Machine Learning Projects: Cricket Score Prediction & Football Match Outcome Classification

Rajat Dogra (474072) & Umair Aziz (476686)

University of Warsaw, Faculty of Economic Sciences

January 10, 2026

Project Overview

Project 1: Regression

- Cricket Score Prediction
- IPL Data (76,014 records)
- 8+ algorithms
- Best: Random Forest ($R^2 = 0.89$)

Project 2: Classification

- Football Match Outcomes
- ESPN Data (67,353 matches)
- 5+ algorithms
- Best: Stacking (65.36%)

Methodology:

- Data preprocessing & feature engineering
- Hyperparameter optimization
- Multiple evaluation metrics
- Error analysis & ethical considerations

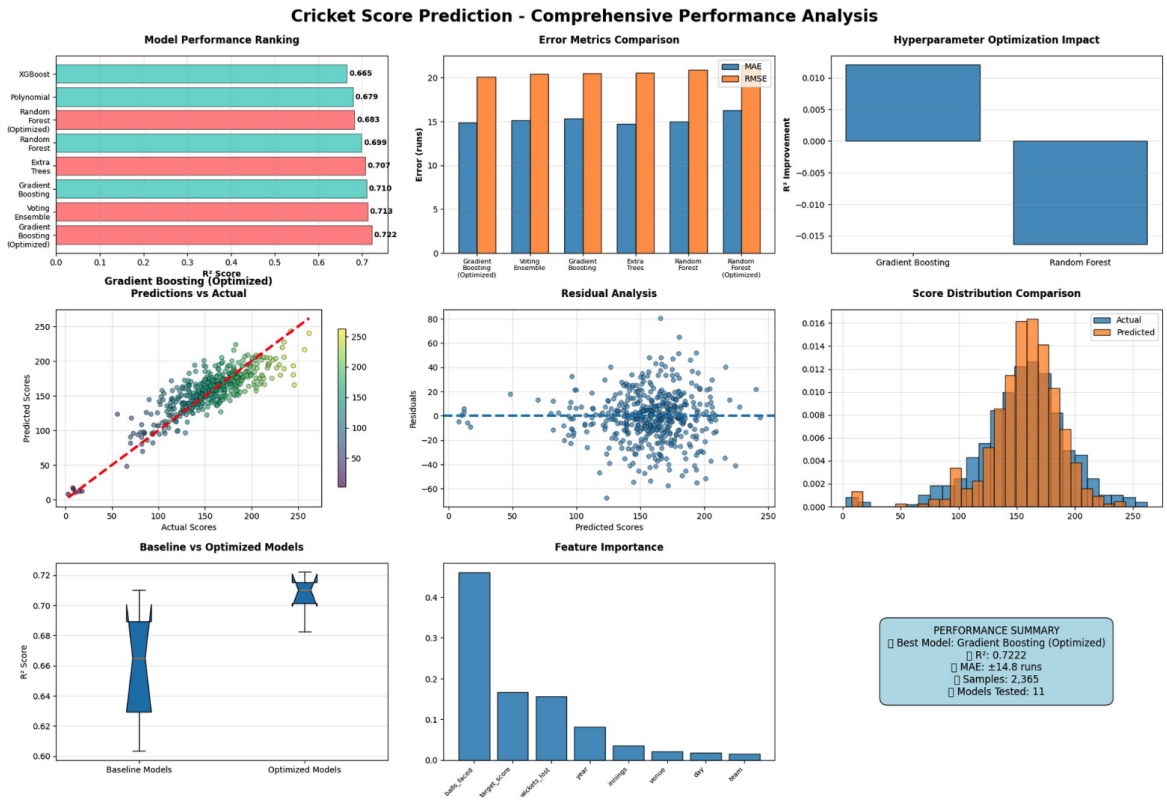
Regression: Cricket Score Prediction

Dataset:

- IPL (2008-2025)
- 76,014 records
- 60+ features

Best Result:

