Consiglio Nazionale
delle Ricerche

The Imitation Game: Reproducing and Explaining Human Evaluations of Soccer Performance with AI

L. Pappalardo, P. Cintia, D. Pedreschi, F. Giannotti, A.-L. Barabasi

Soccer Analytics

“Sports analytics is the process that identifies and acquires the knowledge and insight about potential players’ performances based on the use of a variety of data sources such as game data and individual player performance data”

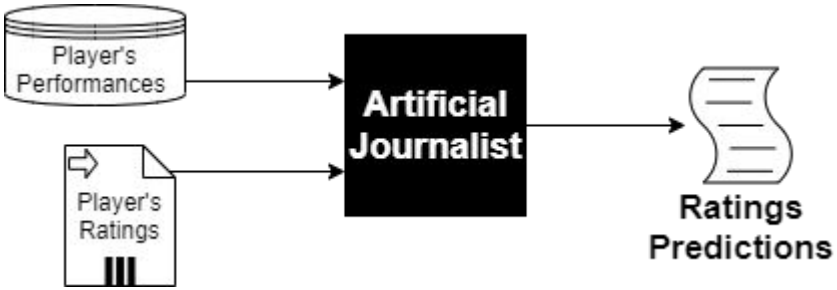
- Increase its application with the presence of massive datasets
- In the literature, there are just a few approaches that evaluate a player's performance quality in a systematic way
- PlayeRank, a data-driven framework that offers a multi-dimensional and role-aware evaluation of the performance of soccer players.

Joachim Gudmundsson and Michael Horton. Spatio-temporal analysis of team sports - A survey. CoRR abs/1602.06994, 2016. URL http://arxiv.org/abs/1602.06994.

Luca Pappalardo, Paolo Cintia, Paolo Ferragina, Emanuele Massucco, Dino Pedreschi, and Fosca Giannotti. Playerank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. ACM Trans. Intell. Syst. Technol. 10(5), September 2019. ISSN 2157-6904. doi: 10.1145/3343172. URL https://doi.org/10.1145/3343172.


TERRIBLE
1
AVERAGE
5
EXCELLENT
10

Can we reproduce, using Artificial Intelligence, the way journalists rate soccer performance?



LE PAGELLE di Antonio Giordano

ZIELINSKI, EREDE DI QUALITÀ

REINA
Per i fantacalcisti e perché sull'unico pallone rischia la salute andando nella pozzanghera.

HYSAJ
Eppure il campo non gli manca (non gli mancherebbe) ma le energie forse un pochino sì.

ALBIOL
Con le ciabatte, in stile salotto, lasciando che la Spal gli vada a battere addosso.

KOULIBALY
Il solito «energumeno»: di forza, di prepotenza e con autorevolezza ritrovata.

MARIO RUI
Rischia il giallo (e la squalifica) e quindi poi si contiene, limitandosi.

MERET
E' bravo, reattivo, istintivo e frena Insigne ma soprattutto Callejon.

SALOMON
Non sceglie: aspetta o attacca Insigne e rischia di finire a gambe all'aria.

VICARI
Sta là dietro e oppone il corpo e la posizione alle rare verticalizzazioni.

FELIPE
Si stacca troppo, aprendo la corsia centrale per Allan, perché Callejon lo distrae.

LAZZARI
Gli mancano le coperture e poi dà un senso di anarchia tagliando sempre, troppo.

ALLAN
Il gol che riconsegna il primato in classifica, prima di correre per sé e per gli altri.

JORGINHO
Geometrie apprezzabili, però senza avere intorno uomini che pedalino come si dovrebbe.

HAMSIK
Il pallido capitano rimane dietro i suoi standard e l'ammonizione gli fa male.

CALLEJON
Apre per Allan e lo manda in porta e poi (sembra) governa i carichi di fatica.

MERTENS
E' la prima sponda nell'1-0 ma è anche un po' vago, quasi distante dalla partita.

SCHIATTARELLA
Si ritrova con Hamsik, lo contiene e persino lo costringe a stargli dietro.

VIVIANI
Gli viene meno il gusto di osare e palleggia con paura addosso che diventa nemica.

GRASSI
Perde lo scatto di Allan, poi dà movimento e pure eleganza ad un centrocampo piatto.

DRAMÈ
Quasi si isola e lascia che da quelle parti, ma senza esagerare, il Napoli vada.

KURTIC
L'unica preoccupazione è Jorginho e spreca non l'occasione ma il suo tempo.

INSIGNE
Insegue il gol, e si vede, però Meret e il palo lo costringono a soffrire ancora.

ZIELINSKI
(25' st)
E' di impatto ma anche di talento (e che ruleta!). Hamsik ha un erede di qualità assoluta.

ROG
(41' st)
Va a coprire il campo, per restringerlo, nel finale da domare con intelligenza.

DIAWARA
(45' st)
L'ultimo argine per il recupero che diventa ampio e comunque pericoloso.

SARRI 6
Piccole tracce di Napoli, qualcosina all'avvio, poi una gestione eccesiva.

SEMPLICI 5,5
Magari un pizzico di coraggio in più, solo quello, per dire di averci provato.

GAVILLUCCI
Già non averla complicata, semplice com'era, sa di buon senso. Comoda così eh

ANTENUCCI
Non gli arriva uno straccio di pallone, ma non ne va neanche a inseguire.

COSTA
(16' st)
In un contesto blando a cui può solo garantire di fungere da cerniera.

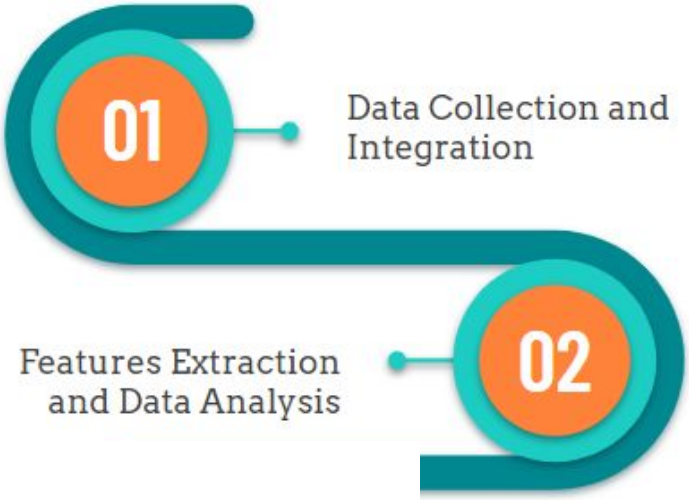
FLOCCARI
(30' st)
E' il jolly che si va a cercare: magari una palla sporca. Ma bisognerebbe arrivare a lui.

PALOSCHI
(37' st)
Aggiunge spiccioli di minutaggio ad una gara in cui l'attacco non esiste.

