Report onData-driven Solution for NAC

Petar Paskalev
Student number 232725
Date: 26/01/2024

DISCOVER YOUR WORLD

Academy/Department

Saturday 01 January 2022

Breda University OF APPLIED SCIENCES

Index

1 Introduction

1.1 Header

NAC Breda aims to enhance their team performance through strategic player acquisitions. The challenge lies in identifying cost-effective players who can significantly contribute to the team's success, particularly focusing on players with soon-to-expire contracts and standout performers.

1.2 Objective

My main objective is to optimize the player scouting process by leveraging machine learning to identify high-performing, affordable players. This involves analyzing player statistics from other promoted teams to identify gaps and opportunities for NAC Breda.

1.3 Approach

To achieve this, I adopted a multifaceted approach incorporating machine learning models to assess player performance comprehensively. This allows for a nuanced understanding of player strengths and weaknesses, aiding in strategic recruitment decisions.

2 Exploratory Data Analysis

2.1 Dataset Overview

The dataset consists of player statistics from 45 Excel files compiled into one data frame using Python, representing players from various European leagues. It includes 105 numerical features related to on-field performance and 9 categorical features such as player names, birth countries, and teams.

2.2 Data Preparation

I identified columns with missing data and removed rows with insufficient player statistics, resulting in the deletion of 237 rows. Key missing values were found in "Weight," "Height," and "Market Value."

Normalization and Ridge techniques were applied to handle outliers and understand the importance of features for central defenders. Categorical variables were encoded to determine player positions. The data was then normalized and categorical variables were encoded to ensure the dataset was ready for analysis. Separate data frames were created for each position and club of interest.

2.3 Summary Statistics

I calculated measures of central tendency and dispersion for numerical features to understand the dataset better. For example:
Market Value Range: 0 to 60,000,000
Average Age of Players: 25 years.

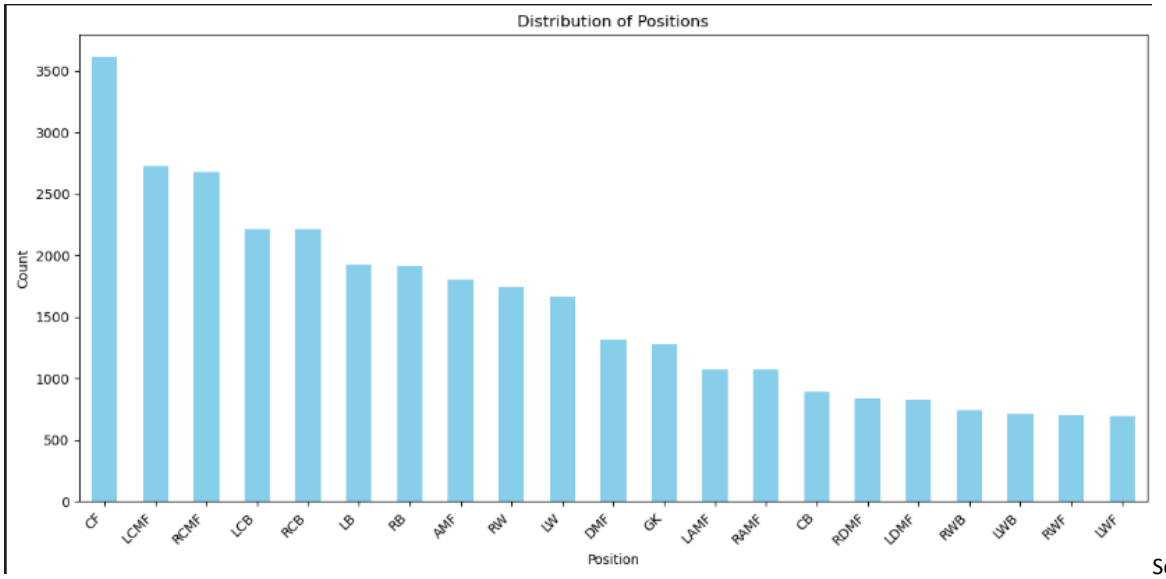
2.4 Frequency Counts

I examined the distribution of players' positions across different teams, revealing insights into player composition.

