

METIS

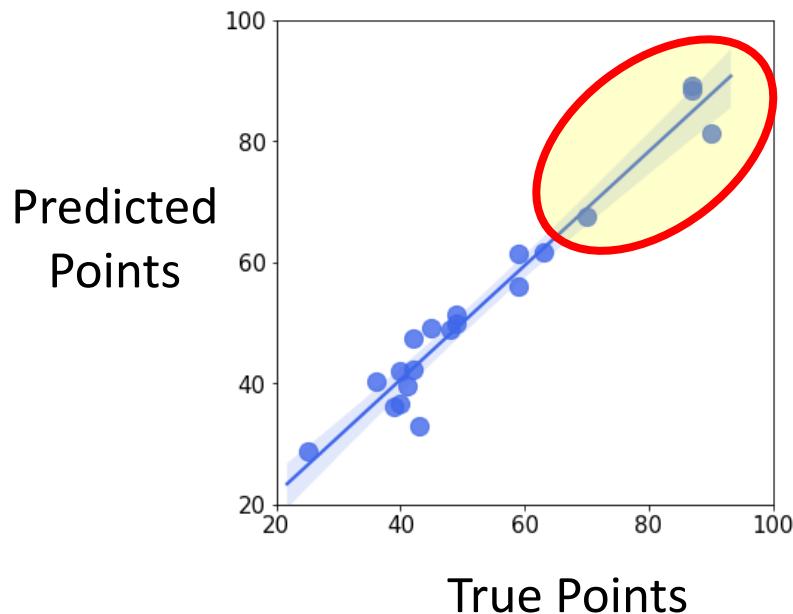
Predicting Top 4 teams in Spanish Soccer League

Ming Tang

10/25/2017

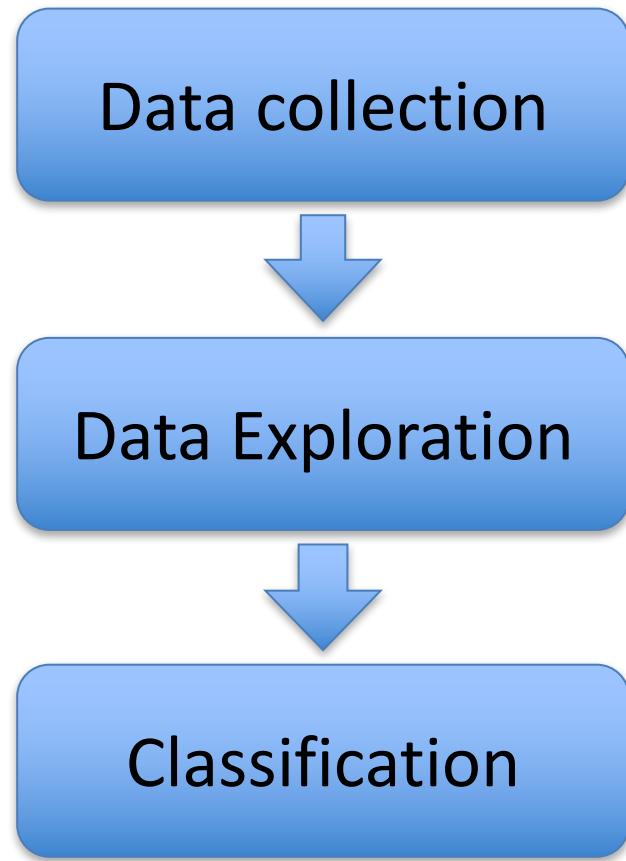
Background

- Project Luther:
 - predicting points
 - Spanish Soccer League (La Liga)
 - 2 primary features: goals scored and goals lost