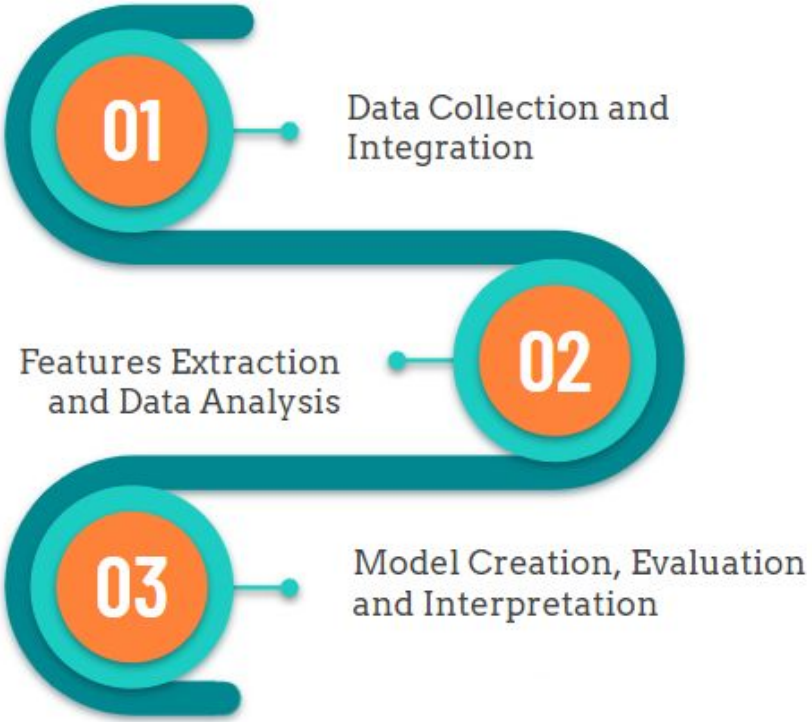
EXPERIENCE

Data Collection and Integration

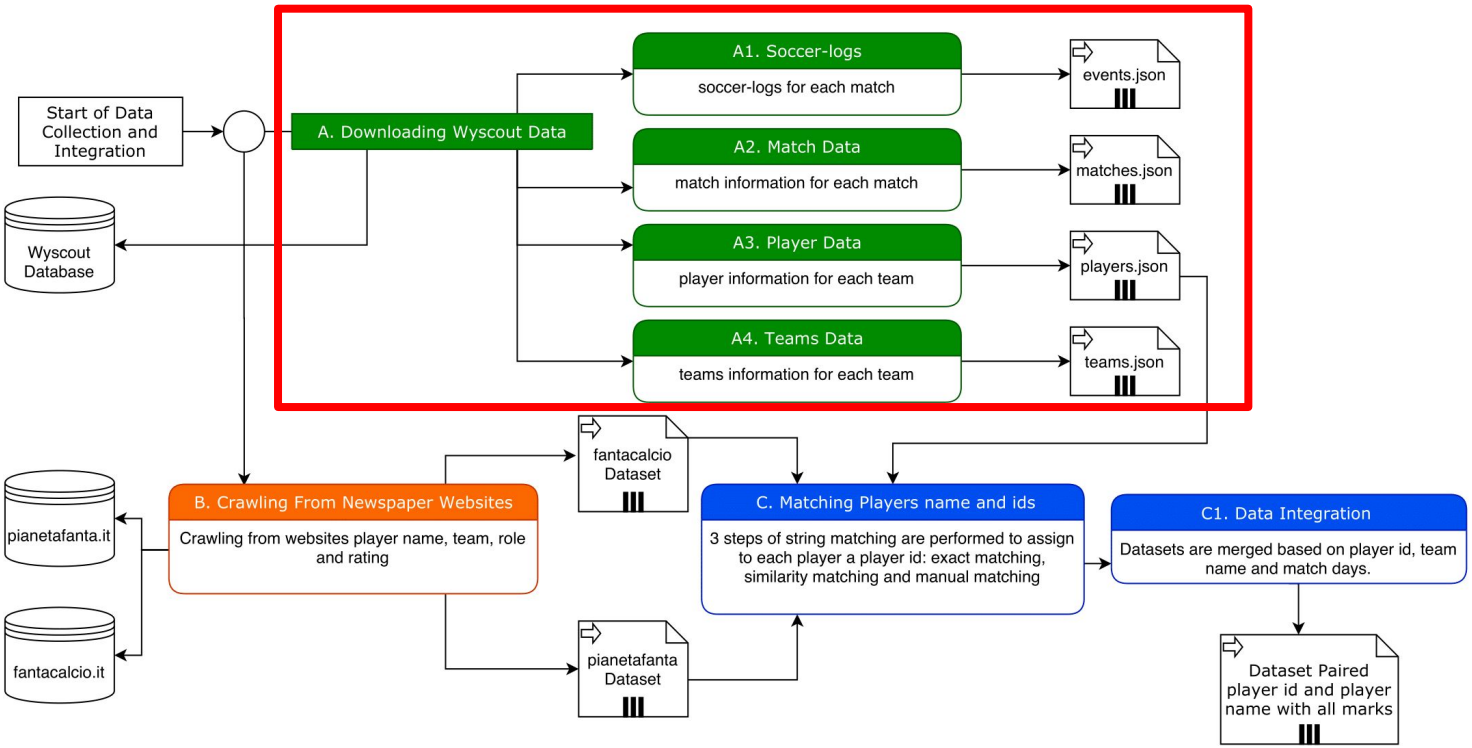
EXPERIENCE


01
Data Collection and Integration
Features Extraction and Data Analysis
02

EXPERIENCE

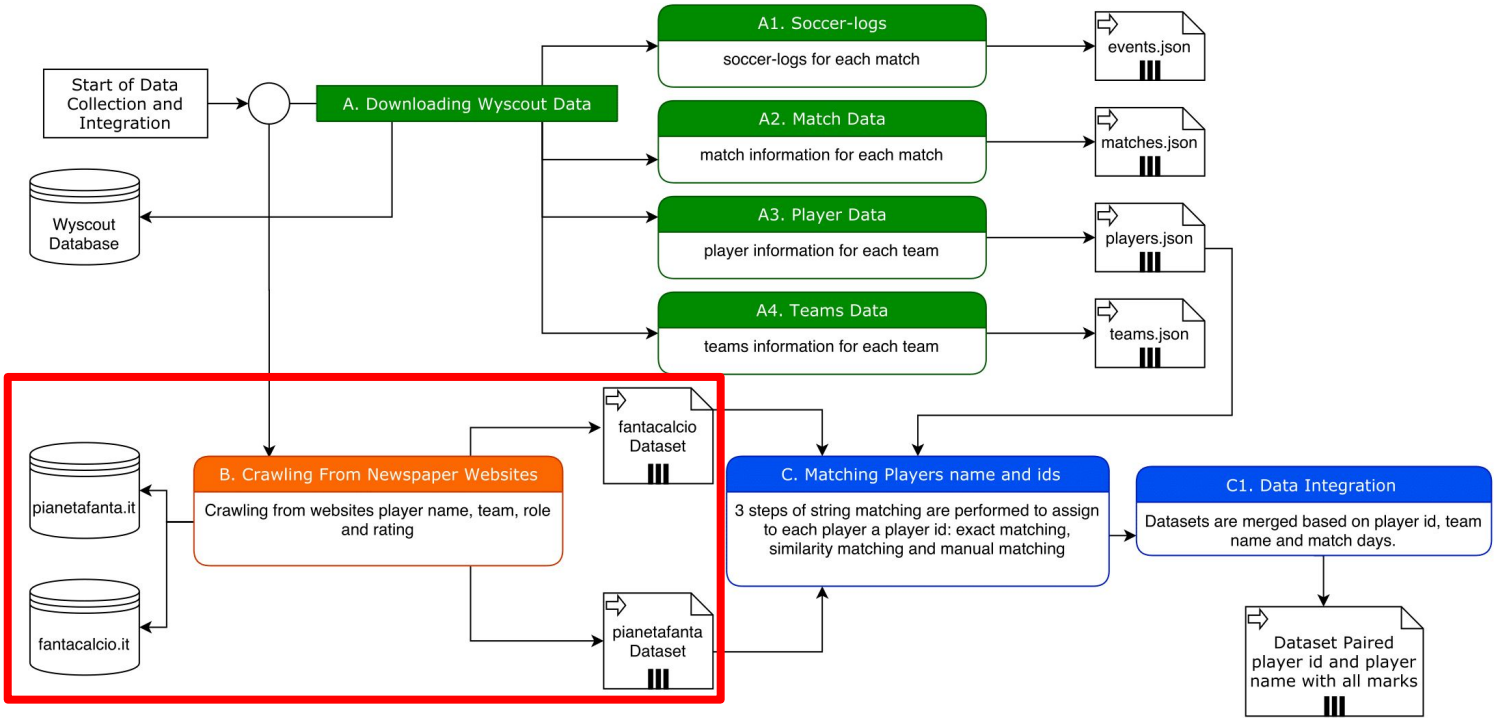

01
Data Collection and Integration
02
Features Extraction and Data Analysis
03
Model Creation, Evaluation and Interpretation