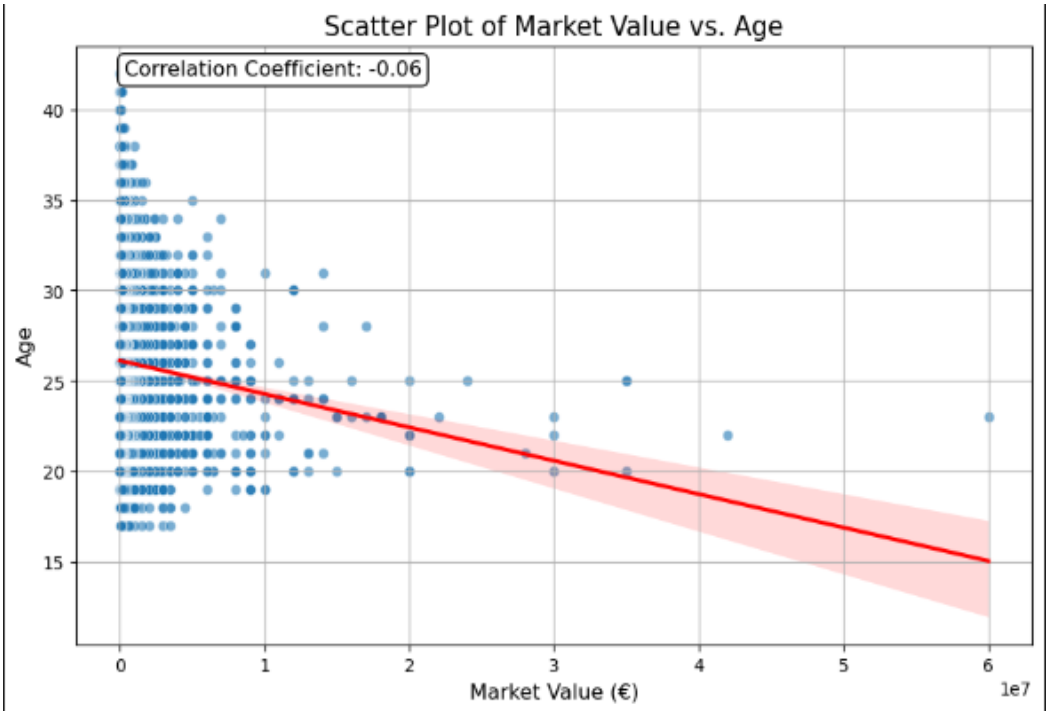
2.5 Data Visualizations

I used various plots to visualize data distributions and relationships:

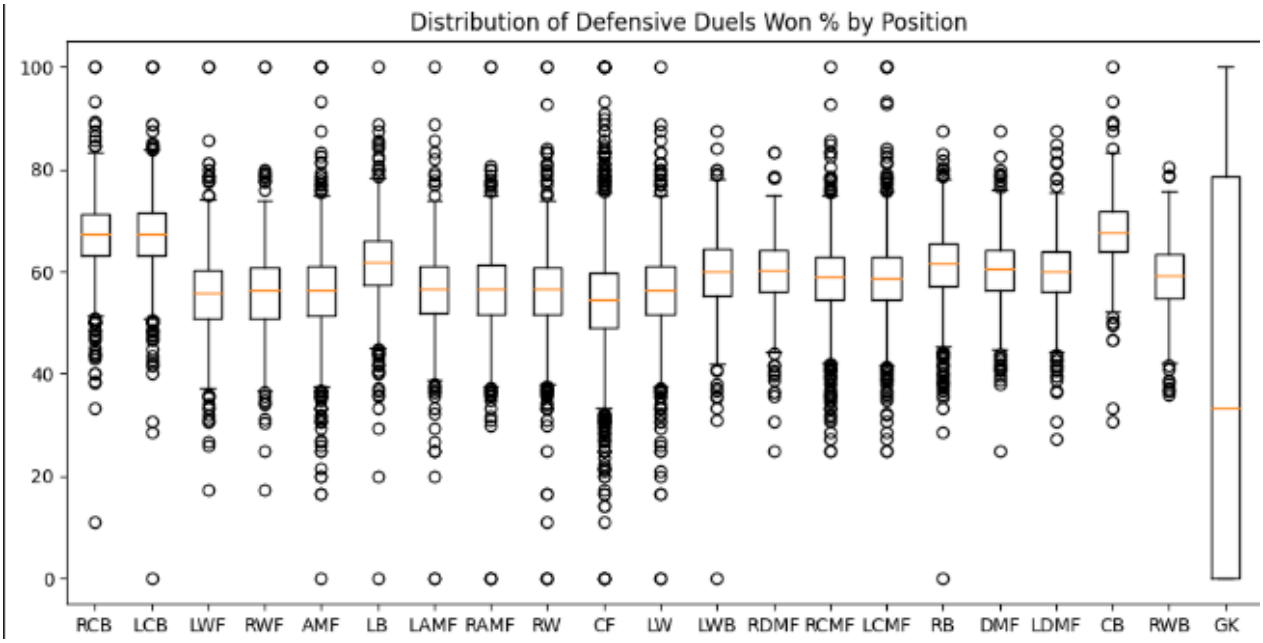
Histogram: To show the distribution of players positions in the dataset.