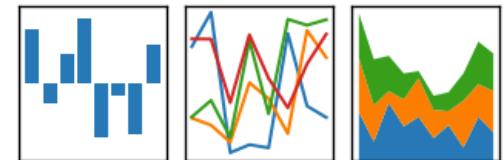
This project:
classify top 4 teams

Procedures



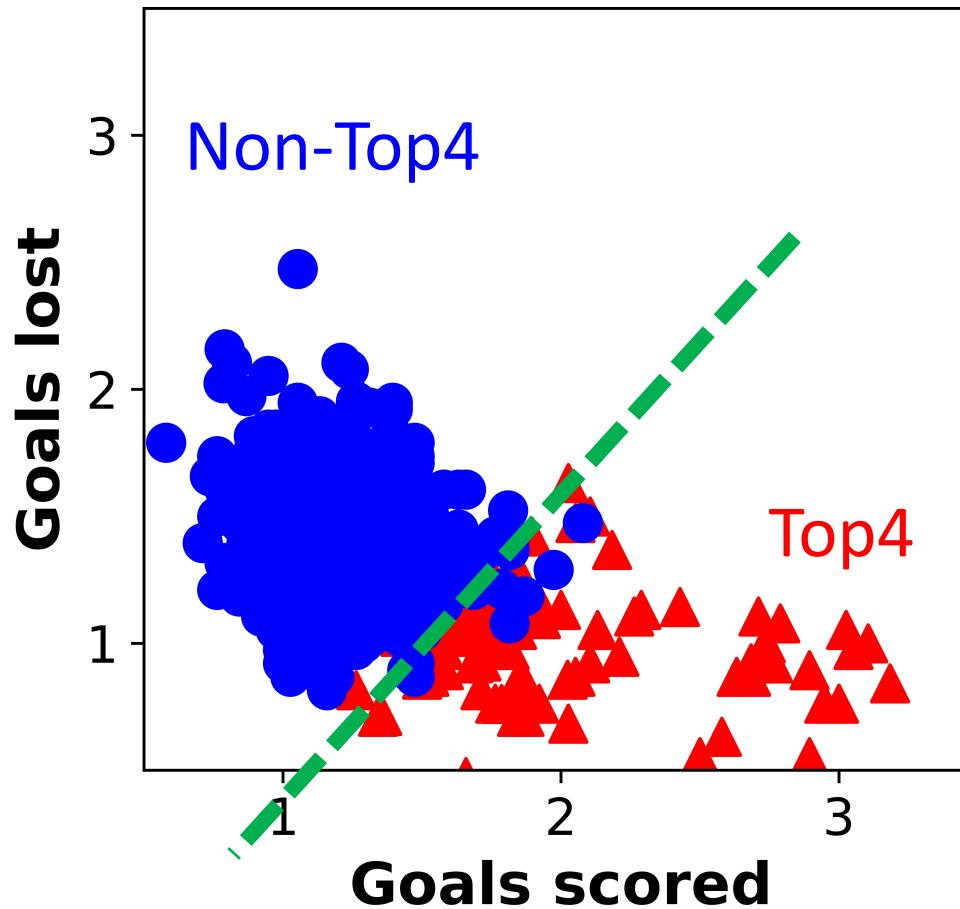
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