Data Collection and Integration



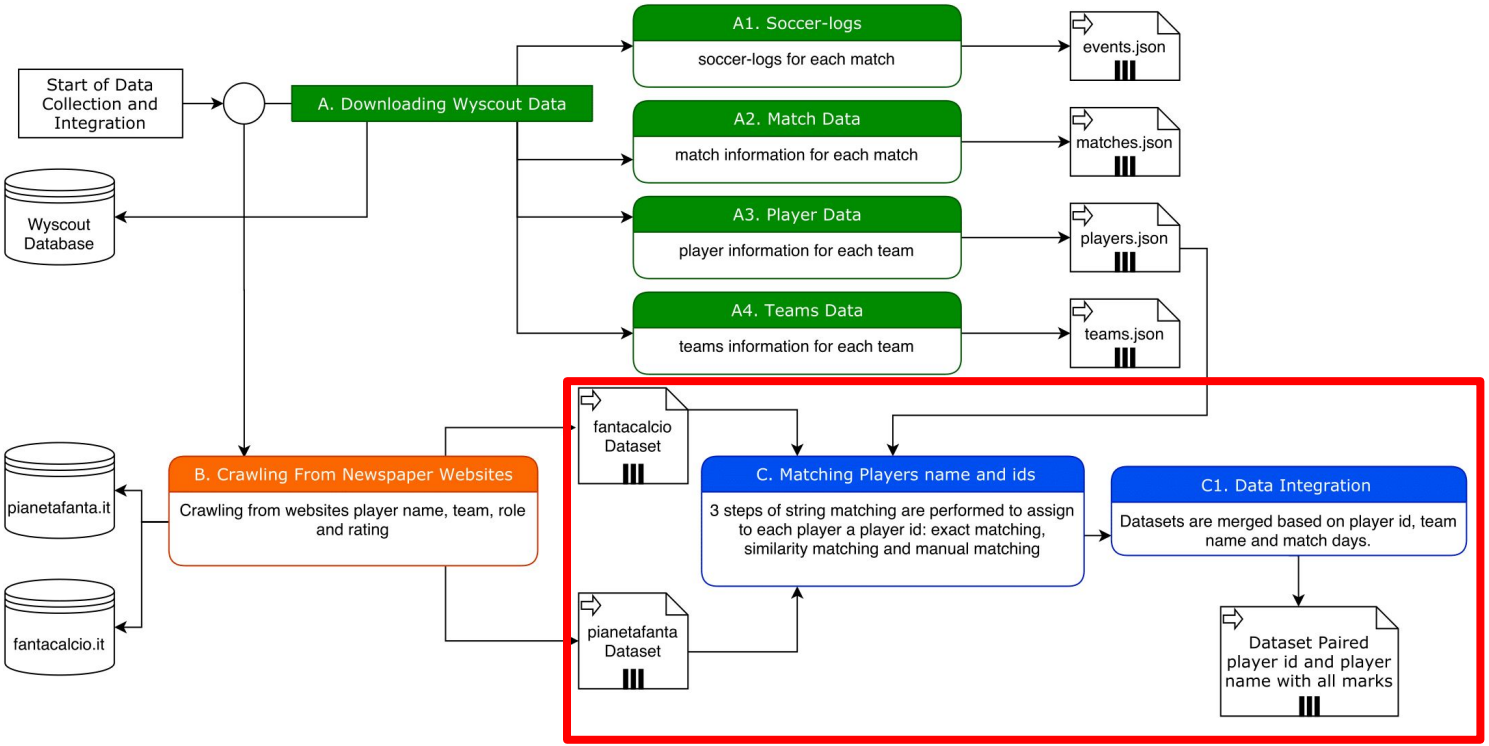
Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. Scientific Data, 6(1):236, 2019. doi: 10.1038/s41597-019-0247-7. URL https://doi.org/10.1038/s41597-019-0247-7.

Data Collection and Integration



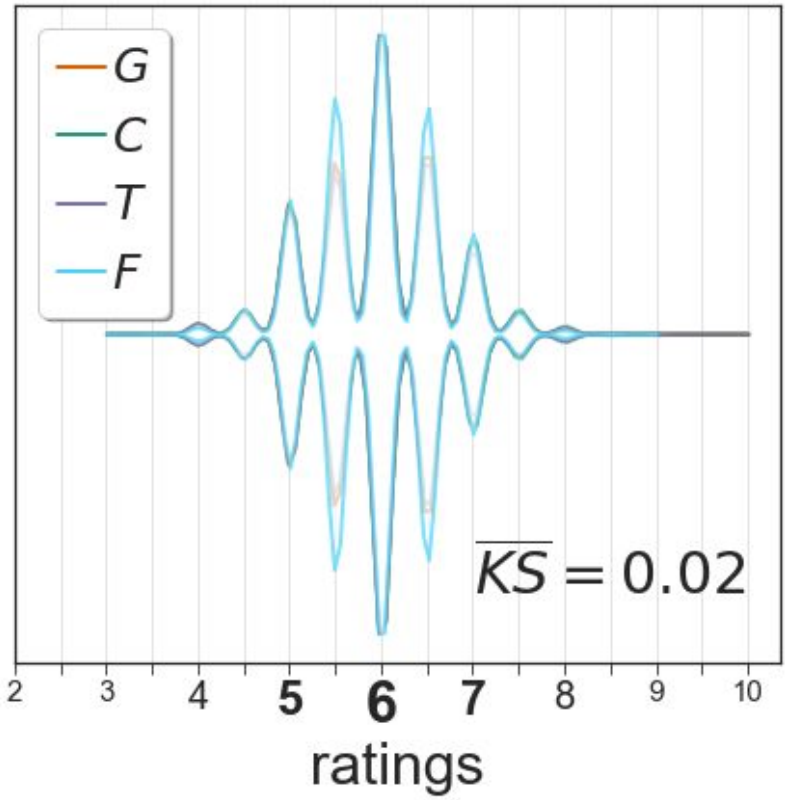
Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. Scientific Data, 6(1):236, 2019. doi: 10.1038/s41597-019-0247-7. URL https://doi.org/10.1038/s41597-019-0247-7.

Data Collection and Integration



Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. Scientific Data, 6(1):236, 2019. doi: 10.1038/s41597-019-0247-7. URL https://doi.org/10.1038/s41597-019-0247-7.

Data Analysis

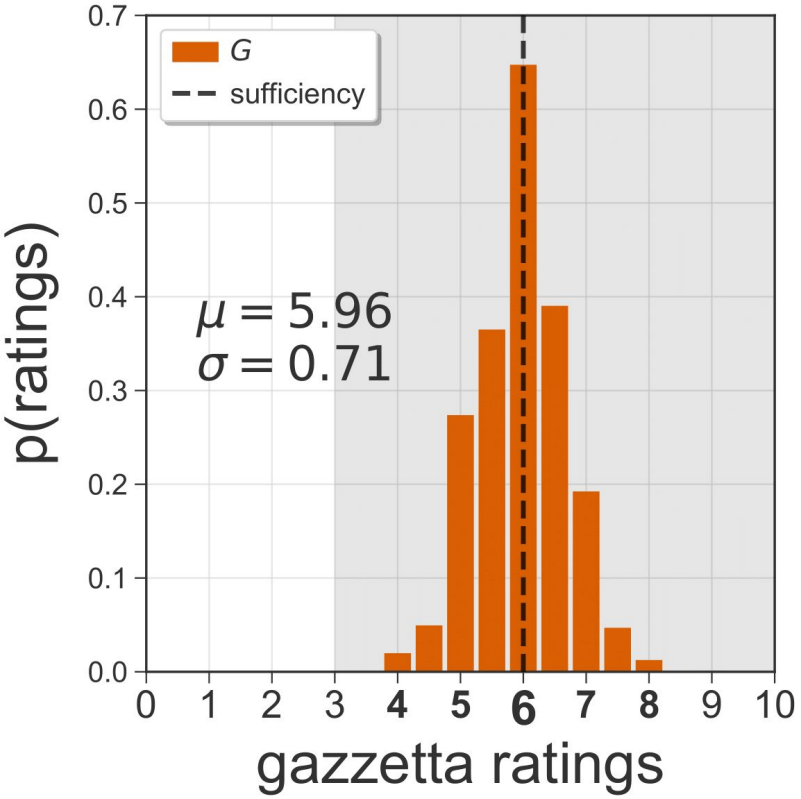


Different Aspects

Principal subjects covered:

(a) Similar Distribution
(b) Peak at Sufficiency
(c) Strong Correlation
(d) Win is an Important Starting Point

Data Analysis

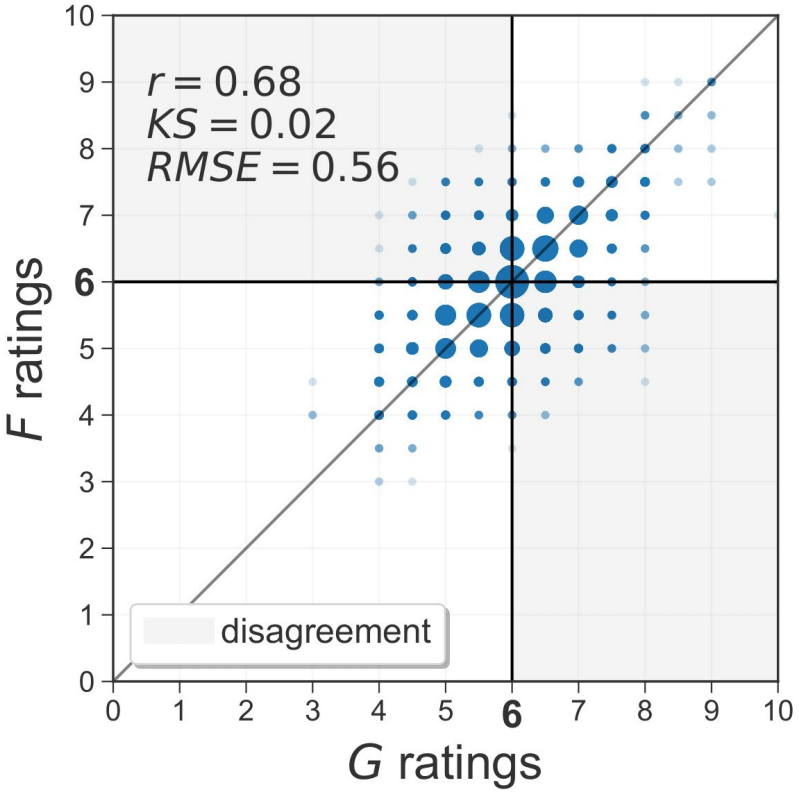


Different Aspects

Principal subjects covered:

