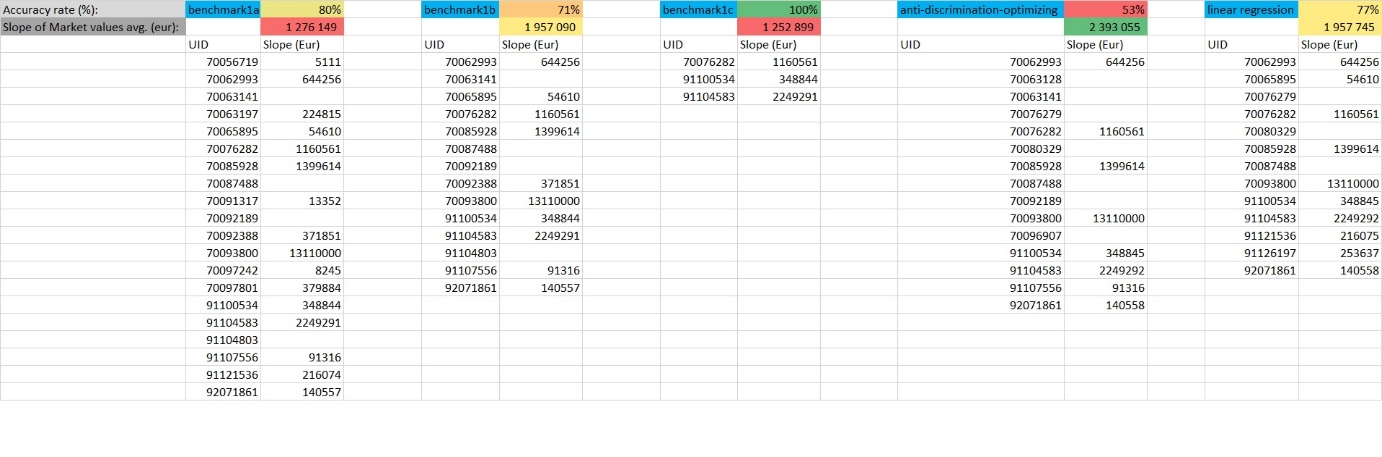
4. Figure: Desired players

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



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)



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, our 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, we hope to contribute to a better understanding of football performance and to support the development of more effective strategies for building competitive teams and maximizing player value.

## 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.

In this study, we aimed to develop a method to predict which young Turkish football players will have a high chance of increasing their value in the future. We started by creating a dataset that initially included 160,000 football players, which we filtered down to only Turkish players. After further filtering, we were left with 129 players who we analysed using 46 professional attributes related to the players' values.

## Experiments and Hypotheses

Our 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. We propose the following experiments to test the efficacy of our 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. We had access to numerous statistical data about these players and focused on their 46 attributes.

Data Analysis: We will use statistical analysis and machine learning techniques 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 our data analysis, we will develop a model for predicting player development and future value, incorporating key performance and physical indicators as well as personal characteristics.

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

Implementation: We will work with football clubs and scouting organizations to implement our methodology in real-world player evaluation and scouting processes, with the aim of maximizing player value and team performance.

Our hypothesis is that by using our methodology, we will be able to accurately identify young football players with high potential for development, and that by investing in and developing these players, we can increase their value and generate profit for clubs and organizations. Furthermore, we believe that by incorporating personal characteristics and other non-performance factors into our model, we can develop a more comprehensive and accurate approach to player evaluation and development.

Based on a sufficient amount of data, we aim to track players' development and identify the future talents. We also intend to investigate whether there are five essential attributes that are crucial for becoming an outstanding football player, and what they have in common. Additionally, we will explore if any trends can be observed in the dataset.

## Outcome

We conducted several experiments 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.

We also conducted experiments 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.

# 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.

# Conclusions

Our experiments show that it is possible to develop a method to select football players who are likely to increase in value in the future. Our 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.

# Future

Based on the results of this study, there are several avenues for future research. First, the dataset used in this study only included Turkish football players, and it would be valuable to expand the dataset to include players from other regions and leagues. This would provide a more diverse set of data and enable us to test whether the same set of attributes and methods are effective in other contexts.

Second, the study only focused on players under the age of 23, and it would be interesting to examine whether the same set of attributes and methods are effective for older players as well. As players age, different skills may become more important, and it would be valuable to understand how age interacts with player value.

