Association Football – Prediction of Team Squads and Positions

By

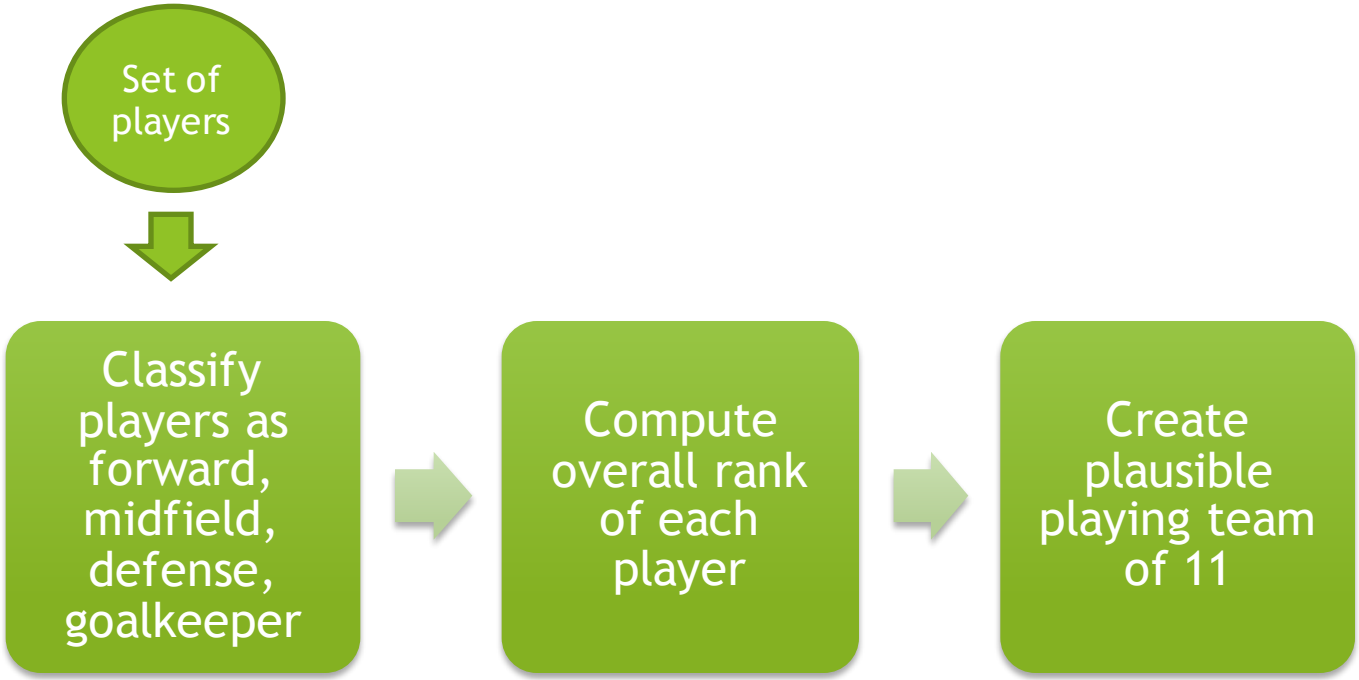
Anand Kamat, Deeksha Arya, Dipanjan Dutta

Motivation

- 250 million players in over 200 countries[1]
- Huge fan following of over 4 billion people[2]
- Statistics of performance of players released, but process of selection of squad confidential
- Football championships give high monetary opportunities as large amounts of bets are placed worldwide on both International matches and league matches
- Great amount of statistical data is studied to predict:
  - best player of a match/tournament
  - team to win a match and ultimately championship
- Though of great interest, not much research has gone into predicting a playing 11 of a team
- Useful especially in fantasy football or video games to generate an opponent team based on easy-medium-hard difficulty level

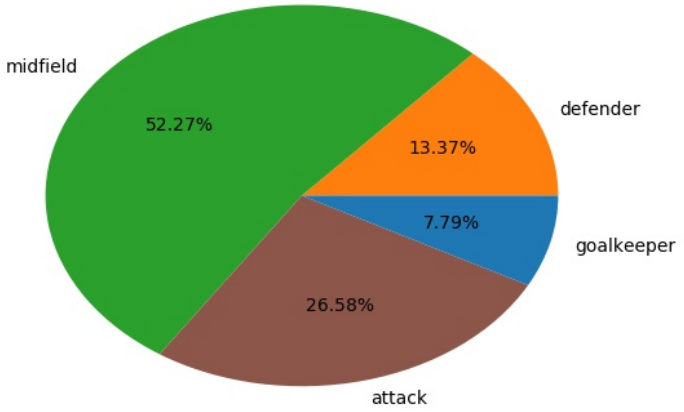
[1] https://en.wikipedia.org/wiki/Association_football
[2] http://www.totalsportek.com/most-popular-sports/

