Results BDS lab 7

Andreas De Lille and Cedric De Pauw

Contents

Q1, Promising Players: Approach

- Create a dataframe containing
  - Info: team, position
  - Stats: for each relevant feature
    - total number of events
    - number of successful events
- For each position: determine a score
  - Striker: #Succesful Pass + #shots on target
  - Defence: #Succesful Take On + #Tackle + #Interception + #Succesful Clearance - #Corner
  - Midfield: #Succesful Pass - #Offside Pass + #Succesful Take On + #Interception
  - Goalkeeper: #Succesful Claim + #Succesful Clearance + #Punch + #Recovery + #Pick Up + #Smother
- Rank players according to this score

Q1, Promising Players: Results

Strikers: 3 of the top 6 strikers played in the Bundesliga later in their carrier Novakovic Milivoje (2011), Patrick Helmes (currently), Francisco Copado(2007)

Midfielders: 3 of the top 4 played in first leagues
Ivica Banovic (2010), Patrick Paauwe (2010), Marko Lomic(2016)

Defenders: 2 of the top 4 defenders play currently in Bundesliga
Niko Bungert, Neven Subotic

Goalkeepers: 3 of the top 5 played in first leagues
Christofer Heimeroth (currently), Florian Fromolowitz (2011), Tomasz Bobel (2010)

Q2, Predict Outcome: Approach

- Similar to Q1, but now scores for each team
- Machine learning in two steps, (mllib only supports binary classification)
  - 2 classifiers:
    - Draw vs win/lose
    - Win vs lose
  - Get train and test set
  - For each classifier:
    - Adjust labels of trainset
    - Train classifier

Q2, Predict Outcome: Results

Classifier one (draw vs non-draw) : train acc of 75%
Classifier two (win vs lose): train acc of 74 %
Test accuracy of 31%

⇒ test accuracy due to a number of reasons:

1. A total of only 90 samples
2. Hard to predict a draw
3. More features might be needed, e.g. players on the field etc
   ⇒ this would require even more training data and more analysis time