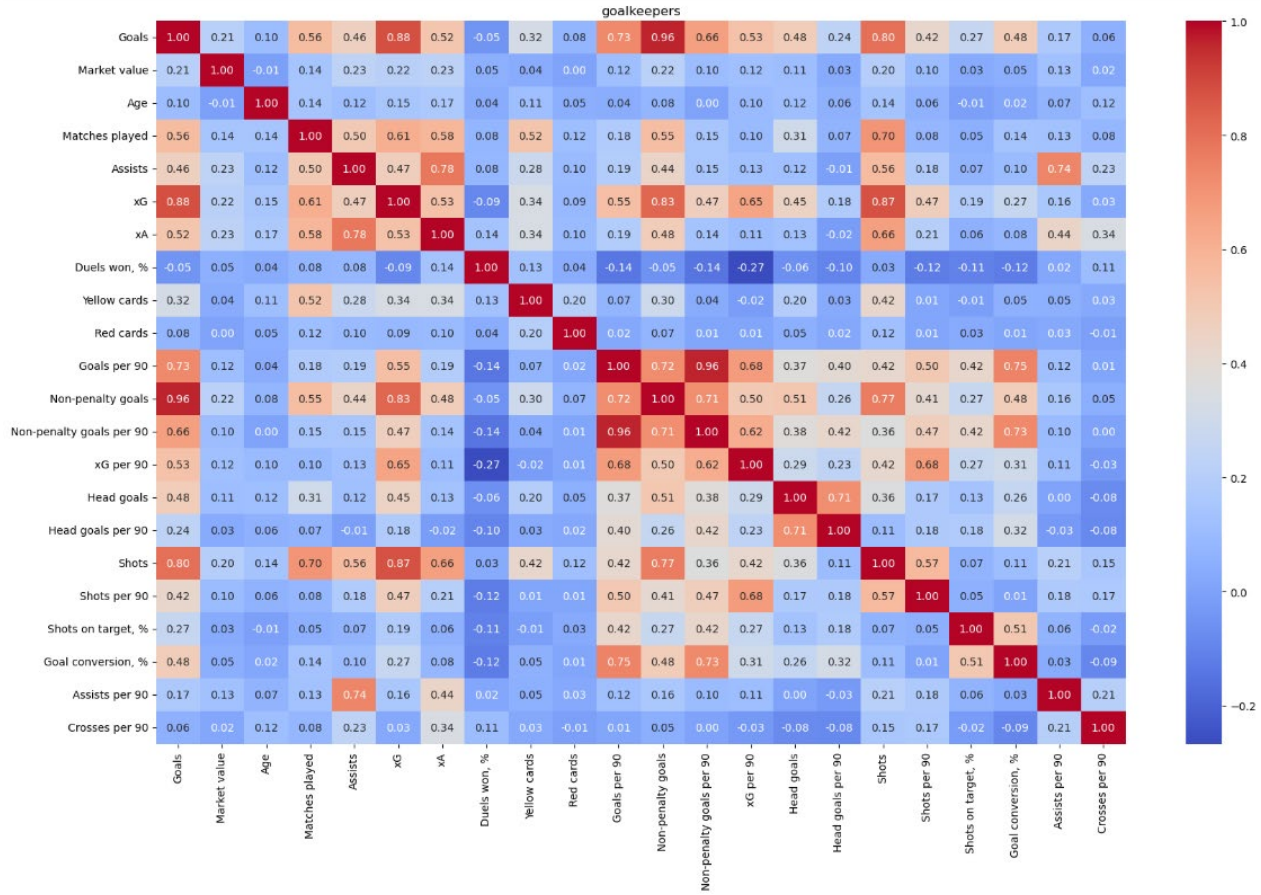
Scatter Plot: To illustrate the correlation between market value and age.



Box Plot: To compare duels won percentages across positions.

