

English Premier League Player Position Prediction



Kaishuo Zhang
Brown University
Dec 6, 2022
<https://github.com/Kingsley024>



Recap





Problem

- Target Value: Player Positions
- Classification Problem
- It plays an important role in
tactics and game lineup
decision making





Data

- All Time Premier League Player Statistics from Kaggle
- EDA: Many statistics are highly correlated with the target variable.
- Preprocessing: OneHotEncoder, MaxMinEncoder, StandardScaler
- 176 features and 502 data points



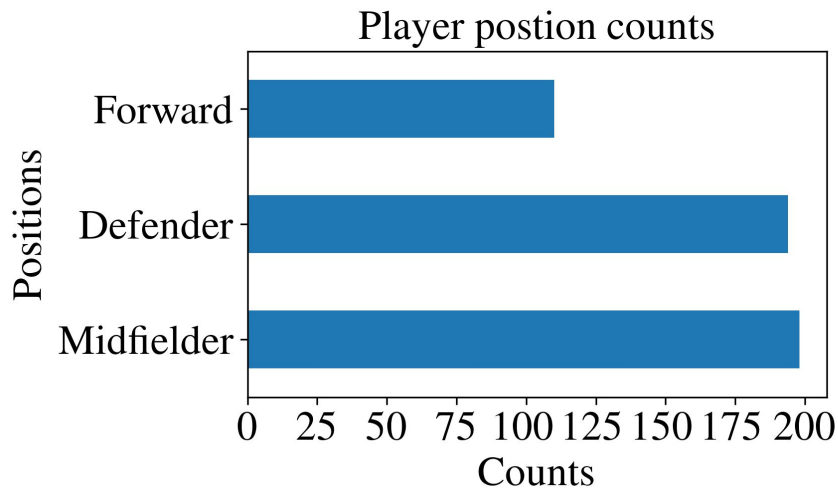
Cross Validation





Splitting

- K fold
- Training, Validation, and Testing contains 60%, 20%, and 20%, respectively
- iid and imbalanced



This graph shows the distribution of players are in each position



Pipeline

- Split
- Preprocessor → Fit model
- GridSearchCV
- Accuracy score



Machine Learning Algorithm

- Random Forest: max depth, n estimators
- K-nearest neighbors: n neighbors, leaf size, p
- Support Vector Machine: gamma, C, degree
- Logistic Regression: C, penalty(l1, l2, elastic net)