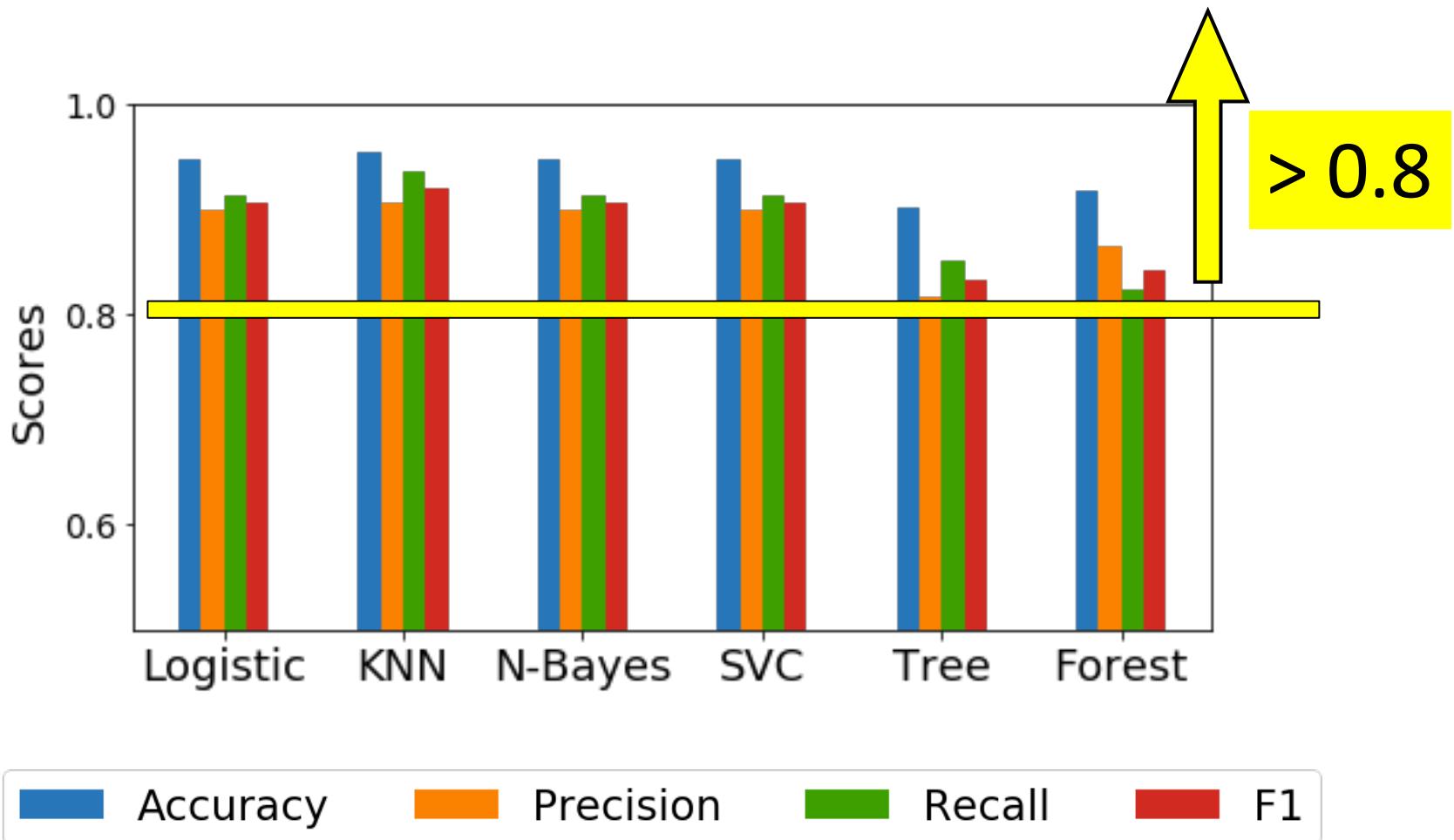


- 1) Logistic regression
- 2) KNN
- 3) Naive Bayes
- 4) SVM
- 5) Decision Tree
- 6) Random Forest