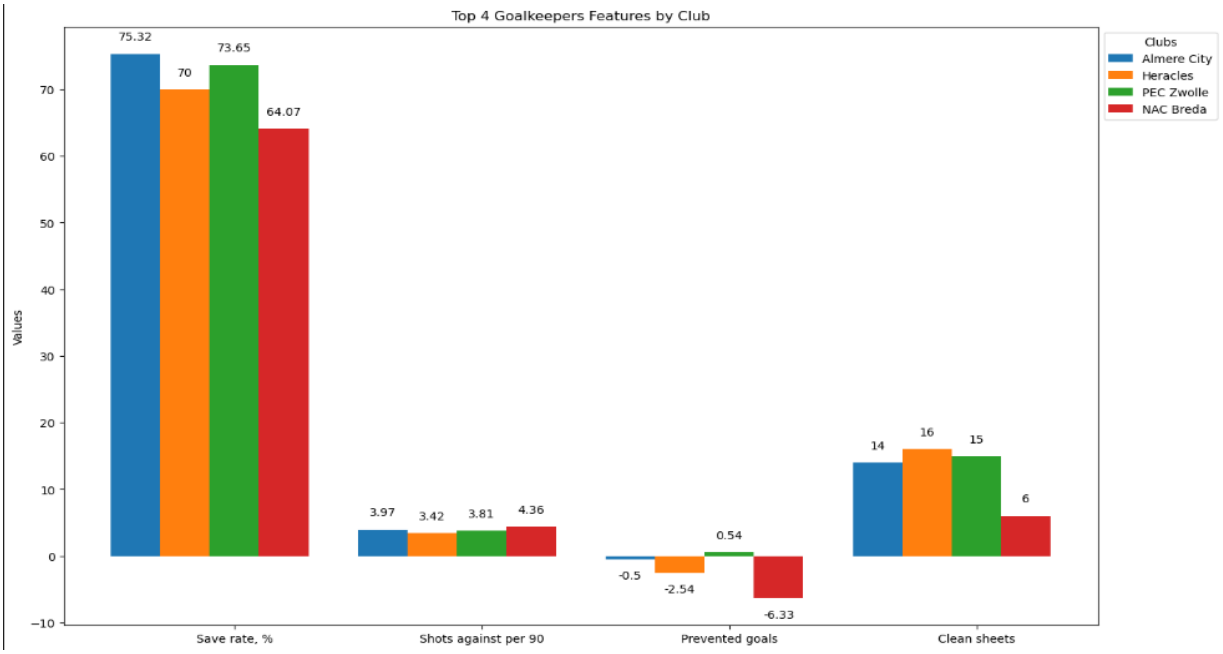


Heatmap: To highlight correlations between key performance metrics (e.g., xA and goals).

