



# Player valuation in European football

Edward Nsolo, Patrick Lambrix, Niklas Carlsson

Linköping University, Sweden

## Published as

Nsolo E, Lambrix P, Carlsson N, [Player valuation in European football](#), *5th Workshop on Machine Learning and Data Mining for Sports Analytics*, [LNAI 11330](#), 42-54, Dublin, Ireland, 2019.



# Outline

- Motivation
- Methods and Results
- Conclusion



# Outline

- Motivation
- Methods and Results
- Conclusion



# Motivation

2018 summer question:

- Will C. Ronaldo do well in Juventus?



# Motivation

- What are the attributes and skills that best predict the success of individual players?
- Which are these for different positions?
- Are these the same across leagues?



# Motivation

- Performance features

- five European top leagues (England, Germany, France, Italy, Spain)
- defenders, midfielders, forwards, and goal keepers
- top 10%, top 25%, top 50%

- Prediction models



# Outline

- Motivation
- **Methods and Results**
- Conclusion



# Data

- English Premier League, La Liga, Bundesliga, Serie A, Ligue 1
- 2015-16 season
- Data provider: WhoScored
- 222 goalkeepers (GKs), 970 defenders (DFs), 1,109 midfielders (MDs), and 305 forwards (FWs).

# Data

Type	attributes
Identifying	player id, nationality, player name, age, position, height, weight, team name, league
Defensive	tackles, interceptions, offsides won, clearance, blocks, own goals, dribbles, fouls committed
Offensive	goals, assists, shots per game, key passes, fouled, offsides committed, dispossessed, bad control, dribbles
Passing	assists, key passes, passes per game, pass success rate, crosses, long balls, through balls
All	man of the match, fulltime, halftime, minutes played, aerials won, red cards, yellow cards
Class	rating

# Data

- Data cleaning
  - duplicates
  - completion
- Normalization of numeric values
- Creation of top 10%, top 25% and top 50% lists using WhoScored ranking
- Use of SMOTE to overcome class imbalance



# Feature selection - filter

Pearson correlation coefficients between each attribute and the class attributes



# Feature selection – wrapper

- Evaluate the performance of the algorithms on the data set using different subsets of attributes (best-first)
- 7 ML algorithms
- Weka for algorithms and merit of subsets
- Support count: how often does attribute occur in selected subset
- Use attributes with support count  $> 1$