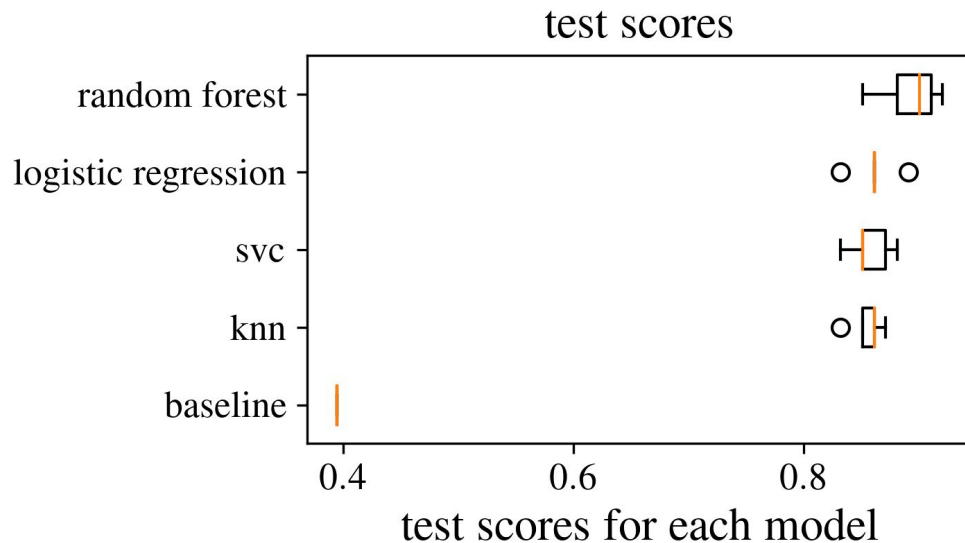
Results



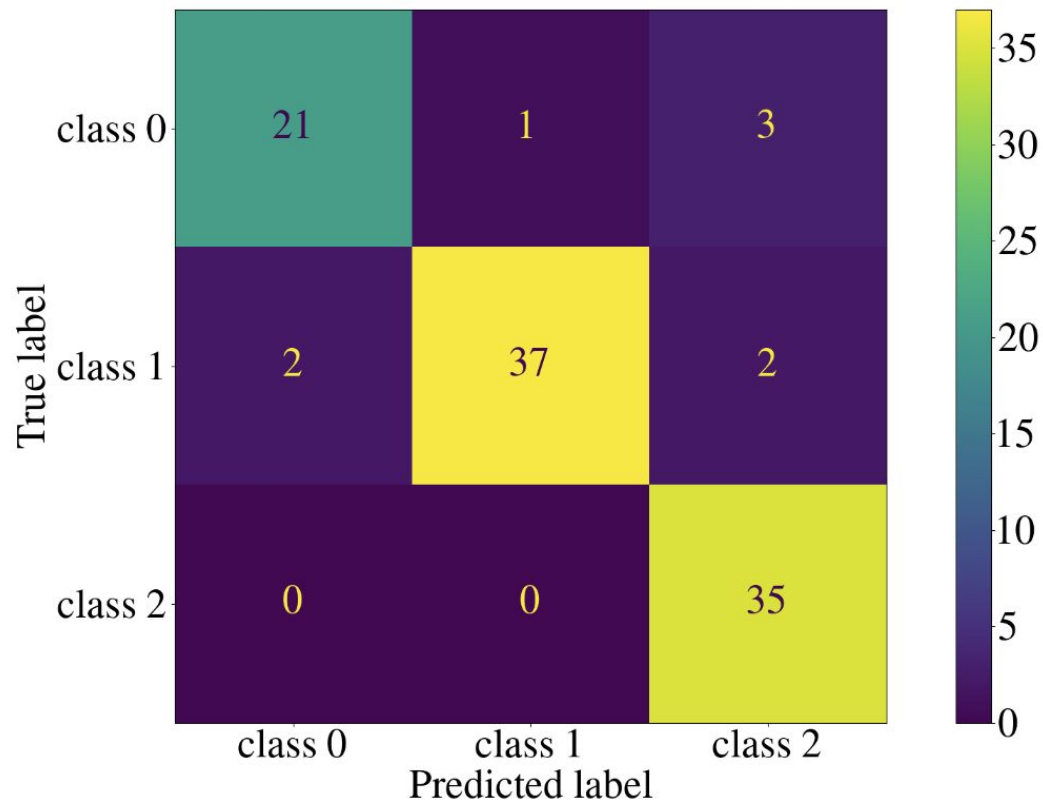


Test Score

- All results are much higher than the baseline score
- Random Forest has the best test score

