



Market value prediction of soccer players

Florian Philippi
April 28th 2017

Approach

- Which factors have the most impact on the market value?
- Which players are over-/ or underrated ?

Data acquisition

The screenshot shows a player profile for Mats Julian Hummels. The top navigation bar includes links for NEWS, TRANSFERS & RUMOURS, MARKET VALUES, COMPETITIONS, MY TM, and LIVE. Below the navigation is a date: 30.06.2017. The main content area is titled 'PLAYER DATA' and contains the following information:

Name in home country:	Mats Julian Hummels
Date of birth:	Dec 16, 1988
Place of birth:	Bergisch Gladbach
Age:	28
Height:	1,91 m
Shoe size:	12
Nationality:	Germany
Position:	Defence - Centre-Back
Foot:	right
Player's agent:	HMH Sportmanagement
Current club:	Bayern Munich
In the team since:	Jul 1, 2016
Contract until:	30.06.2021
Date of last contract extension:	May 23, 2016
Outfitter:	adidas
Shoe model:	ACE LEATHER TURBOCHARGE since Mar 31, 2017
Social media:	Twitter Facebook Instagram

On the right side of the profile page, there are sections for Pronunciation (with a play button and a 0:00 timer), Detailed position (showing a diagram of a soccer pitch with 'Centre-Back' highlighted), and Market value development (a line graph showing the market value increasing from approximately 10 million euros in June 2016 to about 38 million euros by February 2017). The graph also indicates the highest market value was reached on June 1, 2016.

- [Transfermarkt.com](#)
- **Top 5 soccer countries in Europe**
- **First two divisions of each country**
- **About 5000 players**



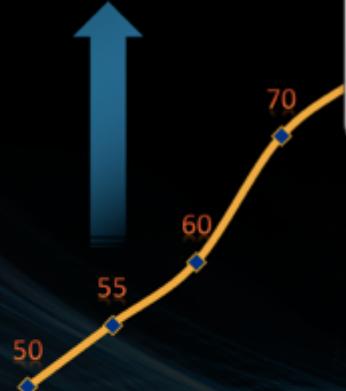
Feature selection

<u>General information</u>	<u>Performance (16/17)</u>	<u>Other</u>
Name	Games	In team since
Age	Goals	Expiry of contract
Height	Assists	Agency
Nationality	Played minutes	Outfitter
Current club	World cup winner	Foot (right/left/both)
Position	Champions league winner	Total transfer proceeds

14 features
45-50% R-squared

Model building

Linear models

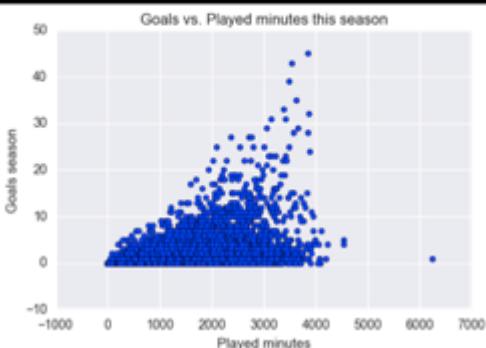


- Linear Regression
- Ridge
- Lasso
- ElasticNet

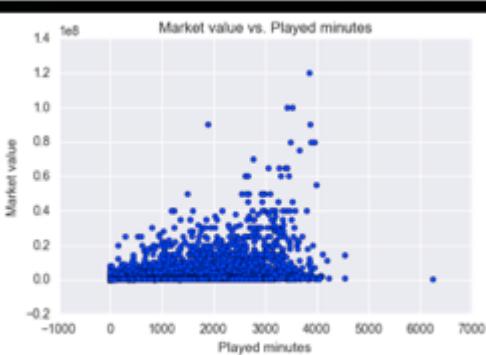
Which additional features make sense?