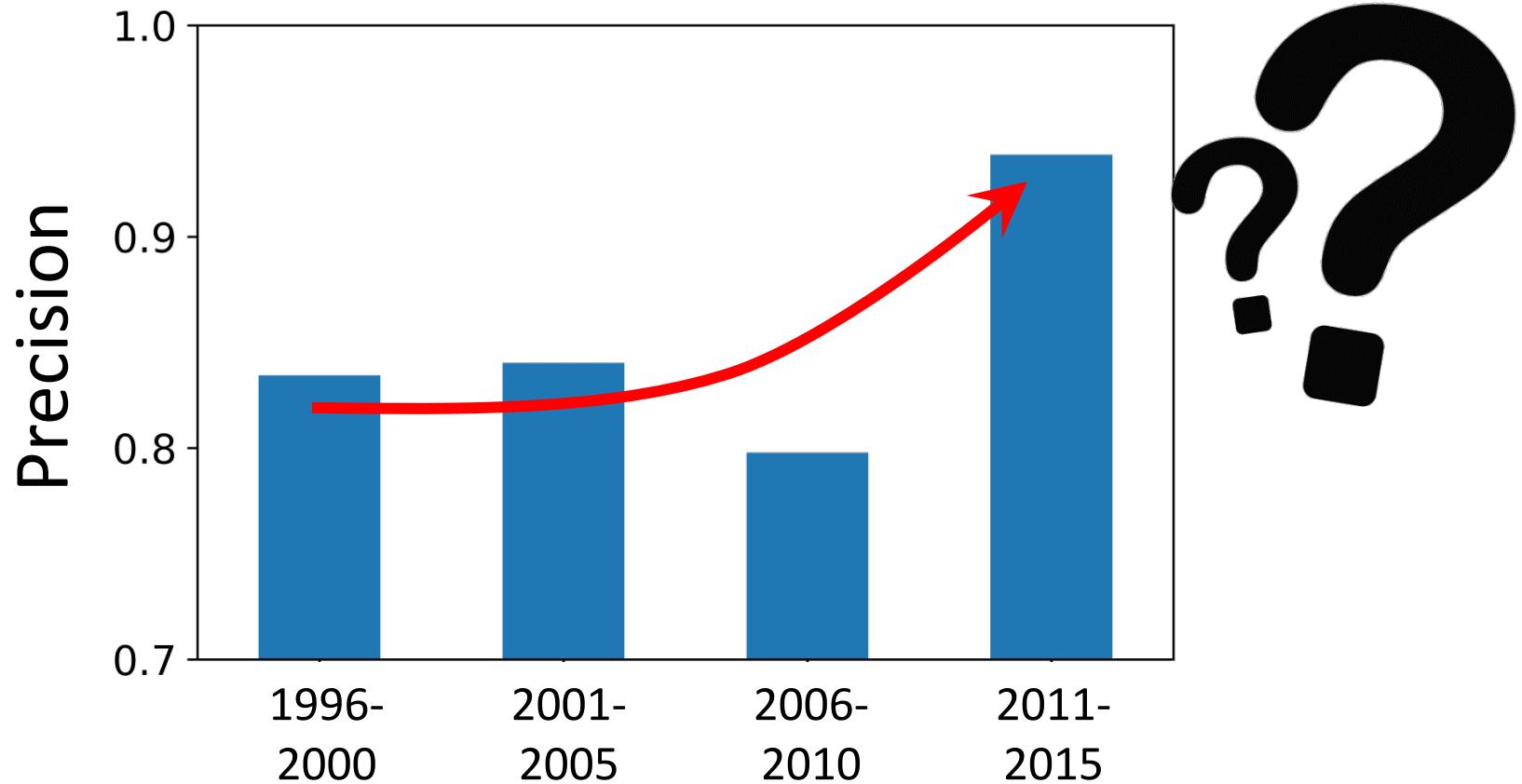
Look at the data



6 classifiers show similar and reasonable results

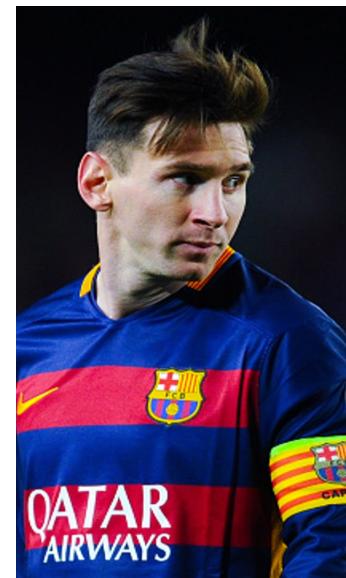