(a) Similar Distribution
(b) Peak at Sufficiency
(c) Strong Correlation
(d) Win is an Important Starting Point

Data Analysis

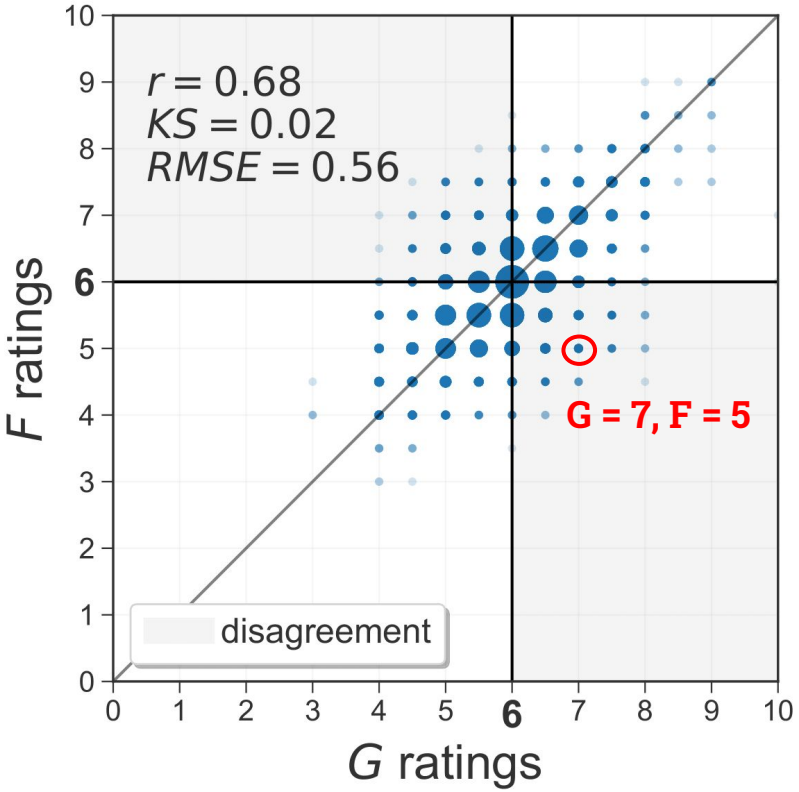


Different Aspects

Principal subjects covered:

(a) Similar Distribution
(b) Peak at Sufficiency
(c) Strong Correlation
(d) Win is an Important Starting Point

Data Analysis



Different Aspects

Principal subjects covered:

(a) Similar Distribution
(b) Peak at Sufficiency
(c) Strong Correlation
(d) Win is an Important Starting Point

Soccer-Logs

wyscout | DATA


```
{"eventName": "Pass",
 "eventSec": 2.4175,
 "matchId": 2576335,
 "matchPeriod": "1H",
 "playerId": 3344,
 "positions":
 [{"x": 49, "y": 50}, {"x": 38, "y": 58}],
 "subEventName": "Simple pass",
 "tags":
 [{"id": 1801}],
 "teamId": 3161}
```


When the event takes place

(i.e. 1H or 2H)

Event starting and ending positions

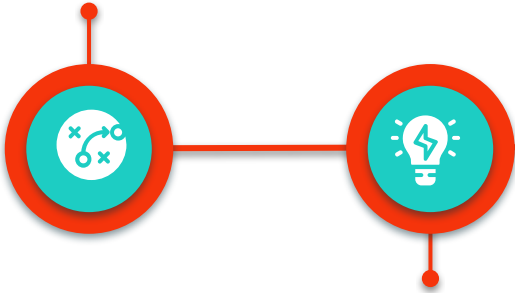
Additional meta-information(i.e. 1801 accurate event)

Features Extraction

Quantity Features
Player's volume of play during a match (e.g. total number of passes, total number of shots).
15 Features.

Features Extraction

Quantity Features
Player's volume of play during a match (e.g. total number of passes, total number of shots).
15 Features.

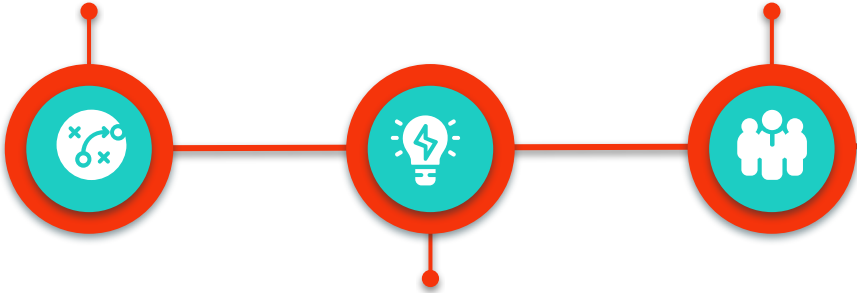


Quality Features
Player's accuracy during a match (e.g. total number of completed passes, total number of failed driblings).
45 Features.

Features Extraction

Quantity Features
Player's volume of play during a match (e.g. total number of passes, total number of shots).
15 Features.

Contribution Features
Player's contribution to its team during a match (e.g. contribution of the player w.r.t. to passes).
45 Features.

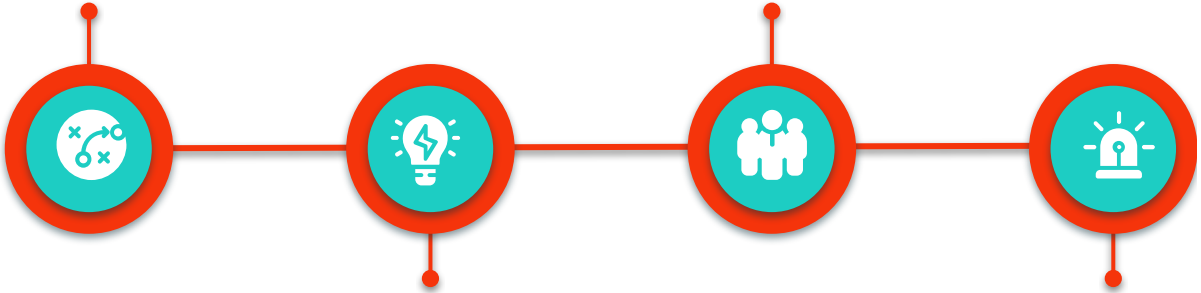


Quality Features
Player's accuracy during a match (e.g. total number of completed passes, total number of failed driblings).
45 Features.

Features Extraction

Quantity Features
Player's volume of play during a match (e.g. total number of passes, total number of shots).
15 Features.

Contribution Features
Player's contribution to its team during a match (e.g. contribution of the player w.r.t. to passes).
45 Features.



Quality Features
Player's accuracy during a match (e.g. total number of completed passes, total number of failed driblings).
45 Features.

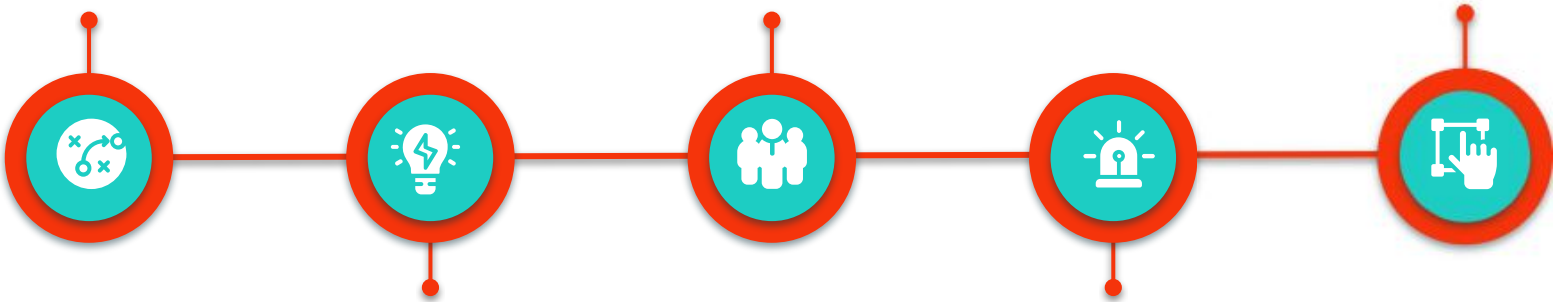
Dangerouseness Features
Player's offensiveness (in terms of positioning) during a match (e.g. position of shot in a match).
45 Features.

Features Extraction

Quantity Features
Player's volume of play during a match (e.g. total number of passes, total number of shots).
15 Features.

Contribution Features
Player's contribution to its team during a match (e.g. contribution of the player w.r.t. to passes).
45 Features.

Contextual Features
Information regarding the context of the match (e.g. expectation of a team win, goal difference between teams). **12 Features**.