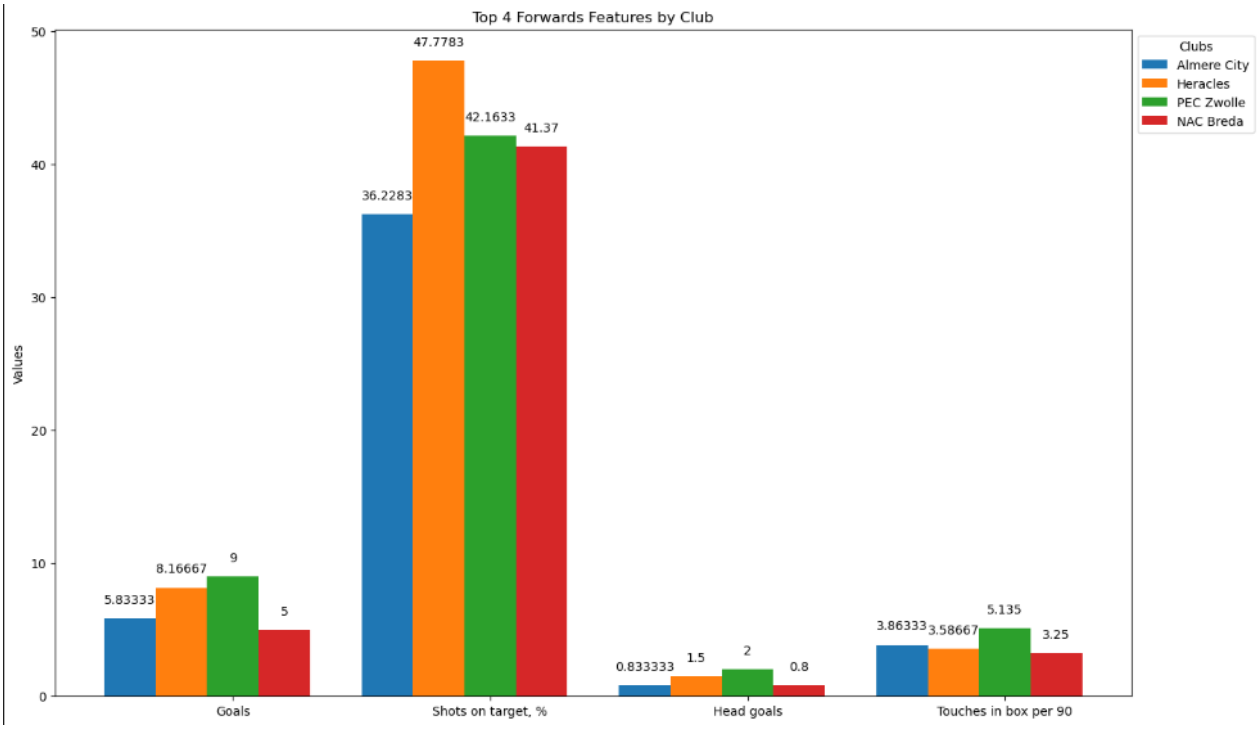


2.6 Key Findings

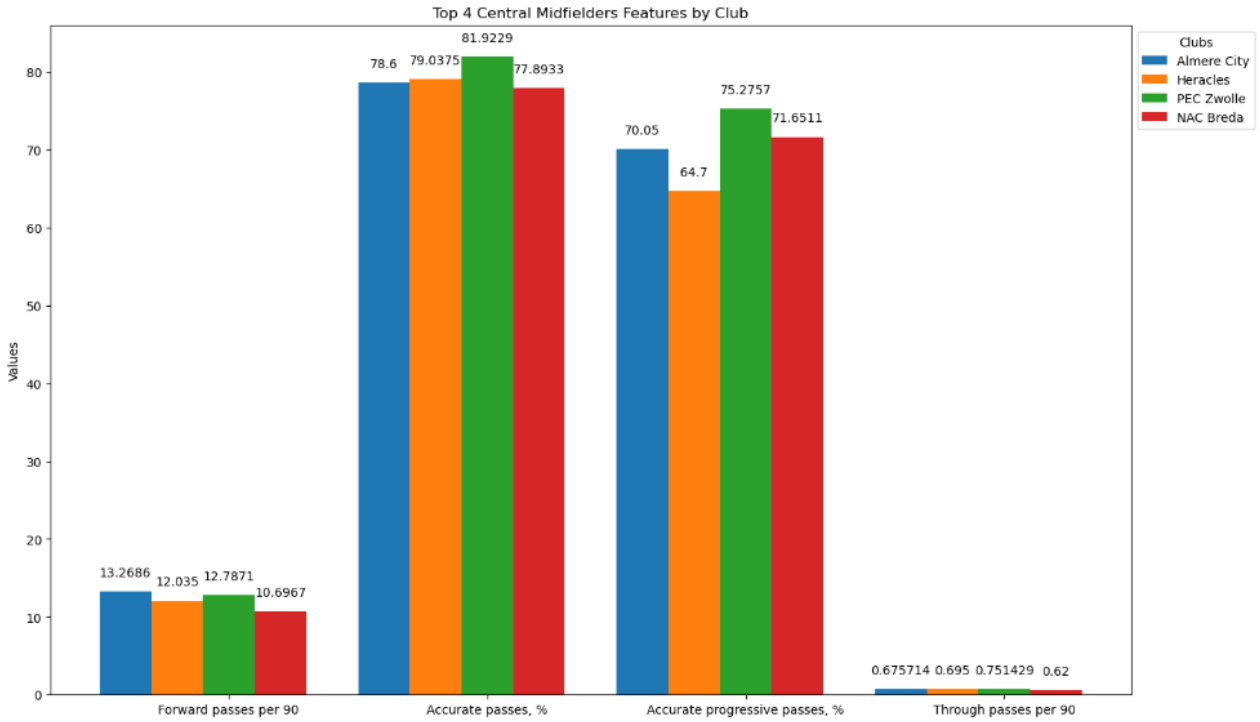
Goalkeepers: The analysis highlighted the need for new goalkeepers to enhance the team's defensive capabilities. Visualizations showed that NAC Breda's goalkeepers have lower save rates and clean sheets and most importantly prevented goals compared to the other clubs.