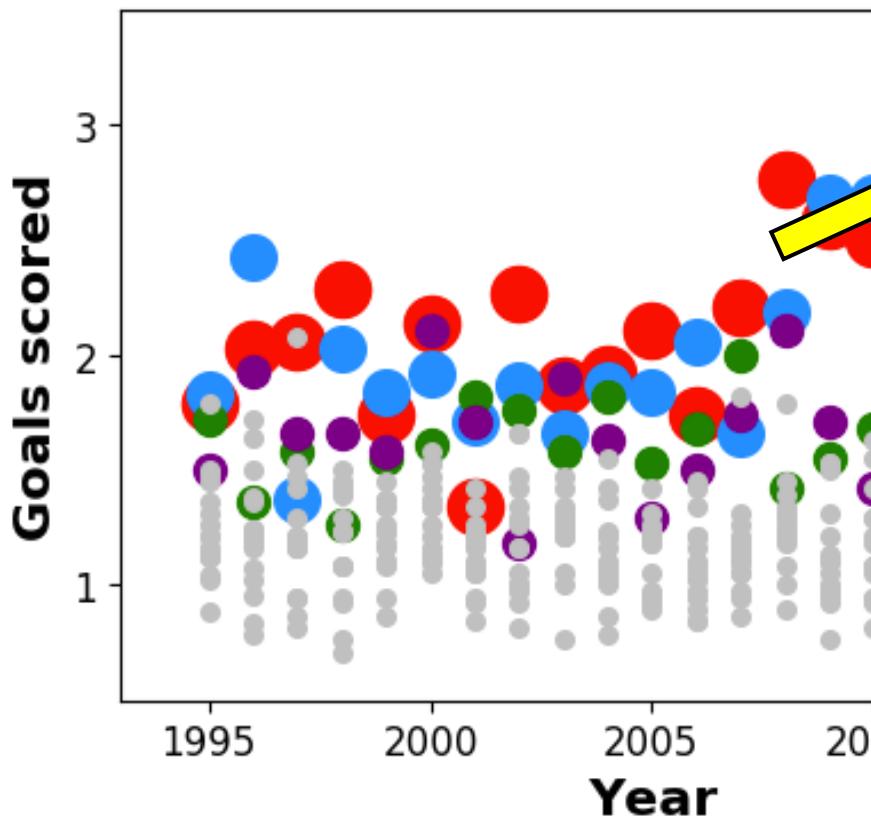


The prediction changes over time



'Precision' value is the averaged results based on 6 models

What happened after 2010?



