14 January 2026 ·

Qeios

Peer Review

Review of: "Accuracy Without Profit: A Statistical Evaluation of Machine Learning Profitability in the English Premier League"

Mustafa Feza Tahmaz[1,2]

1. Faculty of Economics and Administrative Sciences, Beykent University, Turkey; 2. Faculty of Economics, İstanbul University, Turkey

The subject of the paper is very interesting, and I find the topic to be timely and relevant to my expertise. It seems to be written with a good grasp of the previous literature on what methods to apply, what questions to ask, and how to validate tests, which show rigor and promise. However, I would like to point out a few recommendations from my part.

Primarily, the reporting seems a little messy, and some components of the paper seem underexplained. The methodology could have been made clearer with a conceptual model of the paper. I also recommend a unified evaluation table demonstrating the use of the same out-of-sample metrics under the same walk-forward folds, reporting multiclass logloss, multiclass Brier score, and ECE for all models across all folds.

Secondly, the feature selection of Points, Goals Conceded, Shots, Shots on Target, and Corners may be too coarse for the models to capture desired tactical match patterns essential for classification. The logic behind the use of rolling averages is good, as it restricts data leakage. Given the acknowledged feature limitations, if it was possible, adding more sensitive metrics or strength proxies such as ELO Ratings, Poisson/Bayesian hierarchical strength ratings, and xG-metrics to support feature richness and test robustness would have improved the study remarkably.

Third, including market odds in the training data risks the models implicitly learning and may lead to a noisy transformation of the odds rather than adding independent signals. This concern is relevant in the SHAP results and the confusion matrices, which suggest indications of home team bias and draw prediction bias in the results. This may help explain why all models had limited ROI. Additionally, in my

 doi.org/10.32388/881LR9

opinion, the DM test alone isn’t sufficient to demonstrate meaningful distinction from or superiority to market odds. I recommend also including Betting Odds average stats in the Walk Forward Validation Results table just for easier comparison and adding another type of similarity measure between market-implied probabilities and model probabilities of your choice to demonstrate distinction.

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/881LR9