Explainability of AI report

Machine Learning Explanation

ATTIG Léo

leo.attig@edu.univ-eiffel.fr

SAUTIER Sébastien

sebastien.sautier@edu.univ-eiffel.fr

PERRIOT Mathieu

mathieu.perriot@edu.univ-eiffel.fr

TREMELLAT Alexis

alexis.tremellat@edu.univ-eiffel.fr

Master 2 - SIA & SSIO

March 2025

Project Supervisor

Dr. ALAOUI Jawad

Contents

1 Introduction

The initial project aimed to leverage advanced data science methodologies to develop predictive models capable of determining the likely winners of football matches. By participating in the ”Football: Who will win?” competition on ChallengeData, match-related data was analyzed, various modeling approaches were explored, and models were fine-tuned to achieve optimal performance. The primary goal was to build a robust predictive framework while gaining insights into the factors that influence the outcomes of football matches.

This current project builds upon the previous one, shifting the focus towards explaining the decisions made by the predictive models through Explainable AI (XAI) techniques. The aim is to interpret the model's behavior and understand the key factors driving its predictions. To achieve this, various XAI methods, such as SHAP, LIME, and feature importance, are employed to provide both global and local explanations of the model's predictions. The project thus seeks to enhance the transparency and interpretability of the model, helping to identify potential biases and gaining a deeper understanding of the prediction process.

2 Methodology

In this project, we employed several XAI techniques to explain the behavior of the XGBoost model:

2.1 SHAP

SHAP is a method that uses cooperative game theory to explain the contribution of each feature to the final prediction. It provides both local and global interpretability, showing how each feature affects the model's output for individual predictions as well as across the entire dataset.

2.2 LIME

LIME is another model-agnostic technique for explaining individual predictions. It creates a locally interpretable surrogate model that approximates the behavior of the original model in the vicinity of a specific instance.

2.3 Feature Importance

Feature importance refers to evaluating which features are most influential in predicting the outcome. XGBoost provides a way to measure this by analyzing how much each feature contributes to reducing the loss function during training. We also explore Leave-One-Feature-Out (LOFO) as an additional method to assess feature relevance by removing one feature at a time and observing the impact on the model's accuracy.

3 Results

3.1 Global Explanations

3.1.1 Feature Importance with XGBoost

The feature importance of the XGBoost model is assessed by measuring the reduction in the loss function contributed by each feature. The most influential features are related to offensive actions such as shots, possession, and passes. Additionally, features related to the home team have more weight, suggesting a home advantage as commonly observed in sports.

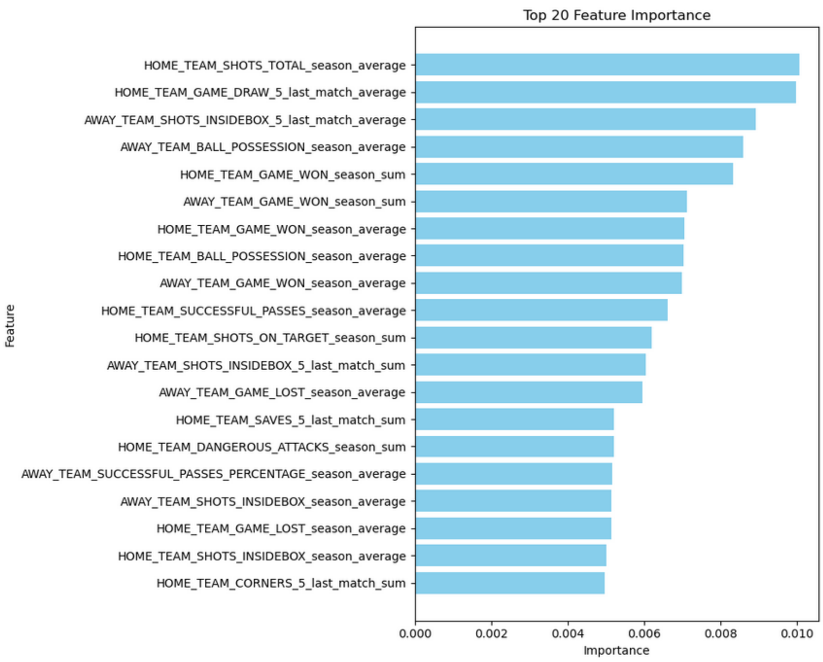

Figure 1: Top 20 Feature Importance based on XGBoost

3.1.2 SHAP

The most impactful feature appears to be the total shots of the home team (indicating how many chances they create), with the number of matches lost by the away team also standing out. The other features seem to have a more tightly clustered importance.

The results from XGBoost and SHAP highlight key characteristics that influence the model's predictions, with both similarities and notable differences. Both methods agree on the importance of the home team's total shots, emphasizing its crucial role in the predictions. However, XGBoost focuses more on global aspects like recent draws and ball possession, while SHAP provides a more detailed view, showing how each feature influences the different outcome classes. Together, these methods offer a comprehensive understanding, with XGBoost providing an overview and SHAP offering a granular analysis of specific impacts.

