Fantasy Football AI: Predictive Modeling and Squad Optimization

Anekha Kallath Giyo Varghese

University College Dublin

Overview

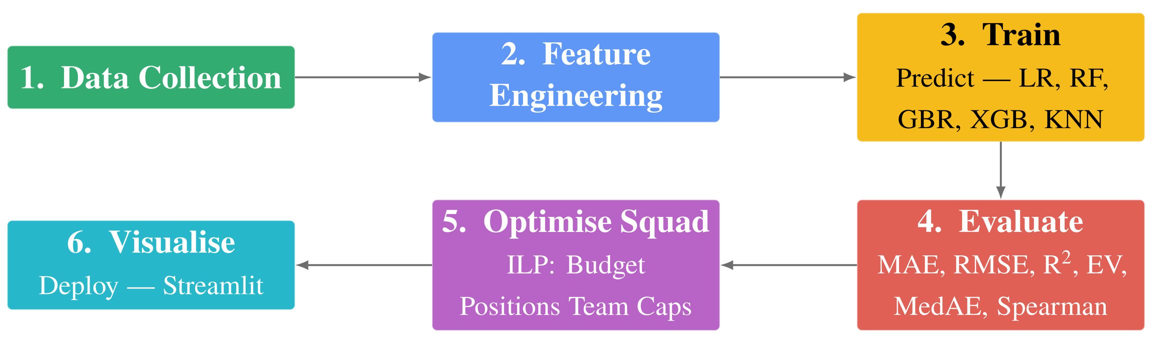
This project delivers an AI-driven Fantasy Premier League (FPL) assistant that:

- Predicts upcoming player points using historical data and engineered features
- Optimises a valid 15-player squad under FPL rules (budget, positions, formations)
- Deploys an interactive Streamlit dashboard for model evaluation, feature analysis, and squad selection
- Integrates machine learning models with integer programming for accurate, rules-compliant team recommendations

Objectives

1. Predict FPL points using Linear Regression, Random Forest, Gradient Boosting, XGBoost, and KNN; also provide an ensemble of tree models.
2. Evaluate models with MAE, RMSE, R^2, Explained Variance, Median AE, and rank correlation.
3. Optimise squad selection using Integer Linear Programming (ILP) via PuLP.
4. Enable captain/vice selection and minutes-based risk filtering.
5. Provide visual outputs: radar plots, residuals, calibration, per-position metrics, and top-K overlap.

Methodology



Standard Constraints:
Budget: 100 value points
Squad: 2GK, 3FWD, 5DEF, 5MID
Min. Spend: 95%

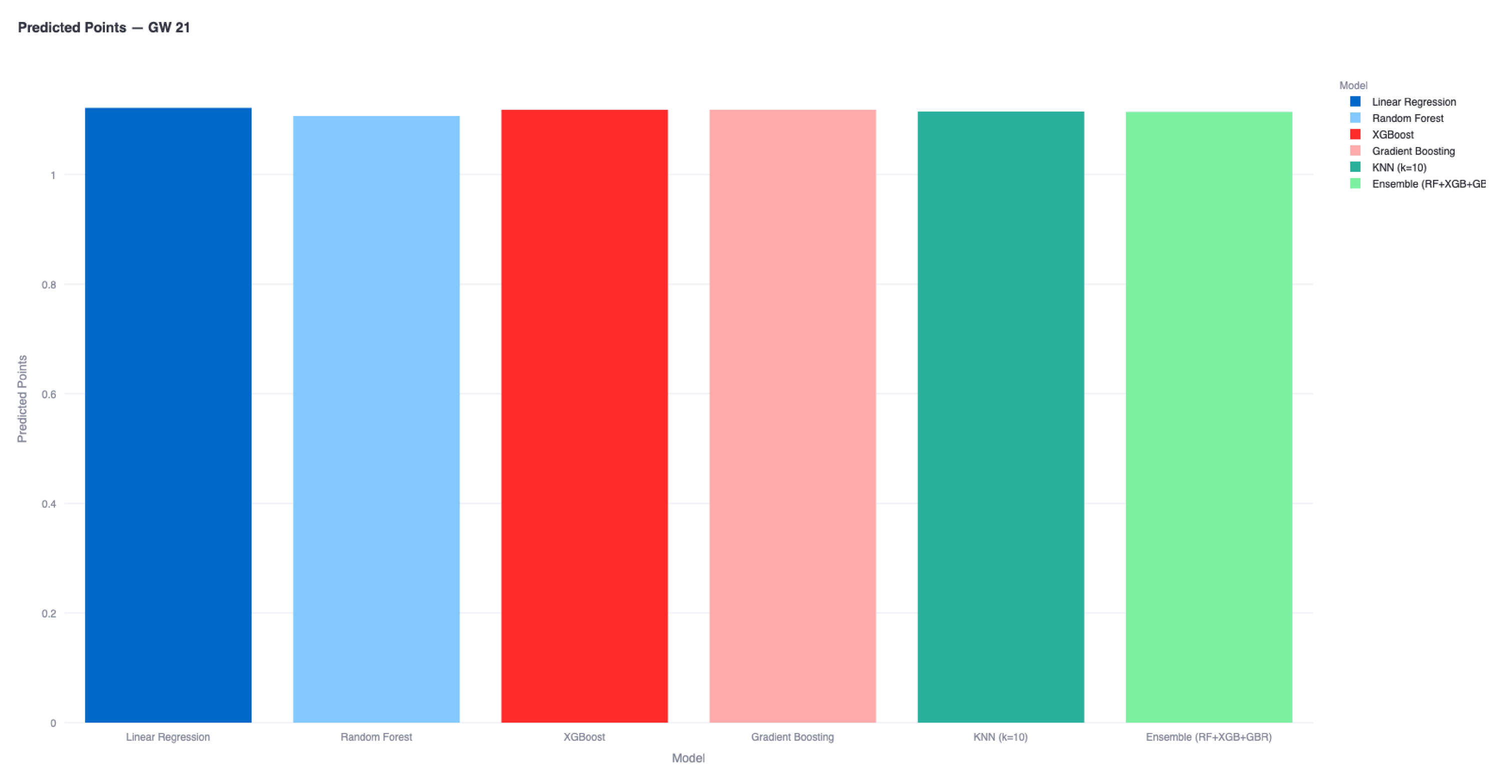
Optimized Squad from the Best Model: XGBoost