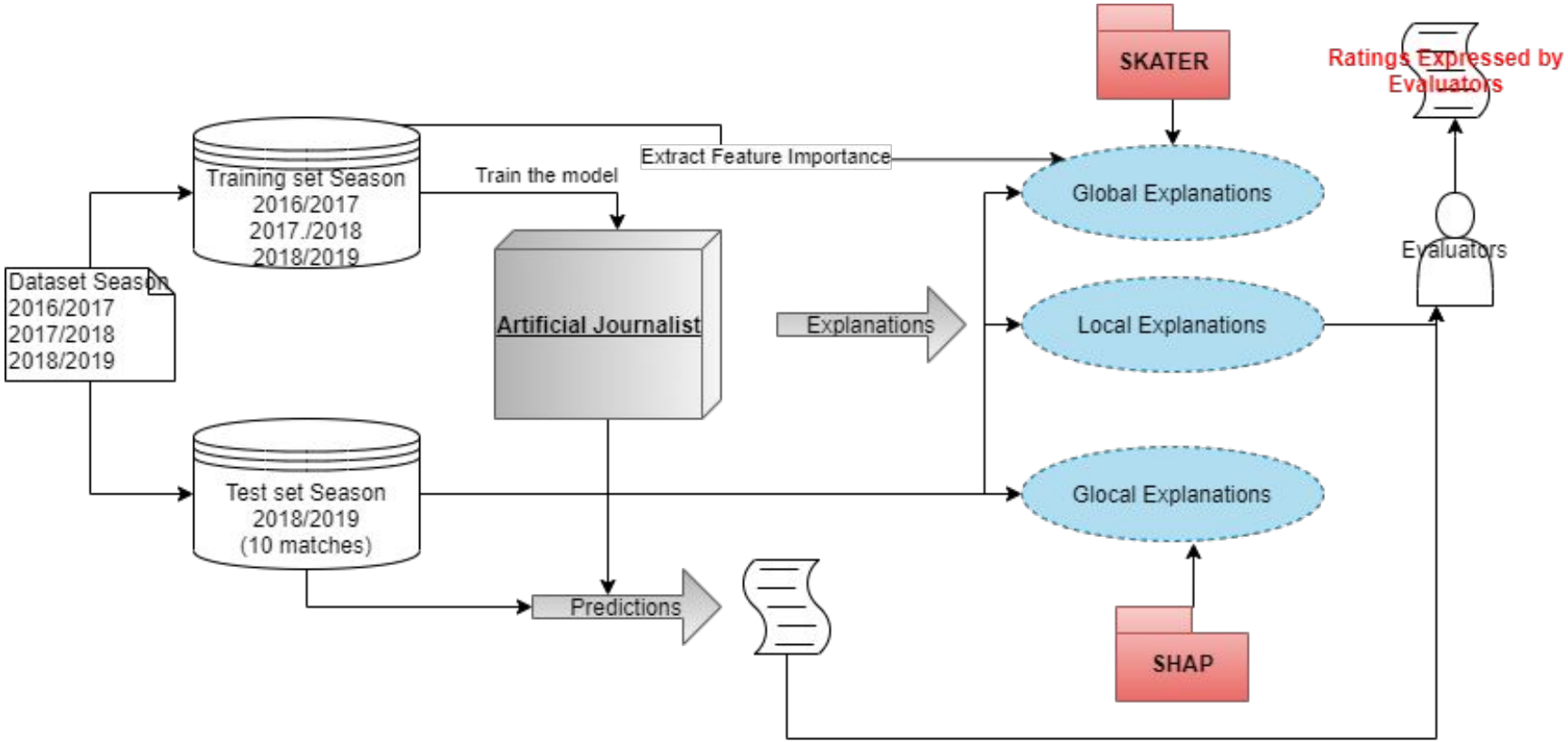


Quality Features
Player's accuracy during a match (e.g. total number of completed passes, total number of failed driblings).
45 Features.

Dangerouseness Features
Player's offensiveness (in terms of positioning) during a match (e.g. position of shot in a match).
45 Features.

169 Features
Computed

Can we reproduce, using Artificial Intelligence, the way journalists rate soccer performance?



Model Creation And Evaluation Methodology

Different Models

- Ordinal Regressor
- Neural Network
- XGBoost
- Decision Tree Regressor

Methodology

- 4 models (one for each role)
- Remove Extreme Outlier
- Encode Contextual Variables
- Hyperparameter Tuning
- Cross Validation For Evaluation

Metrics For Evaluation

- Root Mean Squared Error (RMSE)
- Accuracy
- Kolmogorov-Smirnov statistics (KS)
- Pearson Correlation Coefficient (r).

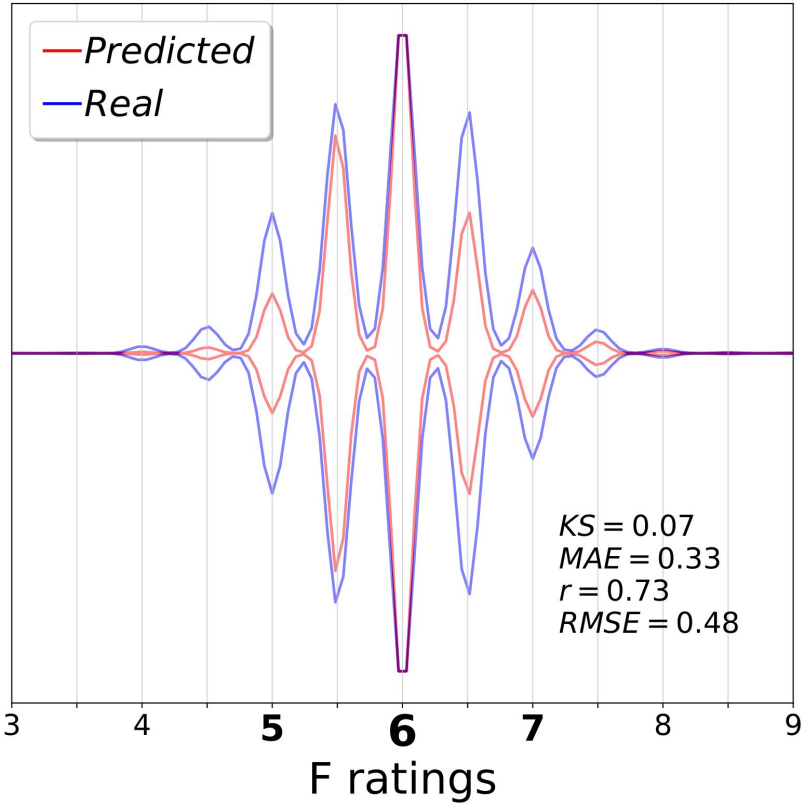
Ordinal Regressor

We use performance and ratings to create an

artificial journalist $AJ_{(P)}$

to predict F ratings from performance

r = 0.73 (0.68, G vs F)
KS = 0.07 (0.02, G vs F)
RMSE = 0.48 (0.56, G vs F)



Ordinal Regressor gives the best results in terms of similar distribution to real ratings.

Regarding *r* and RMSE has high performance.

Special mention to Neural Network created that achieve similar results.