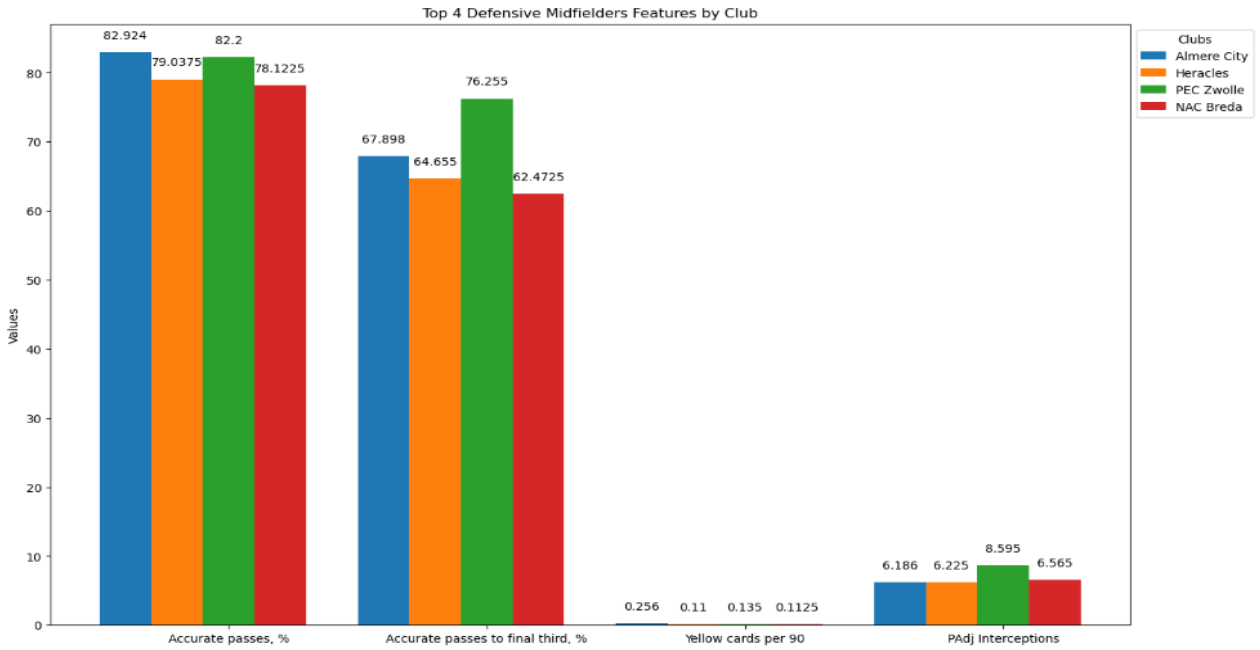
Forwards (Central Forward - CF): Despite modest goal-scoring figures, deeper analysis revealed significant overperformance relative to anticipated data, linked to challenges in build-up play. This was visualized through comparisons of key metrics like goals and assists per 90 minutes across clubs.



Central Midfielders: The analysis of central midfielders showed that NAC Breda lags behind in accurate passes, progressive passes, and forward passes per 90 minutes compared to other clubs.



Defensive Midfielders: Similar trends were observed for defensive midfielders, where NAC Breda performed poorly in accurate passes to the final third and interceptions.



To identify suitable players, I analyzed past transfers of goalkeepers in the team and found that NAC Breda did not spend any money on transfer fees. Therefore, I focused on players with expired contracts. This strategy ensures cost-effective acquisitions while enhancing the team's performance.

3 Machine Learning

3.1 Method

Type of Algorithm: Linear Regression

- Rationale: Linear Regression is a fundamental algorithm that assumes a linear relationship between the independent and dependent variables. I opted for this model initially due to its simplicity and interpretability, providing a baseline understanding of how individual features contribute to goalkeeper performance.

Type of Algorithm: Random Forest Classifier

- Rationale: Random Forest is an ensemble learning technique that combines multiple decision trees to capture non-linear relationships and complex interactions in the data. I chose this model to handle potential non-linear patterns and to assess the importance of different features.

Type of Algorithm: Gradient Boosting Regressor

- Rationale: Gradient Boosting is an ensemble technique that builds decision trees sequentially, with each tree correcting the errors of the previous one. This model provides high predictive accuracy and emphasizes feature importance. It emerged as the model of choice for its comprehensive understanding of feature relevance.