# Feature selection – results

Filter

La Liga

League	POS	PC	Attributes
La-Liga	GK	10	aw, fouled, team_name, name, drib, age, yel, motm, red, flag
La-Liga	GK	25	yel
La-Liga	GK	50	<b>motm</b> , mins, fullTime
La-Liga	DF	10	<i>crosses</i> , <b>motm</b> , mins, fullTime, <i>inter</i> , aw, goals
La-Liga	DF	25	<b>crosses</b> , <b>motm</b> , <i>inter</i> , mins, fullTime, <b>aw</b> , <b>tackles</b>
La-Liga	DF	50	<i>crosses</i> , <b>motm</b> , mins, fullTime, <i>inter</i> , <b>aw</b> , goals
La-Liga	MD	10	<i>crosses</i> , mins, fullTime, <b>motm</b> , assists, <i>thrb</i> , <i>keyp</i> , <i>avgp</i> , <i>fouled</i> , <i>tackles</i> , goals, <i>shotspg</i> , halfTime, <i>drib</i> , <b>team_name</b> , <b>PassSuccPerc</b>
La-Liga	MD	25	<i>fullTime</i> , mins, assists, <b>motm</b> , <i>crosses</i> , <i>keyp</i> , <i>fouled</i> , <i>avgp</i> , <i>tackles</i> , <i>shotspg</i> , goals, yel, <b>BadControl</b> , longb, <i>thrb</i> , <i>drib</i> , halfTime, disposs
La-Liga	MD	50	<i>fullTime</i> , mins, <i>crosses</i> , <i>tackles</i> , <i>assists</i> , <i>keyp</i> , <i>avgp</i> , <i>inter</i> , <b>yel</b> , <i>thrb</i> , <i>fouled</i> , <i>motm</i> , <i>shotspg</i> , goals, longb, <i>drib</i> , aw, clear
La-Liga	FW	10	<b>goals</b> , <i>shotspg</i> , <b>motm</b> , mins, age, <i>fouled</i> , fullTime, aw, <i>crosses</i> , <i>assists</i> , <i>offsidesCommitted</i> , team_name, <i>avgp</i> , <i>disposs</i> , <b>name</b> , clear, <i>keyp</i> , halfTime, <b>owng</b> , flag, <i>BadControl</i> , blocks, <i>yel</i>
La-Liga	FW	25	<b>goals</b> , <i>shotspg</i> , <b>motm</b> , fullTime, <i>crosses</i> , mins, <b>fouled</b> , <i>disposs</i> , <i>BadControl</i> , <i>assists</i> , <i>offsidesCommitted</i> , <i>inter</i> , <i>avgp</i> , <b>age</b> , <i>keyp</i> , <b>drib</b> , aw, <i>tackles</i> , <i>yel</i> , clear
La-Liga	FW	50	<i>fullTime</i> , <i>crosses</i> , mins, <i>shotspg</i> , goals, <i>BadControl</i> , <i>disposs</i> , <i>inter</i> , <i>assists</i> , <i>motm</i> , <i>thrb</i> , <i>offsidesCommitted</i> , aw, <i>yel</i> , <i>avgp</i> , <i>keyp</i> , <i>tackles</i> , <i>fouled</i> , <i>foulsCommitted</i> , clear, <i>drib</i> , blocks

# Feature selection – results

- Filter method:

- no identifying attributes
- Rarely: red cards, own goals, offsides won
- Fewer attributes for GK/DF than for MD/FW

# Feature selection – results

## ■ Filter method:

- Interception for Serie A
- Through balls less in La Liga
- Height only for top 10% GKs, DFs and FWs in the Bundesliga and top 10% GKs in the EPL
- Tackles for all DFs in the Bundesliga, Ligue 1 and Serie A, while only for some in EPL and La Liga
- Offsides committed for all FWs in all leagues, except for the EPL (only top 50%)





# Feature selection – results

- Wrapper method:
  - In general, few selected attributes are common for the data sets related to the same position.