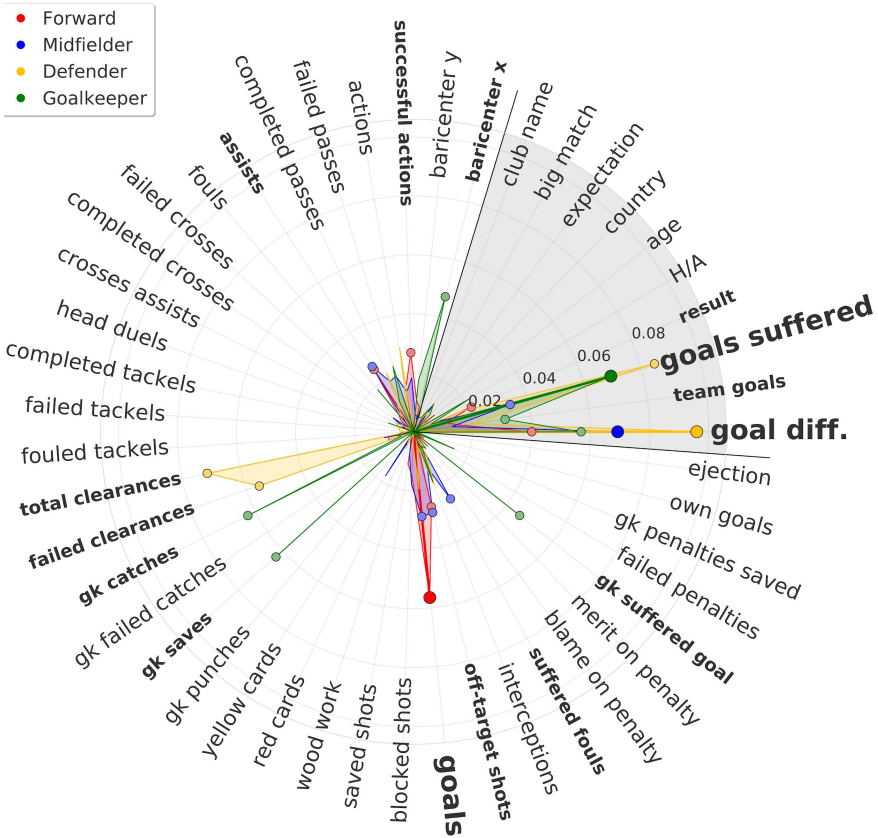
		Mid	For	Def	Gk
RMSE	Ordinal Regressor	0.46	0.45	0.49	0.50
	NN1	0.44	0.45	0.49	0.49
	NN2	0.55	0.58	0.61	0.75
	NN3	0.55	0.69	0.68	0.96
	XGBoost	0.47	0.50	0.53	0.53
	DecisionTreeRegressor	0.49	0.49	0.52	0.50
Accuracy	Ordinal Regressor	0.44	0.49	0.41	0.47
	NN1	0.49	0.48	0.45	0.48
	NN2	0.41	0.39	0.36	0.32
	NN3	0.39	0.39	0.35	0.36
	XGBoost	0.46	0.45	0.42	0.47
	DecisionTreeRegressor	0.43	0.43	0.39	0.46
KS	Ordinal Regressor	0.09	0.08	0.07	0.10
	NN1	0.08	0.12	0.09	0.13
	NN2	0.11	0.18	0.12	0.12
	NN3	0.09	0.12	0.18	0.14
	XGBoost	0.06	0.06	0.07	0.13
	DecisionTreeRegressor	0.14	0.13	0.18	0.10
r	Ordinal Regressor	0.71	0.84	0.68	0.54
	NN1	0.74	0.83	0.69	0.54
	NN2	0.68	0.78	0.64	0.43
	NN3	0.58	0.68	0.45	0.21
	XGBoost	0.70	0.80	0.63	0.42
	DecisionTreeRegressor	0.68	0.81	0.63	0.49

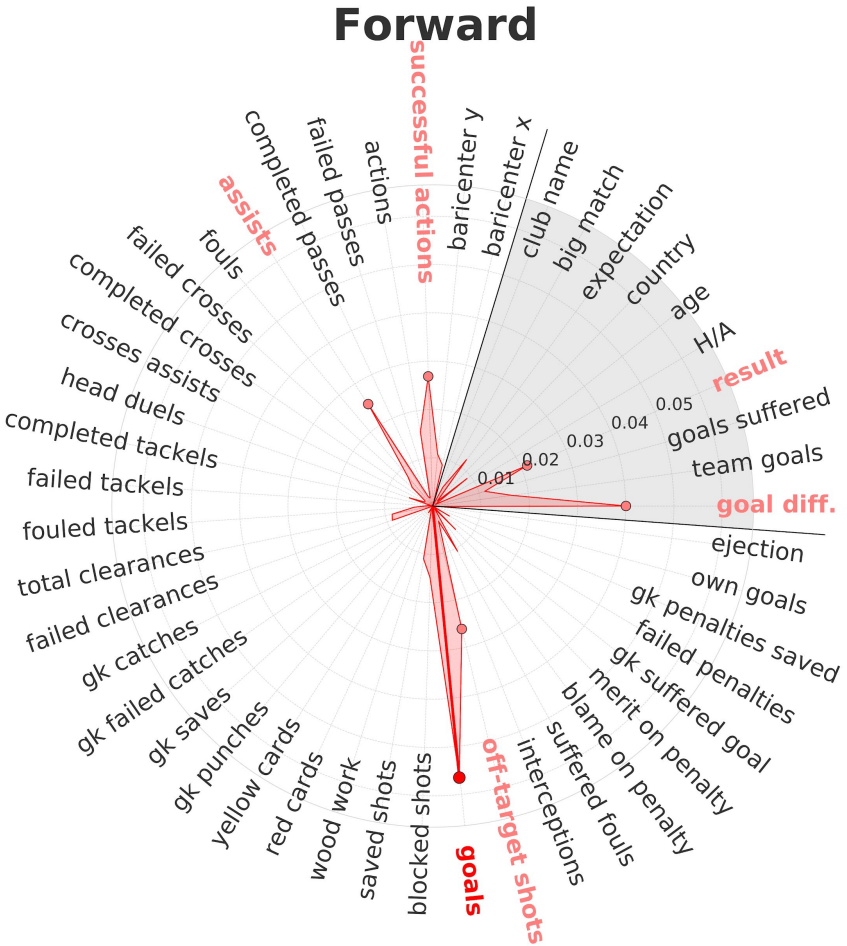
From Black Box to Explanations

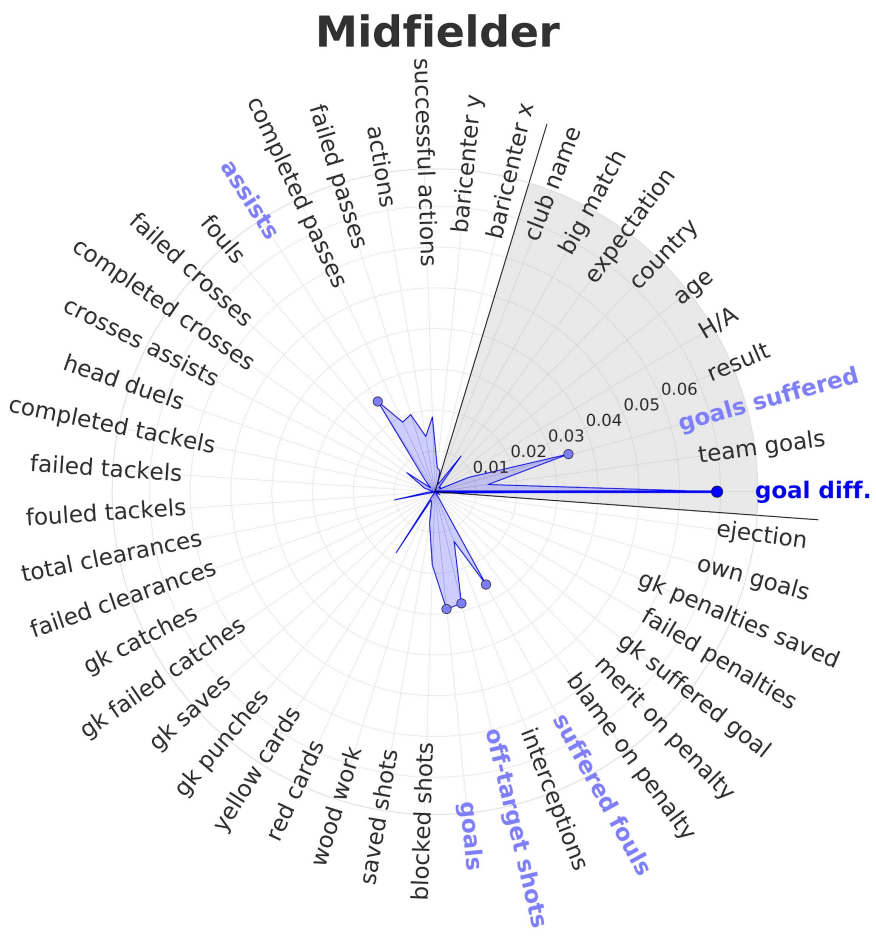
(Global, Local and Glocal Explanations)

Global Explanations

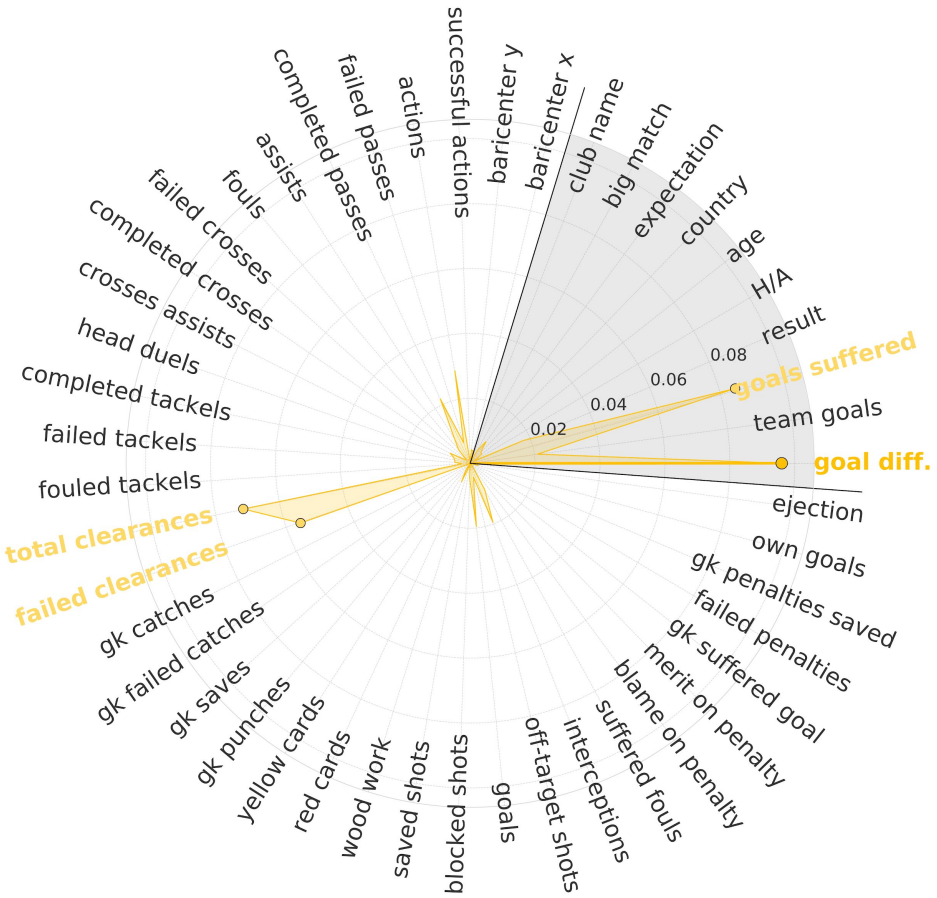
SKATER, https://oracle.github.io/Skater/