3.2 Model evaluation

Linear Regression

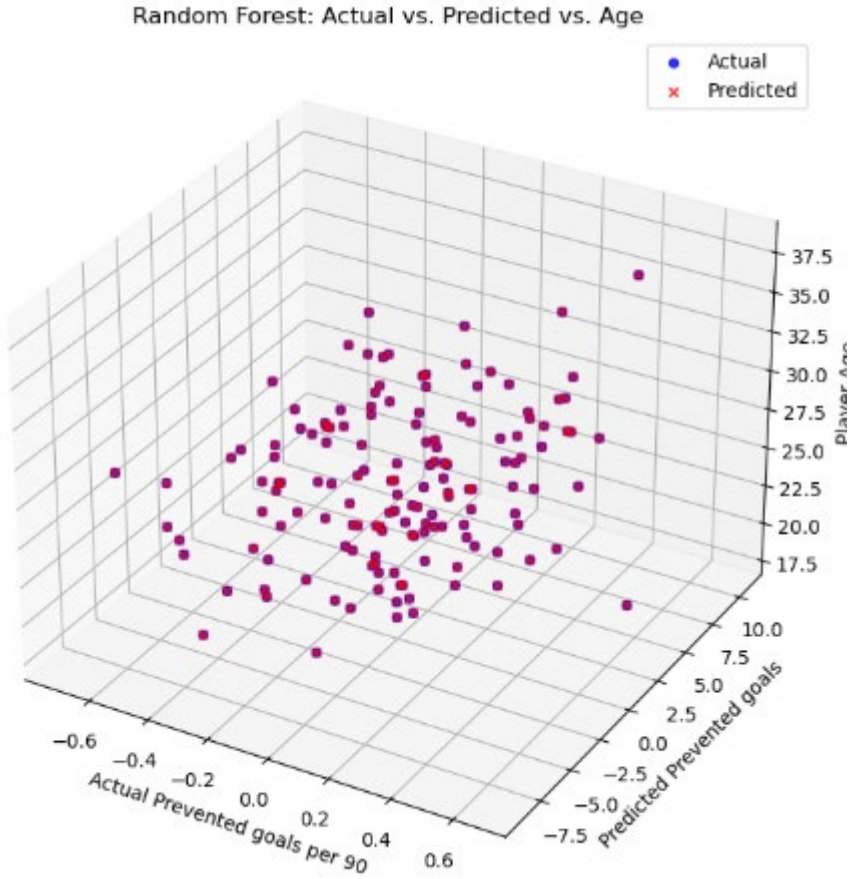
- R-squared: 0.68
- Mean Absolute Error (MAE): 1.54Mean Absolute Error: 1.5437553442716256

Random Forest Classifier:

- R-squared: 0.98
- MAE: 0.29

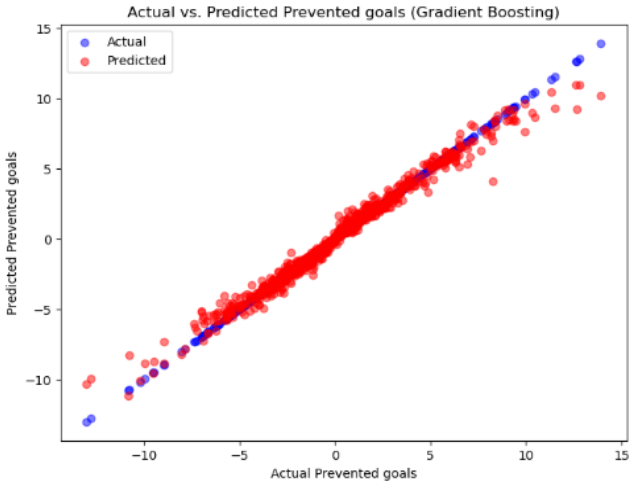
Cross-validation was used to ensure model robustness, particularly focusing on MAE to measure prediction accuracy.

The Random Forest model's performance was visualized through a 3D scatter plot comparing actual versus predicted values, indicating some overfitting issues.



3.3 Model improvement

The Gradient Boosting model demonstrated outstanding performance, making it my preferred choice for predicting prevented goals and other key metrics. The model's accuracy was visualized through a scatter plot showing a strong correlation between actual and predicted values.



4 Ethical Considerations

4.1 Ethical Organizational Capacity

I related transparency, accountability, and integrity to NAC Breda's operations, ensuring ethical practices in player recruitment.

4.2 Responsible Parties

I identified responsible parties within NAC for each ethical element, ensuring clear accountability.

4.3 Research Findings

From discussions with data scientists at NAC Breda, I understood that the club does not recruit players from outside Europe. While this practice aligns with specific strategic or regulatory considerations, it may also introduce a country bias, potentially overlooking talented players from other regions.

4.4 Ethical Decision-Making Framework

I followed an ethical decision-making framework, ensuring GDPR compliance and adherence to ethical guidelines.

4.5 GDPR and Ethical Guidelines

Evidence was provided of GDPR and Ethical Guidelines for Statistical Practice considerations.

4.6 Identified Ethical Problems

Privacy issues and potential biases in recruitment were identified and addressed. The exclusive focus on European players can be seen as a form of geographic bias, potentially limiting the club's talent pool.