Name	Team	Position	Pred. Points	Cost (M)
Tyrone Mings	Aston Villa	DEF	6.739490986	£4.4M
Pervis Estupiñán	Brighton	DEF	7.683516502	£5.0M
Marc Guéhi	Crystal Palace	DEF	14.92946053	£4.5M
Maxence Lacroix	Crystal Palace	DEF	7.724563122	£4.5M
Dan Burn	Newcastle	DEF	7.199381351	£4.4M
Ollie Watkins	Aston Villa	FWD	7.50861454	£8.9M
Jean-Philippe Mateta	Crystal Palace	FWD	7.941020489	£7.2M
Alexander Isak	Newcastle	FWD	17.34338379	£9.3M
Emiliano Martínez Romero	Aston Villa	GK	7.498740196	£5.0M
Martin Dúbravka	Newcastle	GK	11.16487789	£4.2M
Mitoma Kaoru	Brighton	MID	8.787881851	£6.4M
Cole Palmer	Chelsea	MID	8.813951492	£11.4M
Alex Iwobi	Fulham	MID	14.61301804	£5.9M
Phil Foden	Man City	MID	14.20124054	£9.2M
Amad Diallo	Man Utd	MID	21.73747444	£5.4M

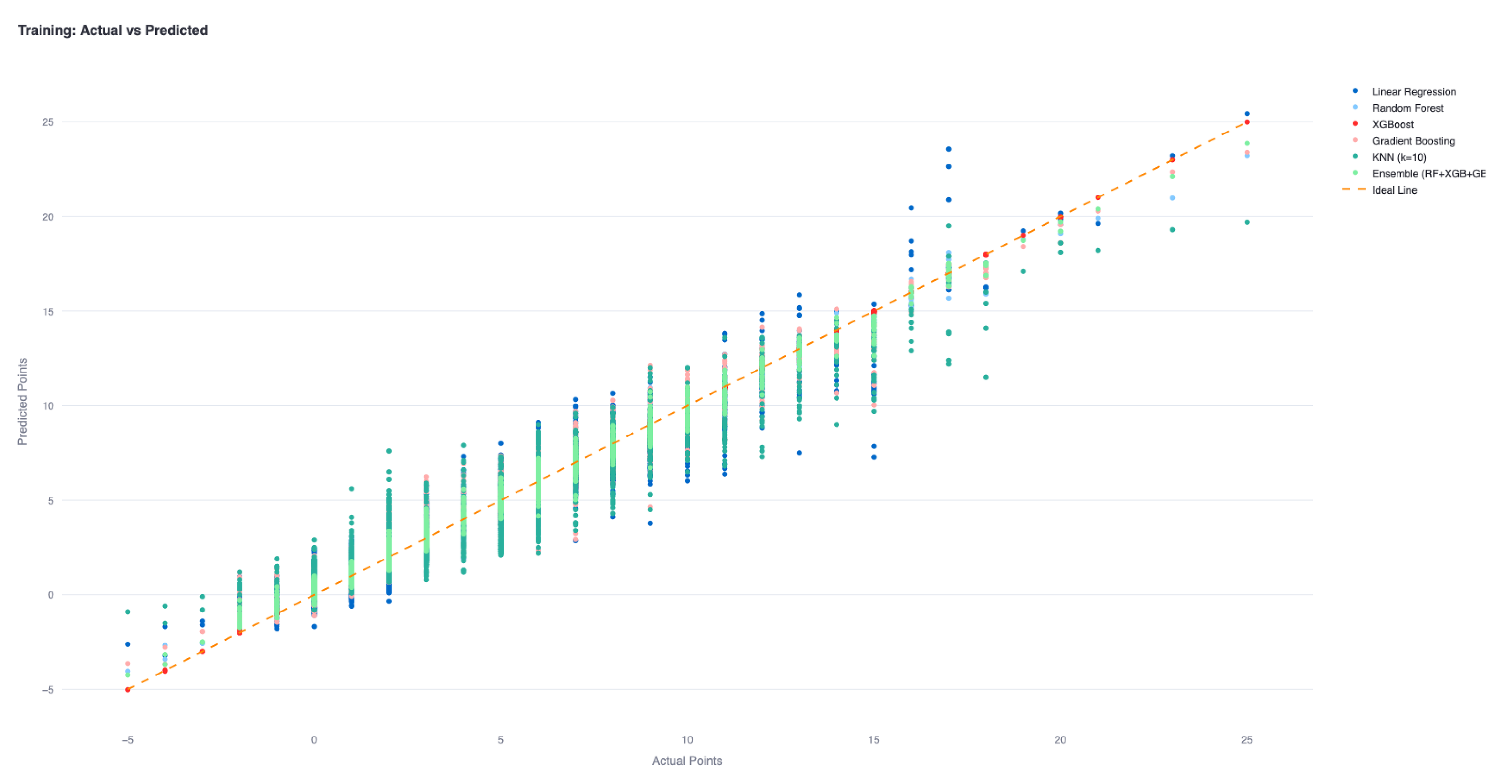
Why XGBoost Outperforms:

- **Lowest MAE**: 0.0287 – indicates highly accurate point predictions
- **Lowest RMSE**: 0.0687 – minimal large prediction errors
- **Highest R^2**: 0.99915 – explains almost all variance in player scores
- **Highest Explained Variance**: 0.99915 – extremely strong model fit
- **Lowest Median Absolute Error (MedAE)**: 0.0000675 – very precise typical predictions
- **Strong Ranking Ability**: Spearman 0.979 and Top-K Overlap 0.90 – excellent at ranking players for squad selection