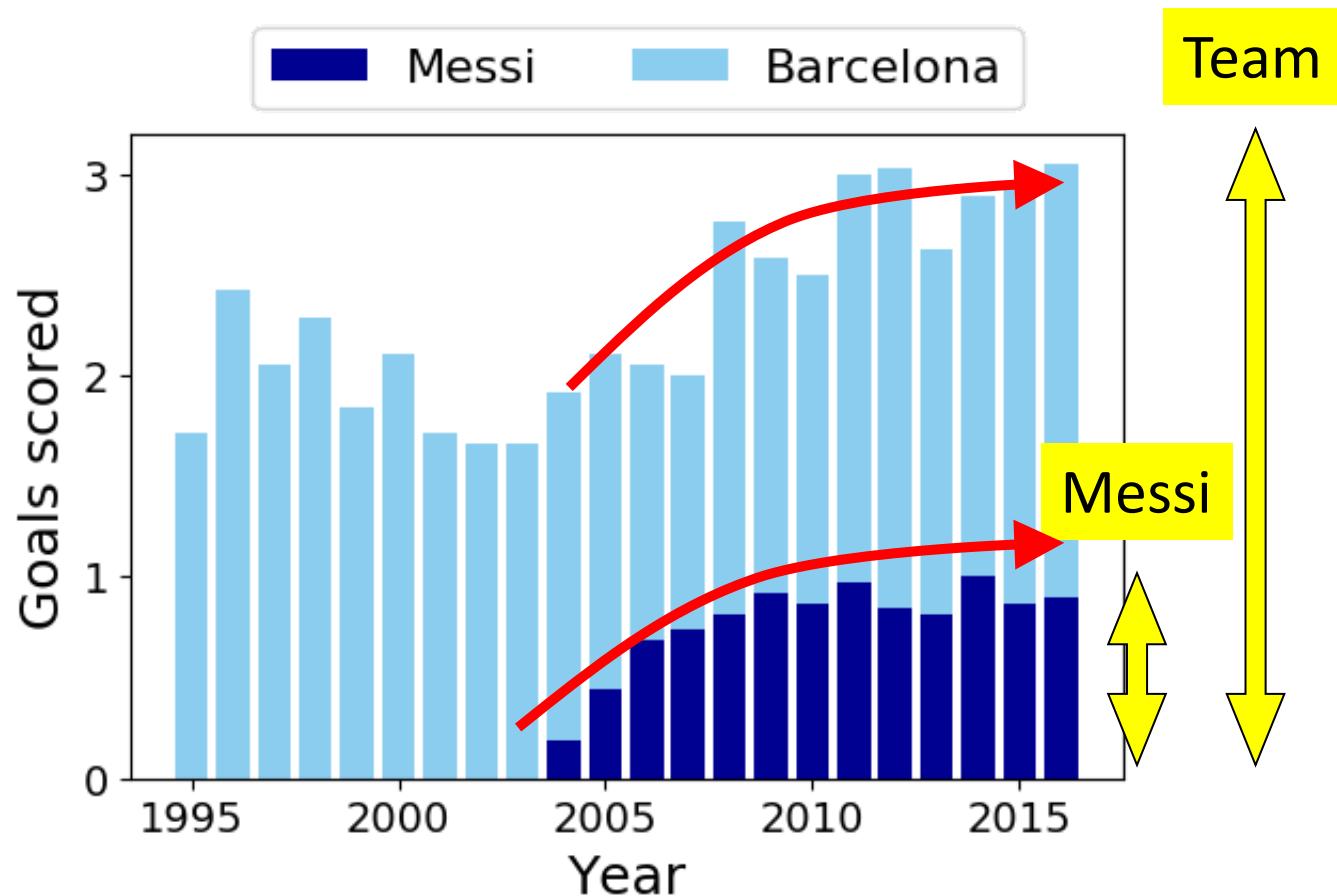
Lionel Messi
FC Barcelona

Cristiano Ronaldo
Real Madrid

Contribution of individual players matters



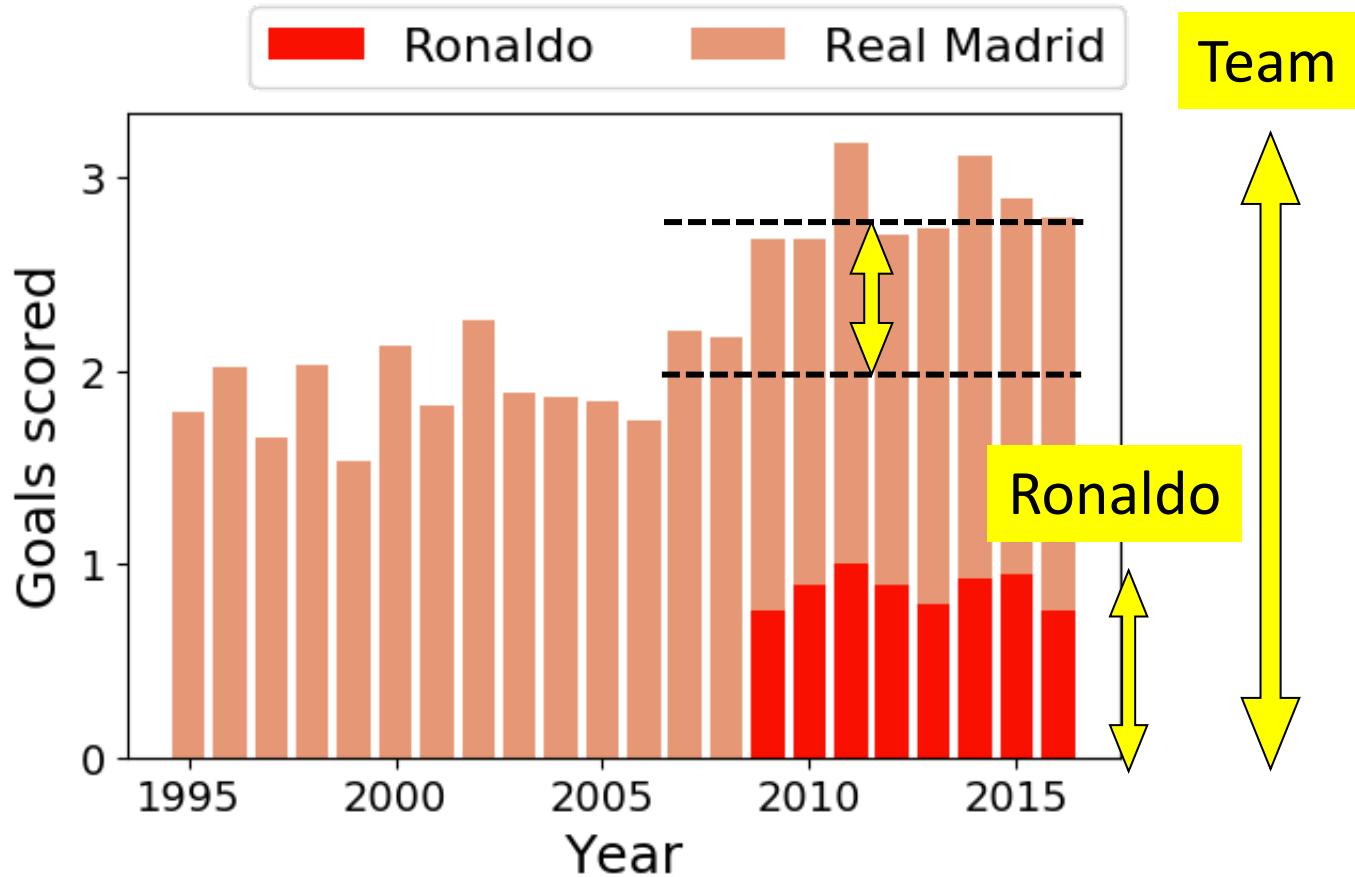
Lionel Messi
FC Barcelona



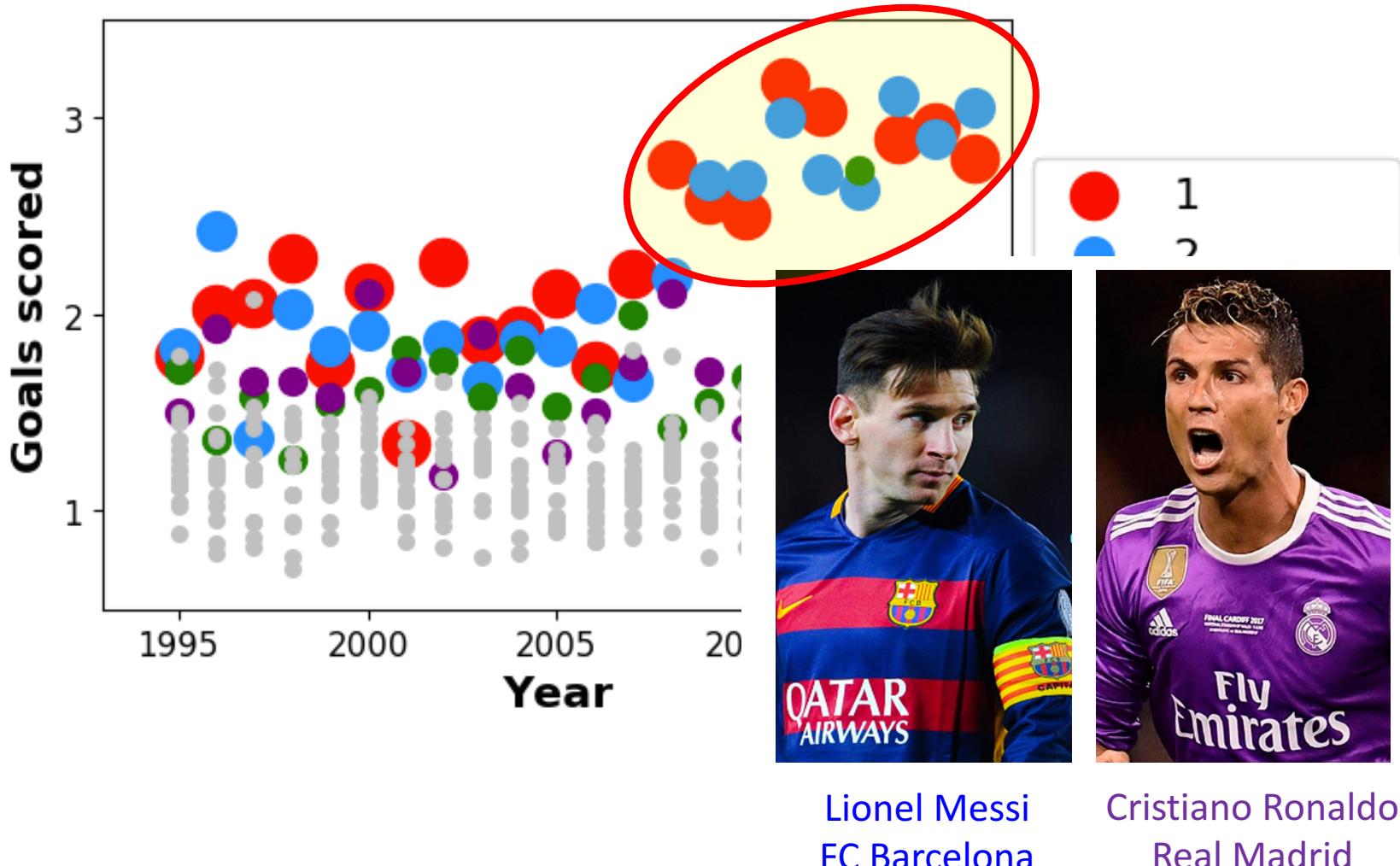
Contribution of individual players matters



Cristiano Ronaldo
Real Madrid



Contribution of individual players matters



Conclusions so far

- It is feasible to classify Top4 teams
 - scores > 0.8
- Contribution of individual players matters

For more technical details, please refer to the jupyter notebook.



Thank you for your attention!



Photo Credit

Key words	Source
La Liga	https://twitter.com/laliga
Selenium	http://selenium-python.readthedocs.io/
Fox sports	http://www.foxsports.com/watch/fs1
Pandas	http://pandas.pydata.org/
Scikit-learn	https://github.com/scikit-learn/scikit-learn
Ronaldo	http://ftw.usatoday.com/2017/06/nike-cristiano-ronaldo-ad-adidas
Messi	https://www.youtube.com/watch?v=xMA6p0jyDYM
FC Barcelona	https://www.fcbarcelona.com/news
Real Madrid C.F.	https://twitter.comrealmadrid