# Feature selection - results

Wrapper

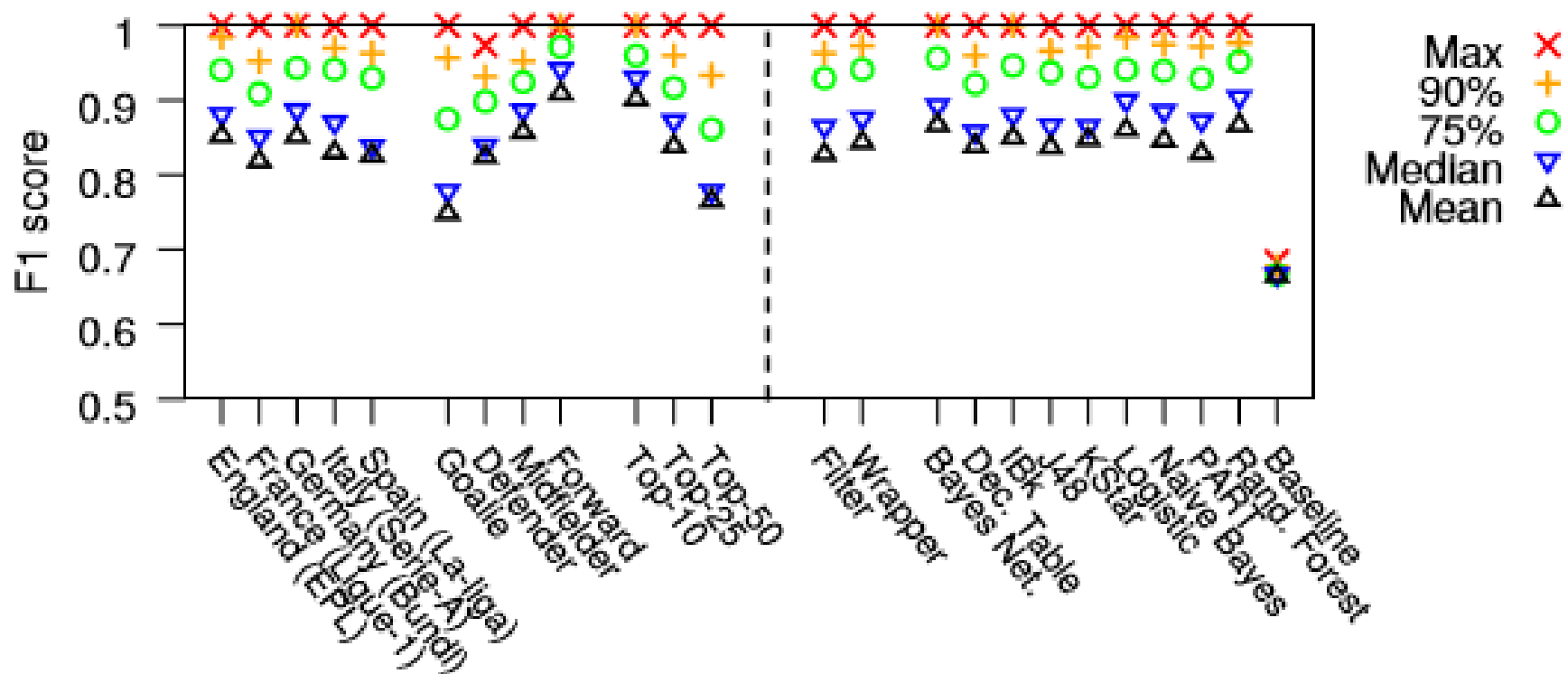
La Liga

League	POS	PC	Attributes
La-Liga	GK	10	tackles, name, red
La-Liga	GK	25	team_name, player_id, inter, name, assists, aw, keyp
La-Liga	GK	50	<b>motm</b> , crosses, inter, owng, tackles
La-Liga	DF	10	<i>team_name</i> , <b>motm</b> , name, <i>inter</i>
La-Liga	DF	25	<i>team_name</i> , fouled, offsidesCommitted, <b>tackles</b> , name, <i>motm</i> , flag, halfTime, <i>inter</i> , <b>crosses</b> , blocks, <b>aw</b> , BadControl
La-Liga	DF	50	<b>motm</b> , <i>crosses</i> , offsidesWon, PassSuccPerc, longb, tackles, <b>mins</b> , <i>inter</i> , <i>team_name</i> , drib, weight, fullTime, disposs, height, age, <b>aw</b>
La-Liga	MD	10	<b>team_name</b> , <b>motm</b> , <b>PassSuccPerc</b> , tackles, player_id, age, name
La-Liga	MD	25	<b>motm</b> , name, player_id, <b>BadControl</b> , <b>thrb</b>
La-Liga	MD	50	<b>fullTime</b> , <b>tackles</b> , offsidesWon, <b>shotspg</b> , <b>assists</b> , owng, <b>thrb</b> , <b>yel</b>
La-Liga	FW	10	<i>player_id</i> , <b>owng</b> , name, <b>motm</b> , <b>shotspg</b> , goals
La-Liga	FW	25	<b>motm</b> , <b>shotspg</b> , <i>player_id</i> , fouled, drib, assists, weight, age
La-Liga	FW	50	<b>fullTime</b> , <b>mins</b> , <b>aw</b> , <i>player_id</i> , assists, halfTime