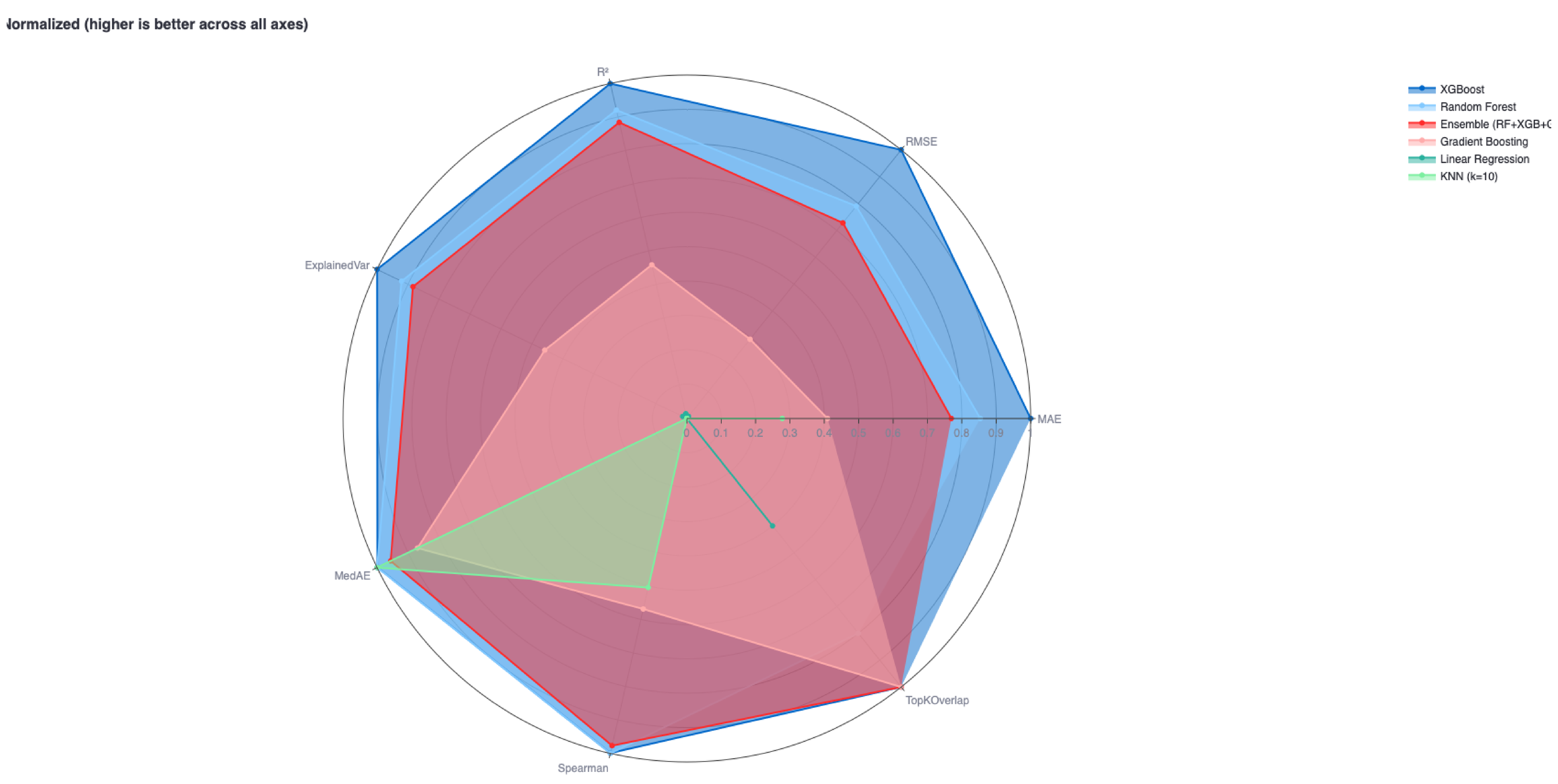
Predicted Points for Next Week



Actual vs Predicted Points



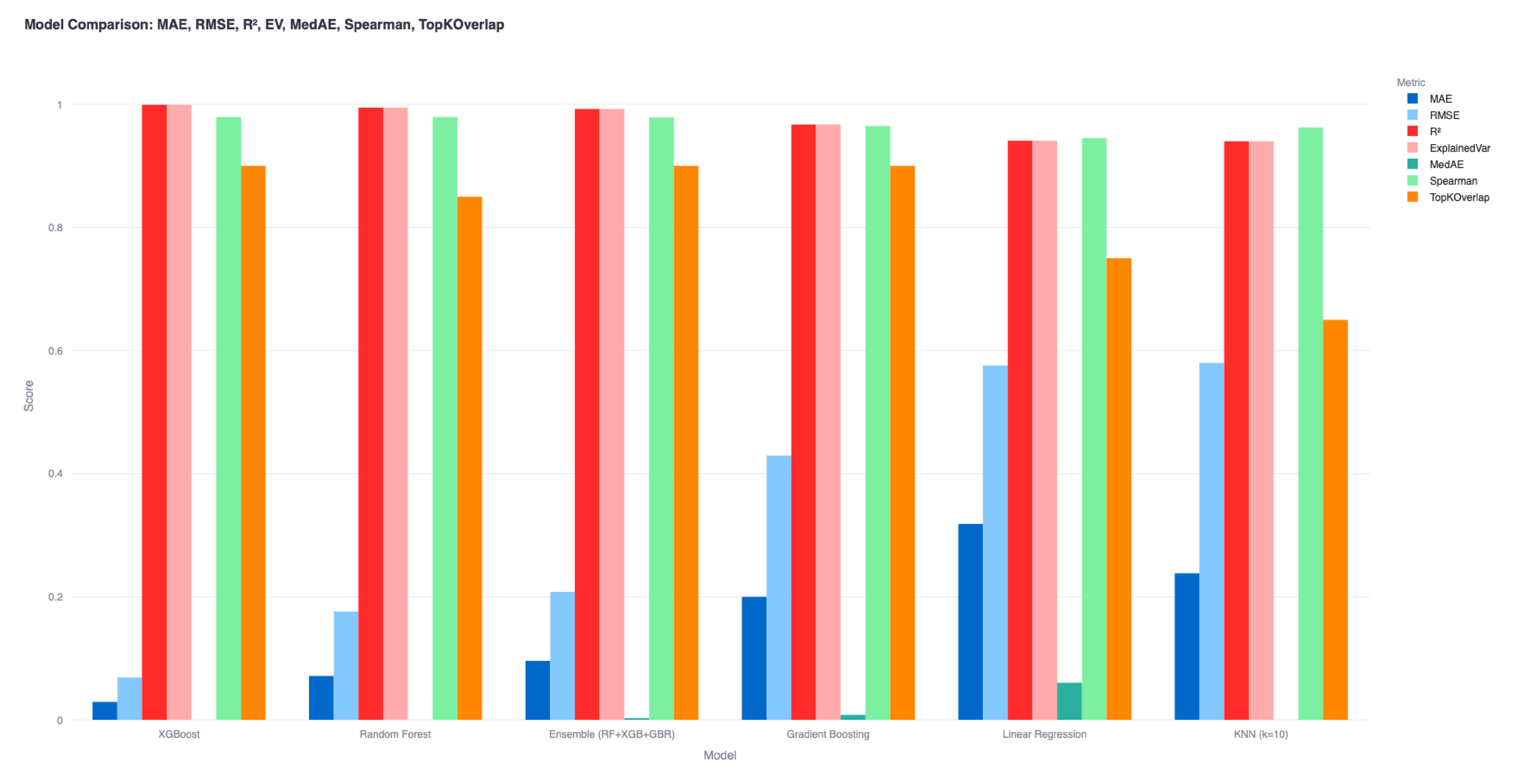
Model Performance Metrics: Radar Chart



Validation metrics for Linear Regression, Random Forest, Gradient Boosting, XGBoost, and KNN (ensemble available).

- Compares models across normalized metrics (higher is better)
- XGBoost & Random Forest lead in accuracy, variance explanation, and ranking
- Ensemble performs competitively across most metrics
- Linear Regression & KNN trail significantly

Model Error Comparison



Metrics Comparison Table

Model	MAE	RMSE	R^2	Expl. Var.	MedAE	Spearman	Top-K Overlap
XGBoost	0.02873	0.06875	0.99915	0.99915	0.0000675	0.97915	0.90
Random Forest	0.07107	0.17572	0.99448	0.99448	0.0000000	0.97919	0.85
Ensemble (RF+XGB+GBR)	0.09546	0.20763	0.99229	0.99229	0.0026275	0.97840	0.90
Gradient Boosting	0.19980	0.42890	0.96709	0.96709	0.0078149	0.96440	0.90
