BasketXplainer

Interactive Dashboard for Interpretable Winning Odds Prediction

Dustin Brunner[1], Jonathan Koch[2], Timothé Laborie[1], Liule Yang[3]
[1]D-INFK, ETH Zurich, CH, [2]D-ITET, ETH Zurich, CH, [3]D-MATH, ETH Zurich, CH

1 Introduction

Applying analytics to sports is nothing new, but the advent of more complex models that are often based on machine learning (ML) has added a new layer of sophistication to this area of work.
Since the main user group of sports analytics usually does not have a technical background, it is paramount to present ML based results in an understandable and logical way. With this newly developed dashboard, we aim to do exactly that in the realm of forecasting match results in basketball based on previous games.

2 Users & Tasks

The typical user of our dashboard is an analyst or coach at a professional basketball team. While their technical knowledge is usually limited, they have a lot of domain knowledge and their own ideas and understanding of basketball tactics. For us it was thus important to build a tool that would satisfy the following requirements:

1. **Easy usability** without the need for technical knowledge to get started.
2. The **ability to experiment** and play around with features to test out hypotheses.
3. **Intuitive explanations** of ML based results where appropriate.

Fran Camba Rodríguez, data analyst at Obradoiro CAB

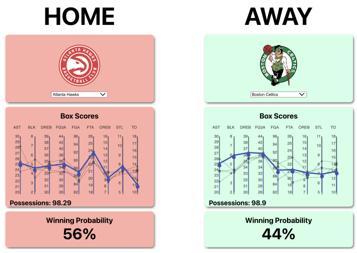
With Fran Camba Rodriguez we had an actual data analyst of a professional Spanish basketball team available who gave us valuable feedback and was eager to get access to the tool we built. The typical tasks for using the tool are the preparation of games and scouting of opposing teams.

3 Data & Features

The dataset used for the project is the NBA Kaggle dataset [2]. We relied on averaged box score statistics of teams as representations of their playing style for predicting the outcomes of future games. The box scores of a team contain the following features: Assists (**AST**), Blocks (**BLK**), Defensive Rebounds (**DREB**), 3-Point Field Goal Attempts (**FG3A**), Field Goal Attempts (**FGA**), Free Throw Attempts (**FTA**), Offensive Rebounds (**OREB**), Steals (**STL**), and Turnovers (**TO**). The original dataset also contained the 3-Point Field Goals Made, Field Goals Made and Free Throws Made, which we intentionally excluded as we considered their respective "Attempts" counterparts to better reflect the playing style of a team.

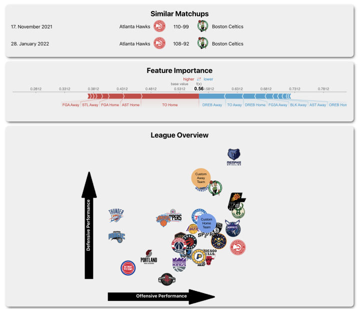
4 Dashboard & Workflow

Our interactive dashboard is meant to guide the user from top to bottom through the whole analysis workflow. After first selecting two teams their respective box score statistics are displayed in an interactive parallel coordinates plot and the predicted winning odds are shown below.



Top components of the dashboard: Team Selection, Box Score Statistics and Winning Probability.

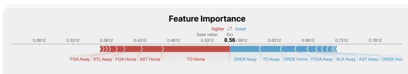
For further analysis of the prediction, the user can use the three bottom components of the dashboard. The first component displays similar previous matchups, the second component highlights the feature importance of all features on the prediction using a SHAP [3] force plot and the last component contains a league overview indicating the offensive and defensive performance of all teams. Finally, the box scores can iteratively be adjusted and new predictions will be computed in order to develop a new strategy for the next games.



Bottom components of the dashboard: Similar Matchups, Explainability (SHAP) and League Overview.

5 Modeling & Explainability

Modeling: For predicting the winning odds based on the aggregated box score statistics of both teams we use a LightGBM (Light Gradient-Boosting Machine) model [1]. LightGBM achieves state-of-the-art performance on tabular classification tasks by utilizing a tree-based learning algorithm with a focus on efficiency and speed which makes it perfectly suitable for our task.
Explainability: In order to help the user comprehend the prediction of the model better we utilize a SHAP (SHapley Additive exPlanations) force plot that highlights the contribution of each feature to the final predicted probability.



SHAP force plot indicating the contribution of features to the prediction.

References

[1] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[2] Nathan Lauga. Nba games data, 2022.

[3] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.