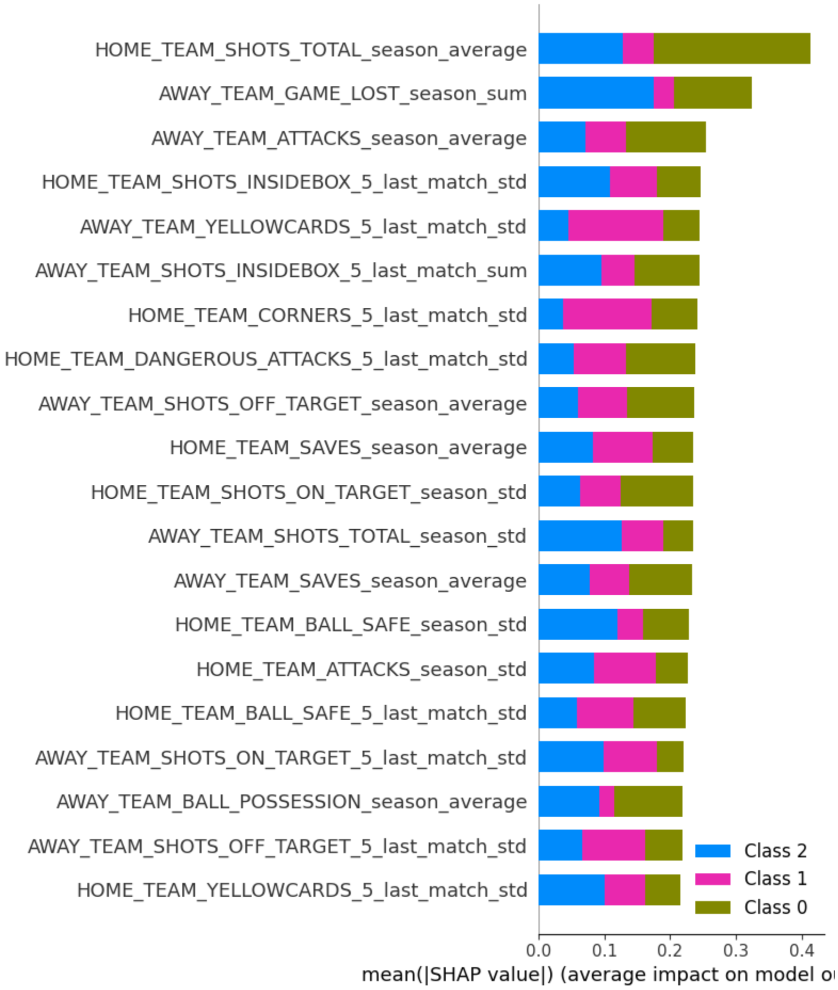

Figure 2: SHAP Summary Plot for Model Explanation

3.2 Local Explanations

Using LIME, we evaluated the model on specific instances to understand the local behavior. For example, on a single prediction, LIME shows how the home team's shots and the away team's losses influence the final prediction.

We iterated this operation on the 50 most important features, giving us the following figure.

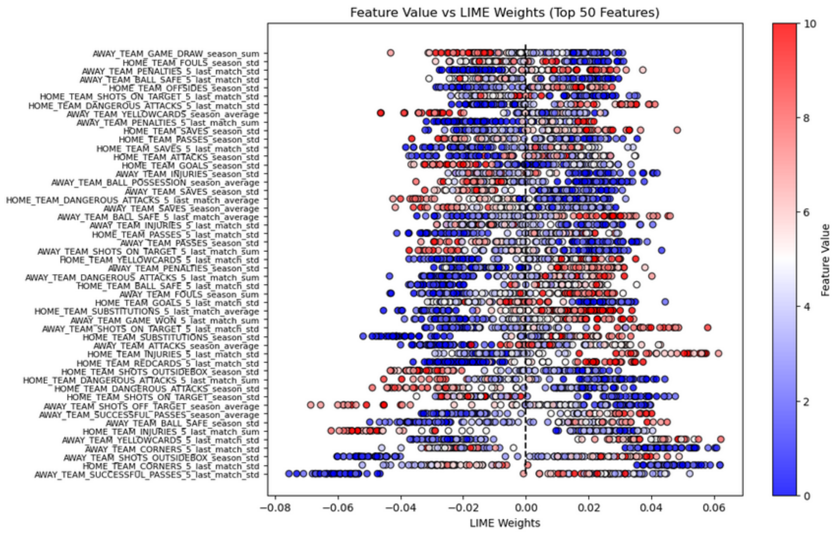

Figure 3: Feature Value vs LIME Weights (Top 50 Features)

4 Discussion

The results from XGBoost, SHAP, and LIME provide deep insights into the model's decision-making process. The consistent importance of offensive statistics such as shots and possession supports the notion that these actions are pivotal in determining match outcomes. The influence of home-team statistics also reinforces the well-known home advantage in sports.

However, there are some biases inherent in the dataset, such as the imbalance in the number of home versus away games. These biases can lead to overfitting, particularly with the home-team features. This can be mitigated by improving data preprocessing and balancing the dataset.

Future improvements could include using other XAI techniques like partial dependence plots to examine interactions between features or experimenting with different model architectures to reduce the model's dependence on certain features.

5 Conclusion

In this report, we applied several explainability techniques (SHAP, LIME, feature importance) to analyze the predictions of an XGBoost model for sports result prediction. These methods allowed us to uncover the most influential features, such as offensive actions and home-team statistics, and provided valuable insights into the model's behavior. Despite some model biases, these techniques offer a powerful tool for enhancing model interpretability and trustworthiness.