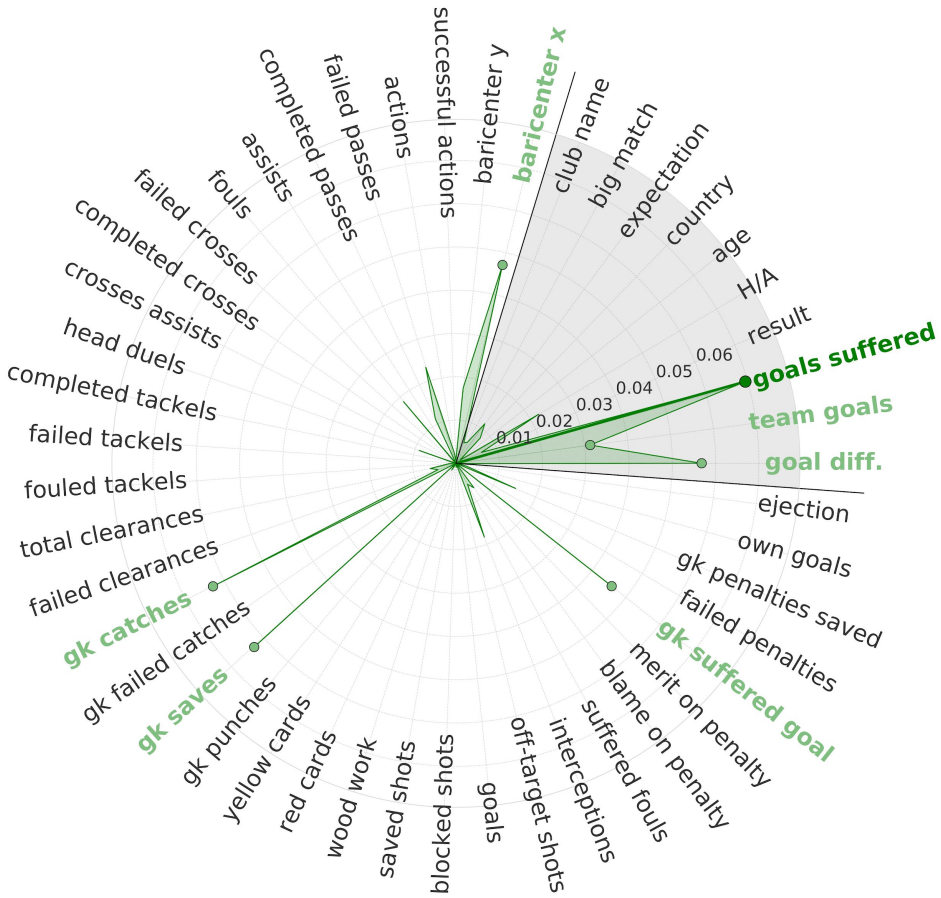

Forward
successful actions
baricenter y
baricenter x
club name
big match
expectation
country
age
H/A
result
goals suffered
team goals
goal diff.
ejection
own goals
gk penalties saved
failed penalties
gk suffered goal
merit on penalty
blame on penalty
suffered fouls
interceptions
off-target shots
goals
blocked shots
saved shots
wood work
red cards
yellow cards
gk punches
gk saves
gk failed catches
gk catches
failed clearances
total clearances
fouled tackels
failed tackels
completed tackels
head duels
crosses assists
completed crosses
failed crosses
fouls
assists
completed passes
failed passes
actions
0.01
0.02
0.03
0.04
0.05


Midfielder
successful actions
baricenter y
baricenter x
club name
big match
expectation
country
age
H/A
result
goals suffered
team goals
goal diff.
ejection
own goals
gk penalties saved
failed penalties
gk suffered goal
merit on penalty
blame on penalty
suffered fouls
interceptions
off-target shots
goals
blocked shots
saved shots
wood work
red cards
yellow cards
gk punches
gk saves
gk failed catches
gk catches
failed clearances
total clearances
fouled tackels
failed tackels
completed tackels
head duels
crosses assists
completed crosses
failed crosses
fouls
assists
completed passes
failed passes
actions
0.01
0.02
0.03
0.04
0.05
0.06

Defender



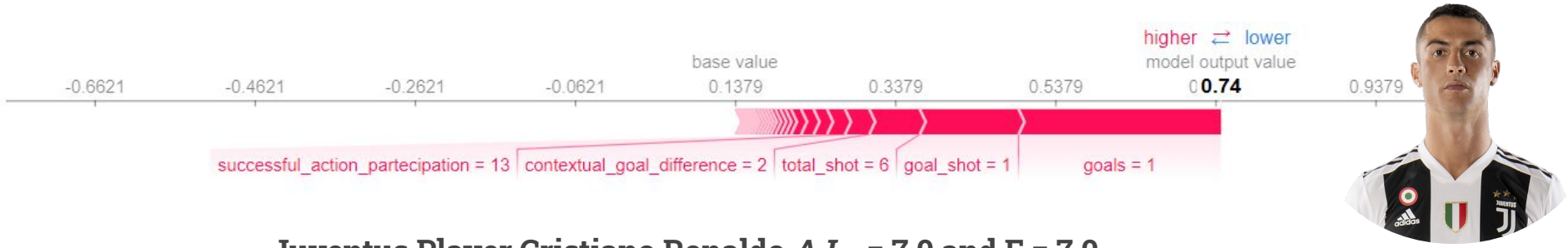
Goalkeeper



Local Explanations

SHAP, https://shap.readthedocs.io/en/latest/

Lazio Player Bastos $AJ_{(P)}$ = 5.0 and F = 5.0



Juventus Player Cristiano Ronaldo $AJ_{(P)}$ = 7.0 and F = 7.0

Local Explanations - Disagreement

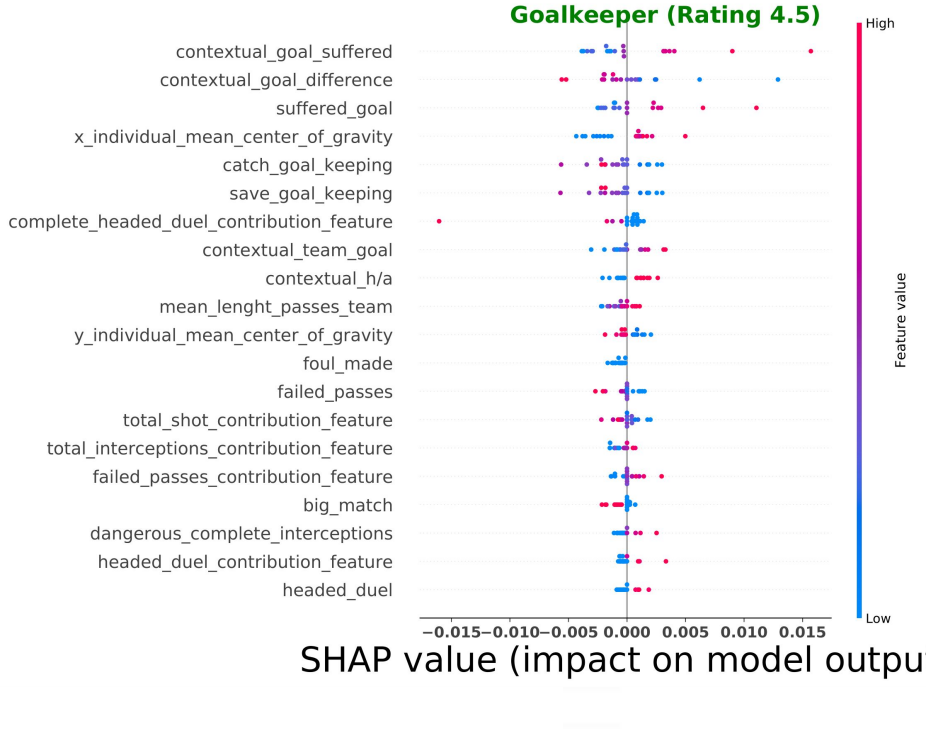
Roma Player Fazio $AJ_{(P)}$ = 7.0 and F = 5.5