Third, while the methods used in this study were effective in generating profits, they did not account for factors such as injuries, player behaviour, or external events that could impact a player's value. Future research could explore how to incorporate such factors into the analysis to generate more accurate predictions.

Finally, the study focused on using a combination of correlation and regression analyses to identify the most important attributes and predict player values. There are many other statistical and machine learning methods that could be used for this purpose, and it would be valuable to compare the performance of different methods to identify the most effective approach.

In summary, this study provides a valuable contribution to the understanding of how to predict and profit from the value of young football players. However, there is still much to be explored in this field, and further research can help to improve our understanding of the complex factors that influence player value and how to effectively predict it.

# Attachments

## 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

MLSZ Magyar Labdarúgó Szövetség (HFF – Hungarian Football Federation)

FIFA Fédération Internationale de Football Association

IT Information Technology

URL Uniform Resource Locator

FC Football Club

## Definition of the attributes

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

## List of figures

[1. Figure: Pool of Turkish players 14](#_Toc129728250)

[2. Figure: Correlation between attributes and market value 15](#_Toc129728251)

[3. Figure: Top 5 Attributes 15](#_Toc129728252)

[4. Figure: Desired players 15](#_Toc129728253)

[5. Figure: Example for human benchmarks 16](#_Toc129728254)

[6. Figure: Human benchmarks vs Robot-eye method 16](#_Toc129728255)

# References

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author | Year | Content | Publisher | URL |
| SI Games | 2017-2022 | Football Manager | Sega | <https://www.sigames.com/> |
| 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> |
| Bennett Miller | 2011 | Moneyball(film) | Sony Pictures | <https://www.sonypictures.com/movies/moneyball> |
| 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/> |
| Miléna Gergics | 2021 | “The Robot-coach” | MY-X research team Hungary | <https://miau.my-x.hu/miau/281/robotedzo.pdf> |
| 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> |
| Ryan Baldi | 2022 | Article on “Moneyball” | The Guardian | <https://www.theguardian.com/sport/2022/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/> |
| 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> |
| Scisports | 2021 | Article on Vahe Tanielian | Scisports | <https://www.scisports.com/scouting-sessions-how-data-analytics-is-used-in-major-league-soccer/> |
| Opta sports | 1996 | Sports analytics | Stats Llc. | <https://statsperform.com/opta> |
| Prozone Sports | 1998 | Sports analytics | Prozone Sports Ltd. | [http://www.prozonesports.com](http://www.prozonesports.com/) |
| Football Radar | 2010 | Sports analytics | Private limited Company | <https://www.footballradar.com/> |
| Wyscout | 2004 | Sports analytics | Hudl | <https://wyscout.com/> |
| Statsbomb | 2016 | Sports analytics | Private | <https://statsbomb.com/> |
| 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/> |
| 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/> |
| GuideToFM | 2018 | Player Attributes | GuidetoFM | <https://www.guidetofm.com/players/attributes/> |
| Espen | 2020 | Football Manager Player Attributes & Hidden Attributes Explained | Passion4FM | <https://www.passion4fm.com/football-manager-player-attributes/?amp=1> |

# Table of Contents

[Abstract 1](#_Toc129762298)

[Keywords 4](#_Toc129762299)

[Introduction 4](#_Toc129762300)

[Goals 5](#_Toc129762301)

[Tasks 5](#_Toc129762302)

[Motivations 7](#_Toc129762303)

[Target audience 7](#_Toc129762304)

[Utility 8](#_Toc129762305)

[Relevant literature 8](#_Toc129762306)

[History of the problem/phenomenon 8](#_Toc129762307)

[The dataset of problem/phenomenon 9](#_Toc129762308)

[Methodology for the problem/phenomenon 10](#_Toc129762309)

[Potential solution alternatives 11](#_Toc129762310)

[Data and methods 13](#_Toc129762311)

[Own dataset 13](#_Toc129762312)

[Own methodology 14](#_Toc129762313)

[Results 17](#_Toc129762314)

[Experiments and Hypotheses 17](#_Toc129762315)

[Outcome 19](#_Toc129762316)

[Discussion 19](#_Toc129762317)

[Conclusions 20](#_Toc129762318)

[Future 20](#_Toc129762319)

[Attachments 21](#_Toc129762320)

[List of abbreviations 21](#_Toc129762321)

[Definition of the attributes 22](#_Toc129762322)

[List of figures 22](#_Toc129762323)

[References 22](#_Toc129762324)

[Table of Contents 25](#_Toc129762325)