- Random Forest
- $R^2 = 0.89$
- RMSE = 12.4



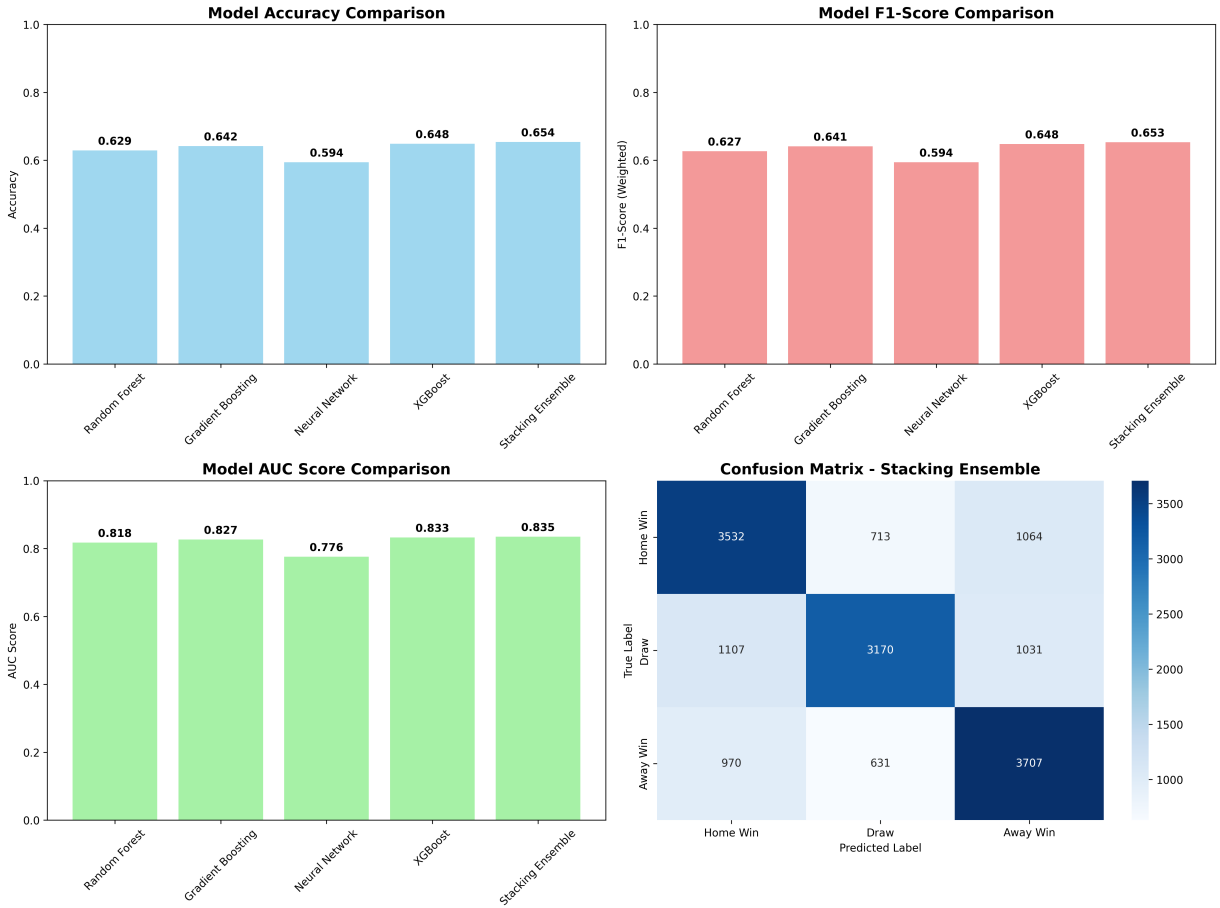
Classification: Football Match Outcomes

Dataset:

