

# **Data Mining 1: Predicting positions of soccer players Project Outline**

Team No. 9

presented by

Jurgen Amedani (1628022), Cara Maria Damm (1631263), Paul Hesselmann  
(1371380), Allwyn Menezes (1671634), Martin Mlinac (1364487), Carmen  
Rannefeld (1631070)

submitted to the

Data and Web Science Group

Prof. Dr. Bizer

University of Mannheim

No Institute Given

## 1 Problem description

There are a great number of analysis in sports, especially soccer: Who is most likely to win the world championship? Who will be the player of the year? Or which position would be ideal so score?

Our group project aims to understand the characteristics of soccer players on a certain field position. Are there specific characteristics that determine if a player will be an amazing goal keeper or a top scorer?

After knowing the characteristics of the positions we want to conclude if the players in available in the FIFA'19 edition are playing in the position which is best suited for him.

In case a player isn't in his ideal spot would changing his position increase his performance?

## 2 Data Set

In order to perform our analysis we will use the FIFA 19 soccer player data which is available on Kaggle.

The dataset consists of 18000 rows and 89 attributes and was extracted from <https://sofifa.com/>.

## 3 Project approach and evaluation method

We want to divide the dataset in k subsets and use one subset as the testing subset and k-1 as the training subset.

The classification results can be compared by looking at the ROC curves.

In order to measure our success we want to use a k - fold - cross validation.

## 4 What do you expect your results to look like? (Model/Clusters/Patterns)

The result is a classification model. The ideal position for each player will be based on his characteristics.

We will be determining clusters by assigning the players in each positions as different clusters.