Our Plan


Set of players
Classify players as forward, midfield, defense, goalkeeper
Compute overall rank of each player
Create plausible playing team of 11

The Dataset
European Soccer Database

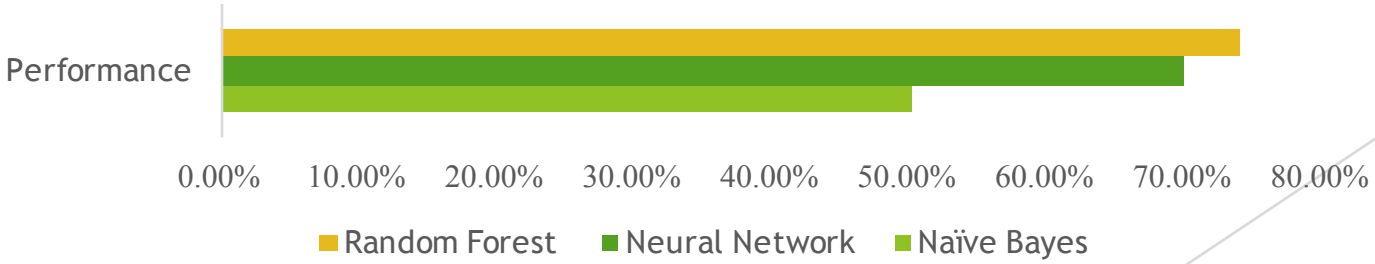
- We focus on association football in Europe
- 10,000+ players with information over the seasons from 2008-2016
- Attributes of players sourced from EA Sports FIFA video game series as of 16th Oct 2016



Percentage of position of players in dataset

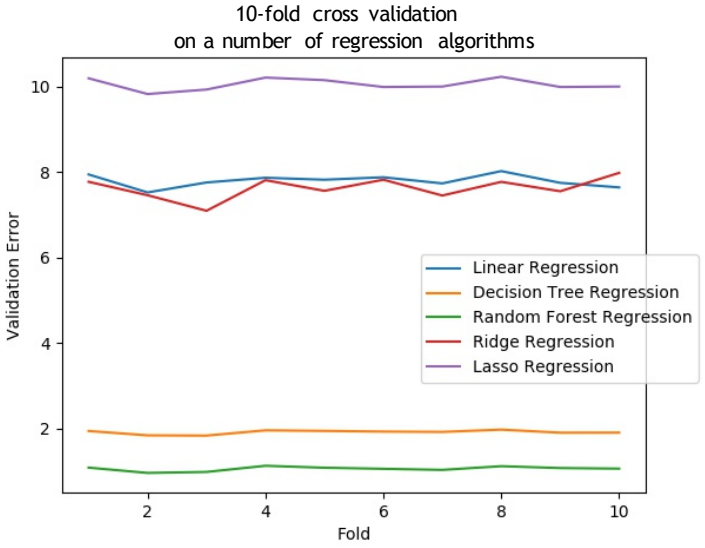
Step 1:
Classifying Players into Position

- 162257 rows of player and their attributes over time
- Classification task to classify a player to be played as one of 4 classes:
  - Attack
  - Midfield
  - Defense
  - GoalKeeper
- 34 features of each player including:
  - crossing, finishing, heading accuracy, short passing, volleys, dribbling, curve, free kick accuracy



Step 2:
Overall Ranking of Players

- Regression task to compute overall rating of player on 100
- 35 features for each player including:
  - potential, ball control, acceleration, sprint speed, agility, reactions, balance, shot power, jumping, strength, long shots and aggression
- Algorithm details:
  - Training: 80%
  - Testing: 20%
  - Error function:
    - Mean Squared Error (MSE)
- MSE on Test Data using Random Forest Regression: 0.94



Our Current Work

- Given a set of features of multiple players, pick the best combination for the playing 11
- Players chosen from:
  - Aaron Ramsey
  - Alex Iwobi
  - Alexandre Lacazette
  - Alexis Sanchez
  - Calum Chambers
  - Danny Welbeck
  - Francis Coquelin
  - Granit Xhaka
  - Hector Bellerin
  - Jack Wilshere
  - Laurent Koscielny
  - Mathieu Debuchy
  - Mesut Oezil
  - Mohamed Elneny
  - Nacho Monreal
  - Olivier Giroud
  - Per Mertesacker
  - Petr Cech
  - Santi Cazorla
  - Sead Kolasinac
  - Serge Gnabry
  - Shkodran Mustafi
  - Taulant Xhaka
  - Theo Walcott

Further Scope of the Project

- Given a team's composition, location of the match and home ground vs away advantage statistics, predict the likelihood of the team winning or losing to its opponent
- Use sentiment from commentary data, Twitter, and news articles to better adjudge a player and his performance
- Use additional features such as previous history of a team to determine the likelihood of a win