



IDS 2023 Project

Business Understanding (HW10 task 2)

04.12.2023

—

Erko Olumets

Business Understanding

Background

- In the context of sports betting, there is a growing interest in leveraging advanced statistics to improve predictive models and inform betting strategies.
- Expected goals (xG) and expected goals against (xGA) are becoming key indicators in football analytics, providing insights into team and player performance.
- Usage of historic team performance is useful, but looking only at goals scored or goals allowed is not too informative, because of small sample size and high randomness.

Business Goals

- Develop a predictive model that utilizes expected goals (xG), expected goals against (xGA), and post shot expected goals allowed minus goals allowed (PSxG-GA) (a statistic used to show how the goalkeeper prevents goals from opponents) to forecast match outcomes more accurately than historic goals and historic team performance (win/loss/draw).
- Comparing the predictive model outcomes to betting site odds and actual scores to calculate if our model is “profitable” or not (no actual betting will occur, only hypothetical positive or negative return of investment [ROI]).

Business Success Criteria

- **Improved Prediction Accuracy:** Achieve a higher accuracy in predicting match outcomes compared to traditional methods.
- **Profitability in Betting:** Measure the success of the project by its ability to contribute to profitable betting outcomes, ensuring that the predictive model aids in making successful bets.

Situation Assessment

Inventory of Resources

- Main source to provide xG, xGA and PSxG-GA will be <https://stathead.com/fbref/> . This data is open data and free to use for commercial and non-commercial use with the condition of citation to the website and company.

- Computing resource will be my laptop and if need be I can use some cloud computing services, but the need for that is highly unlikely

Requirements, Assumptions, and Constraints

- Assumptions:
 1. A predictive model using xG, xGA and PSxG-GA will be more accurate than historic goal data in predicting football matches
 2. The best predictive model will be random forest classifier.
- Using only English Premier League games will constrain the ability to predict different leagues, because data covered is only from one league.
- Because we are comparing data to betting sites we have to be careful of different betting site rules. We will not be making any bets, just using odds to compare predictions.

Risks and Contingencies

- Data quality issues, inaccuracies or inconsistencies in expected goals, expected goals against, and PSxG-GA.
- Risks linked to the unpredictability of football matches and the inherent uncertainty in betting outcomes.

Terminology

- Establish a standardized terminology for key metrics, including expected goals (xG), expected goals against (xGA) and post shot expected goals allowed minus goals allowed (PSxG-GA). Create a glossary for all specialized language.

Costs and Benefits

- Estimate the costs associated with acquiring and processing detailed football event data, including expected goals, expected goals against, and defensive statistics.
- Successful model might lead to improved accuracy in predicting match outcomes, enhanced betting strategies leading to more profitable outcomes.

Defining Data-Mining Goals

Data-Mining Goals

- **Feature Selection and Engineering:** Identify and select relevant features from expected goals (xG), expected goals against (xGA), and defensive stats that significantly contribute to predicting football match outcomes.
- **Betting Strategy Optimization:** Leverage predictive insights to optimize betting strategies, considering factors such as odds, historical performance, and team strengths.

Data-Mining Success Criteria

- **Model Accuracy and Reliability:** Achieve a high level of accuracy in predicting football match outcomes, validated through comparison against historical match results.
- **Profitability in Betting:** Measure the success of the project by its ability to contribute to profitable betting outcomes. Evaluate the return on investment (ROI) for betting strategies based on model predictions.