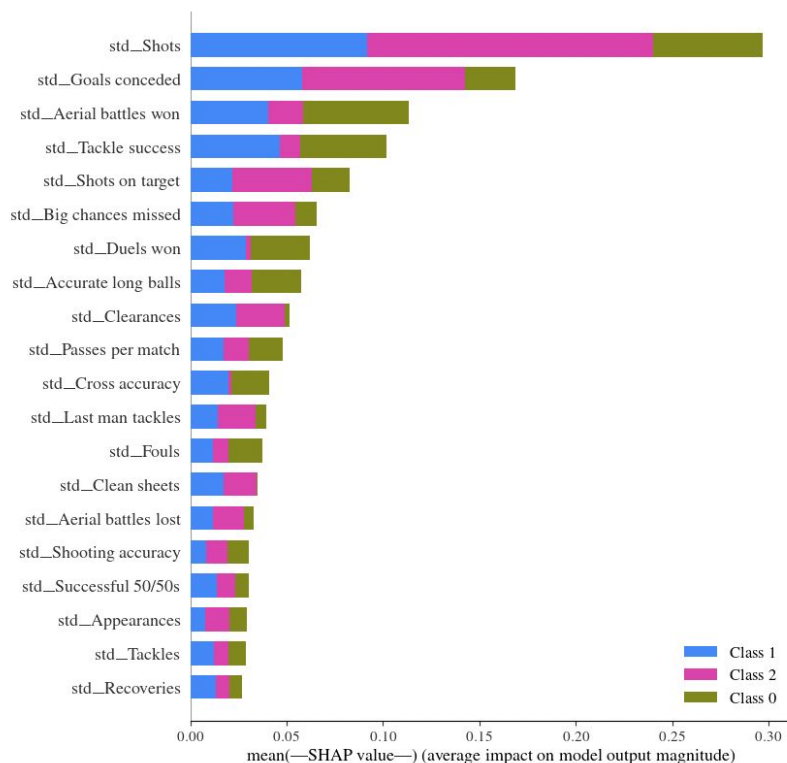
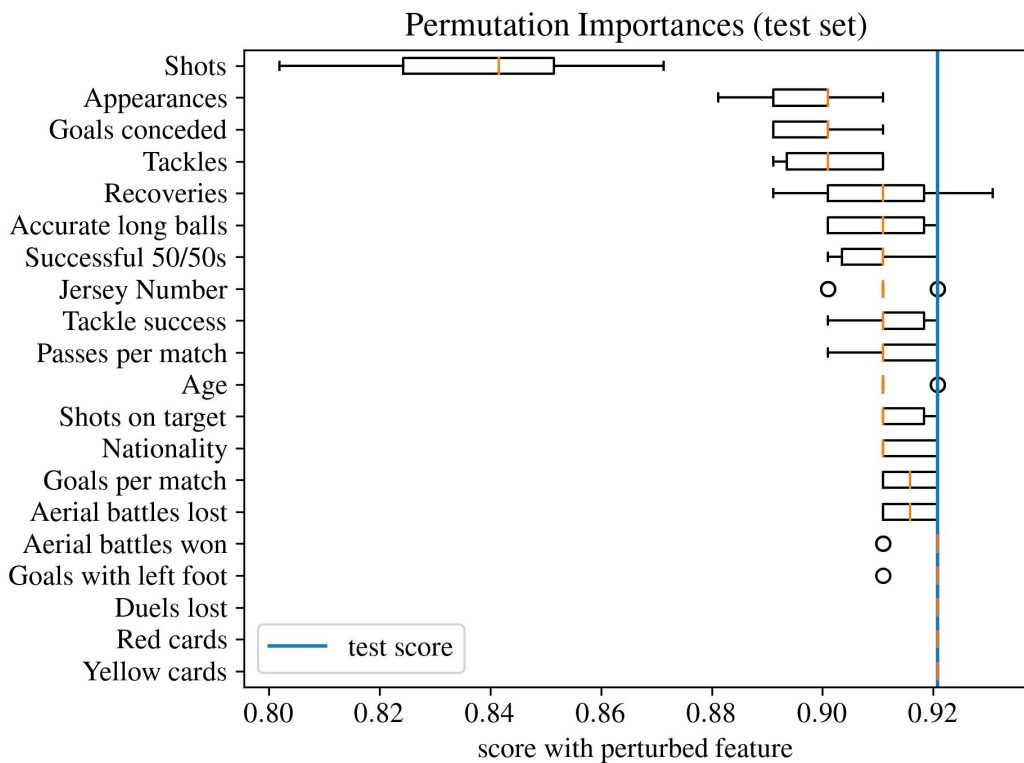


Confusion matrix





Feature Importance-Global



Feature Importance-Local

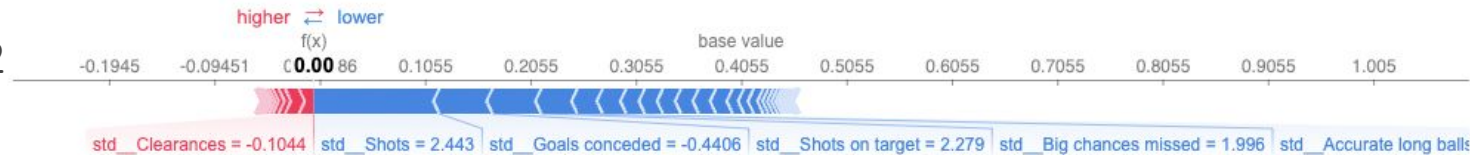
- Class 0



- Class 1



- Class 2



Outlook





Outlook

- Model performance
- Feature importance Analysis



Thank you for listening!