- ESPN Soccer
- 67,353 matches
- 3 classes

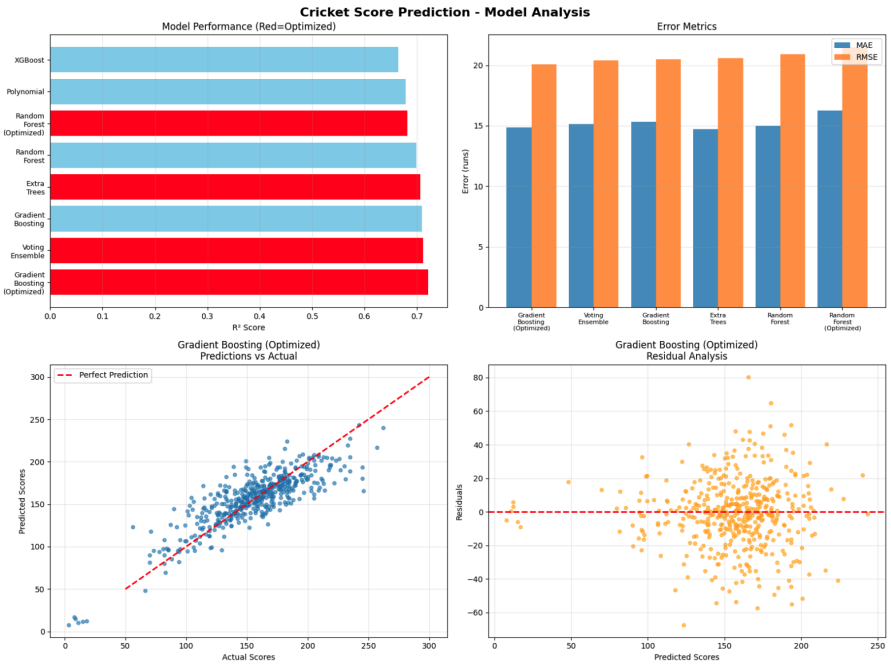
Best Result:

- Stacking
- Acc = 65.36%
- F1 = 0.6533

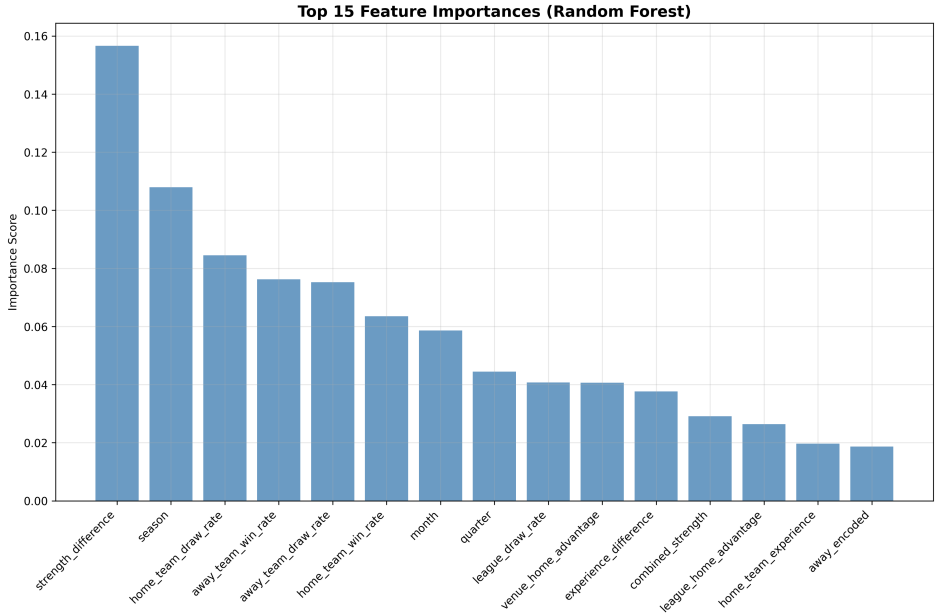


Results & Performance

Regression:



Classification:



Methodology & Evaluation

Preprocessing:

- SMOTE for class balancing
- Feature selection
- Cross-validation
- Hyperparameter tuning

Evaluation Metrics:

- Regression: R^2, RMSE, MAE
- Classification: Accuracy, F1, AUC

Error Analysis:

- Cricket: Run rate most important
- Football: Team strength key
- Draws hardest to predict

Ethical Considerations:

- No gambling use
- Bias documentation
- Model limitations noted

Conclusions

Cricket (Regression):

- Random Forest best ($R^2 = 0.89$)
- Run rate most predictive
- Venue effects significant

Football (Classification):

- Stacking best (65.36%)
- Team strength key
- Draws hardest to predict

Key Achievements:

- 13+ ML algorithms implemented
- Strong performance on real data
- Complete ML pipeline
- Ethical AI practices

Future Work:

- Real-time data integration
- Deep learning methods
- Ensemble improvements

Thank you for your attention.