# Prediction

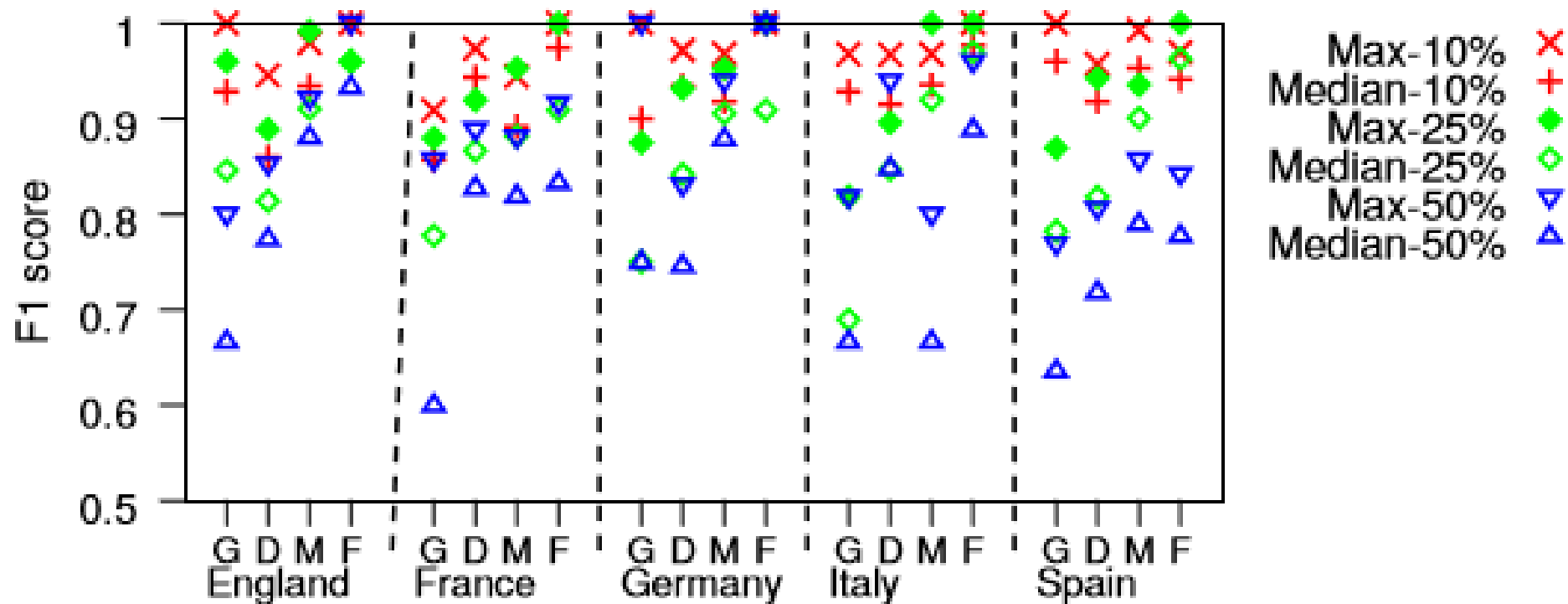
- New datasets per league, position, top X%, wrapper/filter attributes
- 10 ML algorithms
- Training (66%) and testing (34%)
- Run all experiments 10 times
- Measures: accuracy, precision, recall, F1 score, AUC-ROC

# Prediction - results



Distribution Statistics

# Prediction - results



Best/Median predictor results for top X% players



# Outline

- Motivation
- Methods and Results
- Conclusion

# Conclusions

- Attributes and skills for top X% players in different positions in 5 European leagues
- Predictors
  
- More detailed results online (see ref 27)



# Future Work

- Longitudinal drift in the rankings
- Focus on game-related attributes
- Normalization according to game time
- Features for DF and GK



# C. Ronaldo

Different features La Liga – Serie A for top 10% FWs

La Liga

aerials won (f) \*

own goals (f,w)

yellow cards (f)

shots per game (w) \*1

Serie A

through balls (f) -

long balls (f) -

tackles (f) -

dribbles (f) -

interceptions (w) -

\* C Ronaldo top 10 in La Liga in 2017-18

\*1 C Ronaldo best in La Liga in 2017-18