MATH MODELLING FINAL PROJECT

Charvie Kukreja & Shyam Pratap Singh Rathore

University College Dublin

INTRODUCTION

We built a tool that **predicts who will play well** in each match and then **picks a starting XI** you could actually use. It combines two types of models-**XGBoost** and a **neural network**-and averages their confidence to get a more reliable score. We tested it in time order (like the real season) and also week-by-week. The combined model ranked players better and was more consistent than either model alone (ROC-AUC ≈ 0.79, PR-AUC ≈ 0.68). We choose the decision threshold using **F1**, check probability **calibration**, and run a small optimiser to enforce formation and "max three per club." **Form** and **opponent strength** clearly help.

OBJECTIVE

Our objective is to analyse Premier League data (2021–2023) using a hybrid of XGBoost and Neural Networks. By combining both models, we aim to outperform each individually and uncover the key factors that drive goal success.

DATA & FEATURES

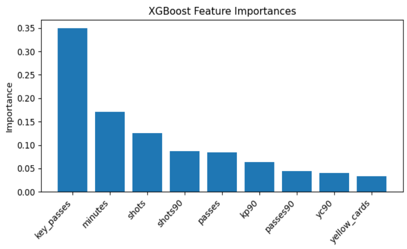
Data: EPL 2022/23 player-match rows (**3555**; **626** players; 20 clubs); keep minutes>0; numeric stats coerced; missing counts→0.

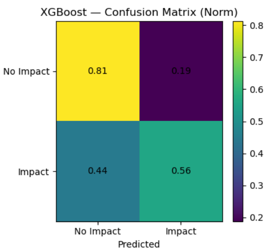
Features: player volume/quality (minutes, goals, assists, shots, passes, key_passes, cards + per-90), **rolling form** (3-match means), and **context** (team_strength_pre, opponent_strength_pre).

Leakage control: all rolling/expanding features are **shifted** so only pre-match information is used.

MEHODOLOGIES

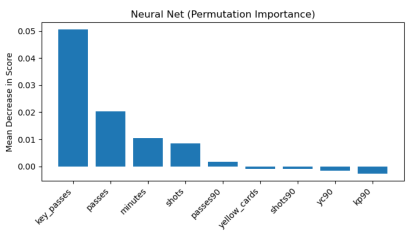
ML (XGBoost) — "Tabular boss"

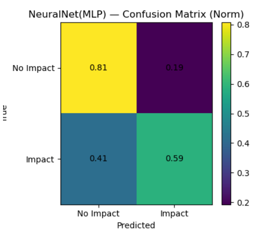


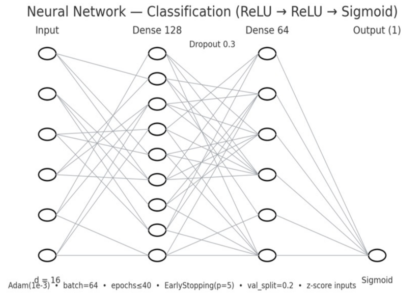


XGBoost serves as our strongest baseline for tabular football data, ranking features by their predictive power. The model highlights *key passes* and *minutes played* as the most influential drivers of player impact, while shots and passing metrics contribute moderately. Disciplinary variables (yellow cards) had negligible predictive value. The confusion matrix reveals that XGBoost is highly reliable at detecting 'no impact' games (81% accuracy) but struggles more with correctly identifying impactful performances (56%). This imbalance underlines both the strengths of tree-based methods in structured data and their limitations in capturing complex, non-linear player dynamics — motivating the use of deep learning and hybrid ensembles.