Feature correlation

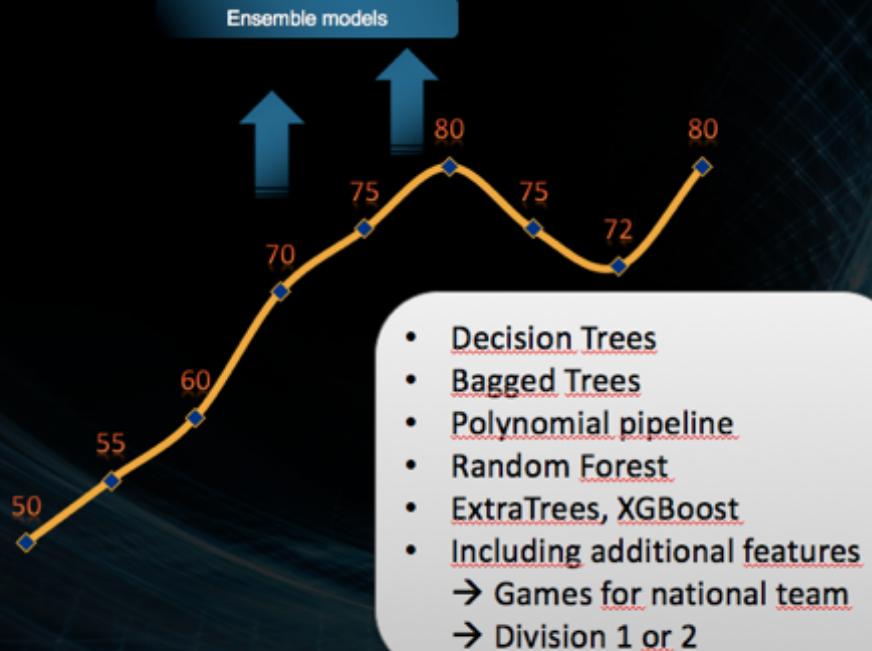


- Relation between variables "Goals scored" and "Played minutes"
- Increase the importance of scored goals.
- Goals * minutes



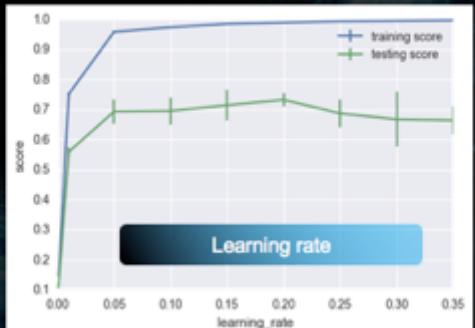
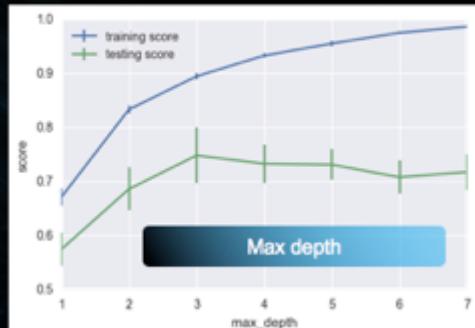
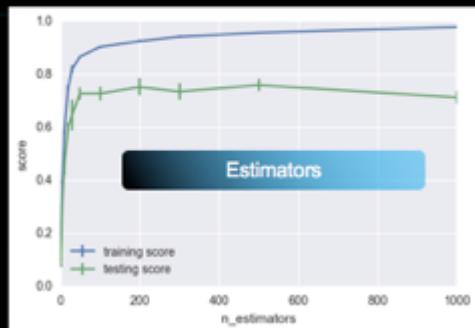
→ Increase of R-squared

Model building



- Decision Trees
- Bagged Trees
- Polynomial pipeline
- Random Forest
- ExtraTrees, XGBoost
- Including additional features
 - Games for national team
 - Division 1 or 2

Model tuning



Model building



Prediction results & feature importance



Total transfer proceeds	33%	Days in team	5.8%
Goals * played mins.	18.7%	Age	5%
Games for national team	12%	Plays in division 1	4.8%
Assists per season	10.7%	Champions League winner	1.5%
Contract expiry	8%		

Prediction results



Name	Current market value	Predictions
Robert Lewandowski	80.0	48.9
Thomas Müller	60.0	42.8
Thiago	40.0	22.1
Jérôme Boateng	40.0	32.6
David Alaba	40.0	34.8
Mats Hummels	38.0	20.8
Arturo Vidal	37.0	23.2
Renato Sanches	30.0	24.1
Douglas Costa	30.0	23.5
Joshua Kimmich	25.0	10.9
Javi Martínez	25.0	26.0
Kingsley Coman	20.0	5.9
Juan Bernat	16.0	8.2
Philipp Lahm	11.0	21.4
Arjen Robben	10.0	24.2
Franck Ribéry	8.0	14.0
Rafinha	5.0	4.4
Xabi Alonso	3.5	-6.4
Niklas Dorsch	0.2	0.7
Fabian Benko	0.2	0.2

Improvements

Include other features such as:

- number of injuries
- In-game stats like pass precision, successful tackles, etc.
- popularity / social media presence
- Add performance history of each player

Use a different kind of data set:

- All first division leagues in Europe (exclude second division)
- Look at players by position separately

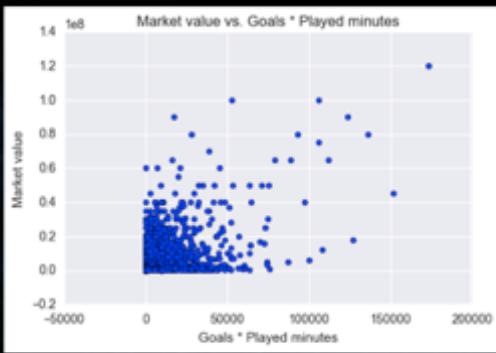
Thank You.

Florian Philippi
florian@philippi.info

Sources

- [Transfermarkt.com](#)





LR, Lasso, Ridge, EN ~50-60%

Decision Tree 45-50%

Bagged Tree 66-71%

Polynomial RF model 70-77%

Random Forest 70-77%

ExtraTrees 73-79%

Gradient Boosting 77-80%