4.7 Recommendations

My recommendations include:

- Strengthening Corporate Governance: Ensuring transparency and accountability in all recruitment processes.
- Enhancing Transparency: Clearly communicating recruitment policies and practices to avoid perceptions of bias.
- Increasing Community Engagement: Engaging with a broader community to understand diverse perspectives and reduce bias.

5 Recommendations

Leveraging the analytical insights derived from both the Linear Regression and Random Forest Classifier models, I can strategically enhance NAC Breda's player recruitment process. The following goalkeepers have been identified based on their predicted prevented goals, providing valuable options for reinforcing the team's defensive capabilities.
Linear Regression Model:

![Recommended Goalkeepers - Linear Regression]

Recommended Goalkeepers:

	Player	Team within selected timeframe	Age	Market value	Contract expires	Matches played	Minutes played	Clean sheets	Prevented goals	Save rate, %	Predicted Prevented goals
11943	V. Černiauskas	Panevėžys	34.0	200000	2023-12-31	28	2705	21	7.04	86.30	7.719118
11894	L. Paukste	Šiauliai	24.0	300000	2023-12-31	22	2114	12	3.79	75.95	4.661955
16378	A. Fagerström	Västerås SK	31.0	175000	2023-12-31	21	2082	12	3.38	78.67	4.299383
7883	Á. Ólafsson	Stjarnan	32.0	50000	2023-11-16	25	2418	9	5.76	75.00	3.724005
12141	L. Wahlstedt	Odds	24.0	1400000	2023-12-31	17	1658	8	3.06	75.00	4.451844
16327	R. Wallinder	Gefle	24.0	225000	2023-12-31	20	1984	5	9.92	79.31	6.373220
1068	S. Lammens	Club Brugge II	20.0	600000	2023-06-30	11	1050	3	4.72	79.10	5.387587

Selection Criteria:

- Predicted Prevented Goals: Prioritize players with higher predicted prevented goals for a robust defensive strategy.
- Contract Expiry: Focus on players whose contracts have already expired to explore cost-effective acquisitions.

Random Forest Classifier Model:

![Recommended Goalkeepers - Random Forest Classifier]

	Player	Age	Market value	Contract expires	Matches played	Minutes played	Predicted Prevented goals RF
16327	R. Wallinder	24.0	225000	2023-12-31	20	1984	8.7757
4182	O. Forsman	35.0	50000	2023-12-31	26	2499	7.6878
11943	V. Černiauskas	34.0	200000	2023-12-31	28	2705	6.3689
7883	Á. Ólafsson	32.0	50000	2023-11-16	25	2418	5.8017
1068	S. Lammens	20.0	600000	2023-06-30	11	1050	4.1232
11894	L. Paukste	24.0	300000	2023-12-31	22	2114	4.0039
16378	A. Fagerström	31.0	175000	2023-12-31	21	2082	3.5209
12141	L. Wahlstedt	24.0	1400000	2023-12-31	17	1658	3.2223
8098	D. Lyness	32.0	150000	2023-11-30	27	2647	3.0134

Selection Criteria:

- Predicted Prevented Goals: Emphasize players with significant predicted prevented goals for reliable defensive contributions.
- Contract Expiry: Prioritize players with expired contracts for strategic and cost-efficient acquisitions.

Next Steps:

In-Depth Player Evaluation: Conduct thorough assessments of each recommended goalkeeper, considering additional performance metrics, playing style, and compatibility with team tactics.

Contract Negotiation: Initiate negotiations with identified players, leveraging the advantage of their expired contracts to secure favorable terms.
Seamless Integration: Upon successful acquisitions, strategically integrate the new players into the team through well-planned training sessions and onboarding.
Future Considerations:

Extend the analysis to other playing positions, such as midfielders and forwards, to achieve a holistic improvement in team dynamics.
Regularly update predictive models to adapt to evolving player performances and market dynamics.
By adopting a data-driven approach to player recruitment, NAC Breda can make informed decisions, ensuring a resilient and competitive team in the dynamic landscape of football. Composition and elevate its competitive standing.

6 Reference

Paskalev, P. (2024). Data-driven Solution for NAC Breda. Breda University of Applied Sciences.

Transfermarket. (2024). Player market values. Retrieved from https://www.transfermarkt.com

Games

Leisure & Events

Tourism

Media

Data Science & AI

Hotel

Logistics

Built Environment

Facility

Mgr. Hopmansstraat 2
4817 JS Breda

P.O. Box 3917
4800 DX Breda
The Netherlands

PHONE
+31 76 533 22 03
E-MAIL
communications@buas.nl
WEBSITE
www.BUas.nl

DISCOVER YOUR WORLD

Breda University
OF APPLIED SCIENCES