NN (Keras MLP)

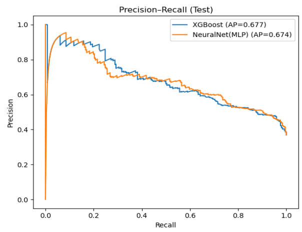


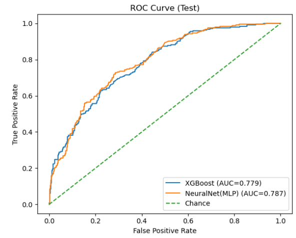


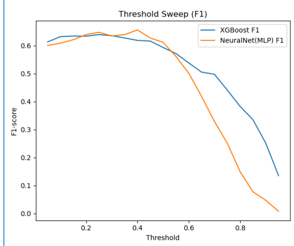


e neural network adds a different angle mpared to XGBoost. Instead of relying ly on structured splits, it picks up on ore subtle patterns in the data. We see that *passes* and *overall passing* still drive ost of the signal, but the NN spreads portance across features in a way that nts at deeper interactions. It's a bit better spotting when a player truly has an pact (higher recall), though it sometimes sfires with extra false positives. In short, NN doesn't just rank features — it rns hidden connections that make the alysis richer."

COMPARISON







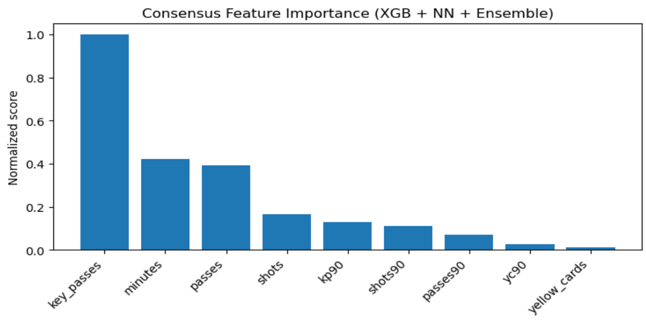
Feature Importance: XGBoost highlights *key passes, minutes, and shots* as the most decisive factors, while the Neural Network spreads its focus more evenly. This shows the tree model is sharper on obvious stats, while the NN picks up subtle signals.

Threshold Sensitivity: When we adjust the decision threshold, XGBoost drops off faster, whereas the Neural Net stays more stable. This means the NN is slightly more reliable when we raise the bar for what counts as "impact."

Precision–Recall Tradeoff: Both models handle the imbalance between impact vs. no-impact cases reasonably well, with XGBoost holding a tiny edge. The curves confirm how tough it is to predict rare impactful performances.

ROC Curve: The Neural Net edges ahead here, showing a slightly better ability to separate impactful players from non-impactful ones. Both models clearly perform better than random guessing.

RESULTS



XGBoost's strength: excels at structured signals like *key passes* and *minutes played*, but can overlook subtle feature interactions.

Neural Network's strength: better at uncovering hidden, non-linear patterns, though less consistent on straightforward cases.

Ensemble effect: by averaging both, we reduce false positives from the NN and improve recall compared to XGBoost.

Result: more stable accuracy across thresholds and more reliable insights for Fantasy Football player selection.

Takeaway: the ensemble doesn't just improve metrics — it delivers predictions that are consistent and practically useful.

REFERENCES

- **Atta Mills et al. (2024)** — Machine and deep learning for soccer predictions https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-01008-2
- **Wong (2025)** — Predictive framework for soccer using ML/AI models https://www.sciencedirect.com/science/article/pii/S2772662224001413
- **Moya (2025)** — Neural networks for football player performance optimization https://www.mdpi.com/2504-4990/7/3/85
- **Ouyang et al. (2024)** — XGBoost + SHAP for real-time sports predictions https://pmc.ncbi.nlm.nih.gov/articles/PMC11265715/
- **Rane (2024)** — Review of ensemble learning methods in ML/DL https://sabapub.com/index.php/SMHS/article/download/1225/631/4974
- **Bunker & Thabtah (2022)** — Survey of machine learning approaches in team sports prediction https://jair.org/index.php/jair/article/download/13509/26786/30289
- **Granitto et al. (2005)** — Neural network ensembles: evaluation of accuracy and diversity https://arxiv.org/abs/cs/0502006