Linear Regression	0.31827	0.57553	0.94074	0.94074	0.0599263	0.94486	0.75
KNN (k=10)	0.23793	0.57952	0.93991	0.93994	0.0000000	0.96219	0.65

- **XGBoost** – Best overall (lowest MAE/RMSE, highest R^2 & rank metrics)
- **Random Forest** – Competitive accuracy, slightly higher errors than XGBoost
- **Ensemble (RF+XGB+GBR)** – Strong ranking performance, small trade-off in errors
- **Gradient Boosting** – Good balance, but weaker than XGBoost/RF
- **Linear Regression** – Higher errors, weaker ranking alignment
- **KNN (k=10)** – High MAE/RMSE, decent R^2, but weakest ranking overlap

Conclusions

- Tree-based models (XGBoost, Random Forest) deliver the highest accuracy and ranking consistency
- Low error rates (MAE, RMSE) with high R^2 and Explained Variance ensure robust predictions
- Strong rank alignment (Spearman, Top-K overlap) supports reliable top-player identification
- Well-calibrated predictions with minimal deviation from the ideal line
- Stable next-week forecasts across models enable confident, data-driven squad optimisation

References

[1] Vaastav Anand. Fantasy Premier League Data Repository. `https://github.com/vaastav/Fantasy-Premier-League`, 2024. Accessed: 2025-06-13.

[2] Malhar Bangdiwala, Rutvik Choudhari, Adwait Hegde, and Abhijeet Salunke. Using ML models to predict points in Fantasy Premier League. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–6. IEEE, 2022.

[3] IBM. IBM Watson Fantasy Football. `https://www.ibm.com/thought-leadership/fantasy-football/index.html`, 2024. Accessed: 2025-06-13.

[4] Prof. Dr. S. N. Sarda, Rishikesh Sahu, Atharva Ingole, Aditi Patil, and Akansha Tarpe. Fantasy Football Team Prediction Using ML. `https://www.doi.org/10.56726/IRJMETS54188`, 2024. Accessed: 2025-06-13.

[5] Mitchell, S., Mason, A., & OSI Community. PuLP: A Linear Programming Toolkit for Python. https://coin-or.github.io/pulp/