Saban Akay, Ioan Gulyas, Ardit Luzi, Tobias Haider

Data-oriented programming

14th of January 2023

# Exercise 2 - Report

## Topic

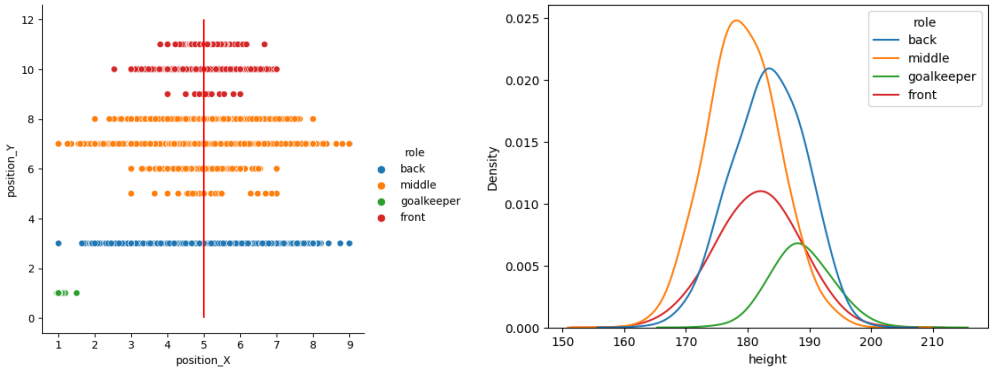
Soccer, one of the most popular sports in the world, was chosen as a project topic. The idea is to analyze specific details of players like physical attributes in the context of soccer teams, leagues and match outcomes. The following four experiments aim to give insights into data available about this sport and answer various questions soccer fans would be interested in.

## Dataset

All experiments performed during the project are based on “The European soccer database” dataset. It was simply not necessary to include more sources, since the dataset is very detailed, large and comprehensible. The database consists of the relational tables Country, League, Match, Player, Player_Attributes, Team, Team_Attributes.

## Player positions (Tobias Haider)

Soccer players usually have a specific role assigned to them. Of course, the role can change over time, but it still is mostly a rather constant fact about players. With the suspicion that attributes like height, weight etc. of a player are related to this, a model was built predicting the role a player plays in based on his physical and skill features.




The left plot shows the manual labelling done based on the coordinates stored for the matches. The right plot shows differences in the player height between the population of different roles.

Using a support vector classifier for one model and linear discriminant analysis for another model, it was shown that it is possible to predict the role of a football player based on his features. The testing accuracies vary for the roles. The highest accuracy is achieved for goalkeepers (99%), the lowest for front and center players (70% - 75%).

Topic (Ardit Luzi)

Topic (Ioan Gulyas)

Topic (Saban Akay)

Links

European soccer database: https://www.kaggle.com/datasets/hugomathien/soccer

Project github repository: https://github.com/arditluzi1/DoppExercise2