Human: severe error

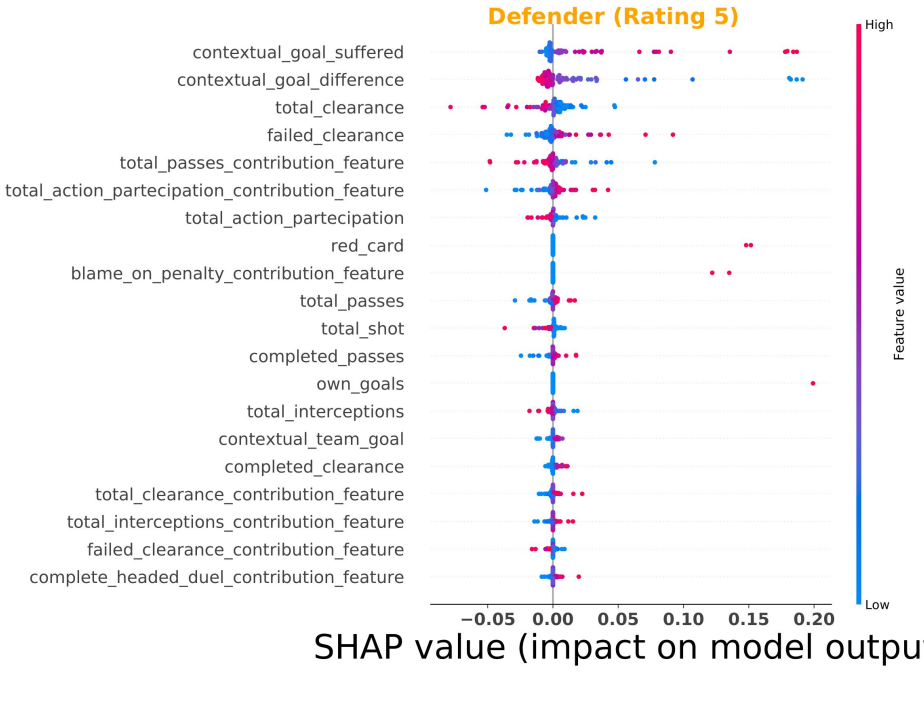
AJ: failed pass + 1


HIGHLIGHTS
ROMA
1927
S.S.LAZIO
SERIE A
TIM

Glocal Explanations

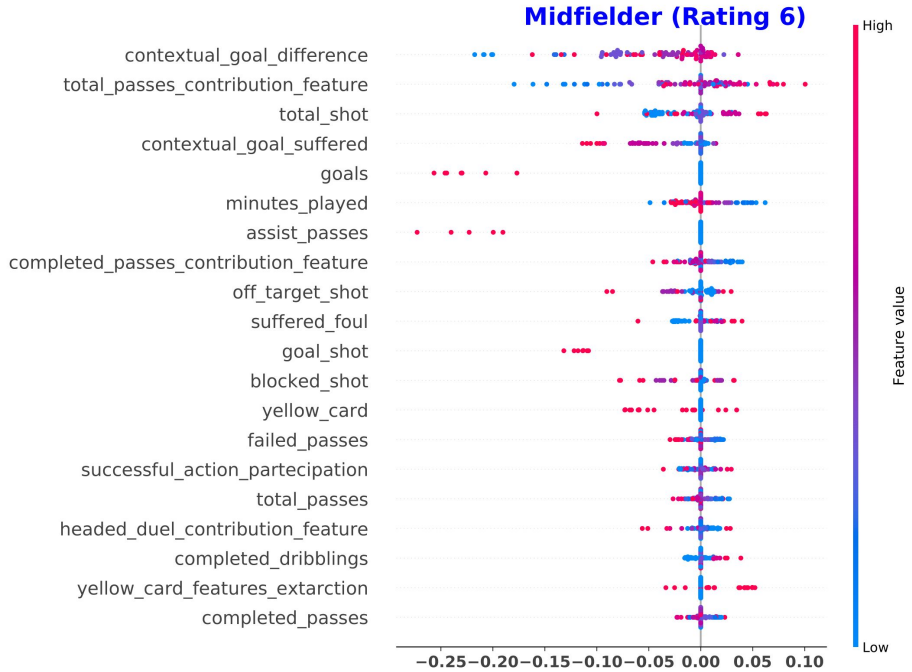


SHAP value (impact on model output)

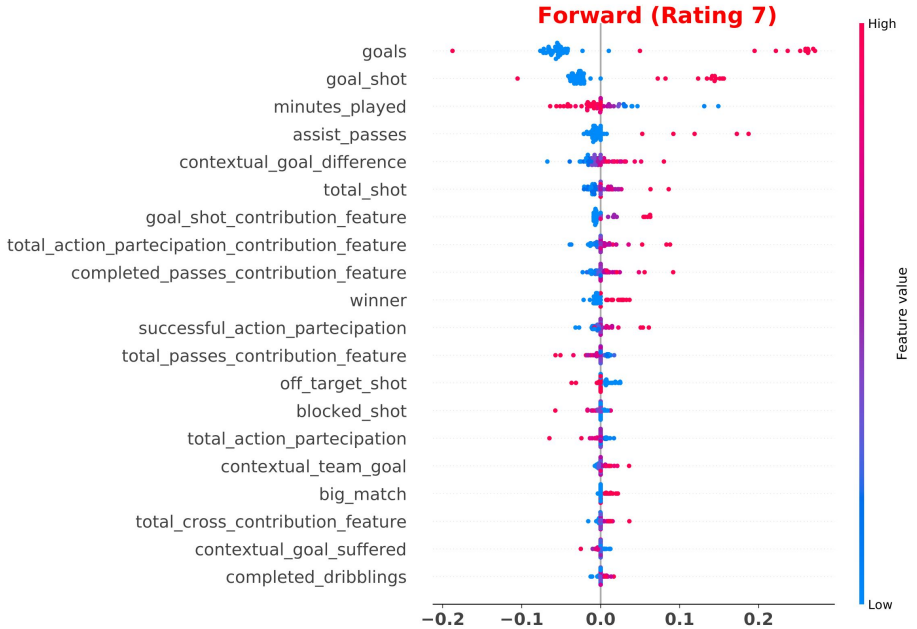


SHAP value (impact on model output)

Glocal Explanations



SHAP value (impact on model output)



SHAP value (impact on model output)

Summary of Interpretations and Explanations

- Midfielders and Forwards are strongly conditioned by crucial features such as goals and assists.
- Defenders and Goalkeepers are influenced by contextual variables.
- Predictions are influenced by the absence of crucial informations.
- The features that condition player's evaluation represent events that catch the viewer's attention.

Experiment

Setup the Environment
Find participant and organized the whole experiment. We selected **12 participants** that have to evaluate **10 matches**

Extract Ratings and Explanations for the Matches Selected
Each participant need to evaluate in mean 3 games. In total participant evaluated **19 forwards, 18 midfielders, 28 defenders and 3 goalkeeper**

Create Surveys
Using Google Surveys, create different document for each group of participant. The 12 participants were divided into **4 different surveys** (3 for each one).

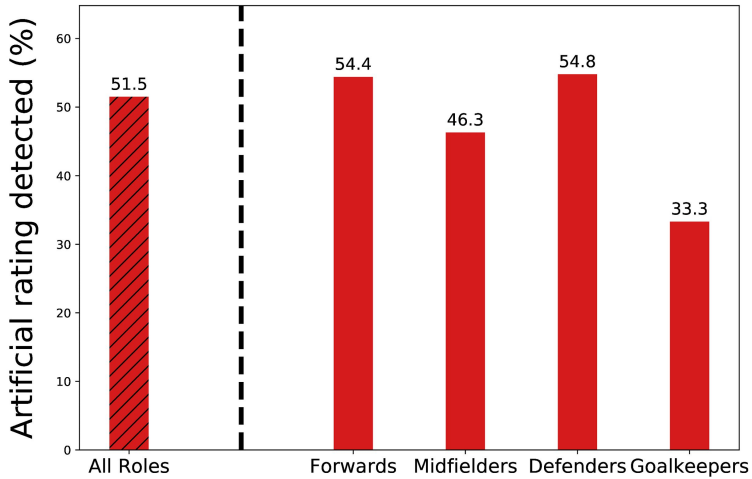
Analysis of Results
Results (**204 responses**) are investigate in order to find out the efficiency of the artificial journalist

Experiment Results

Different Experiment

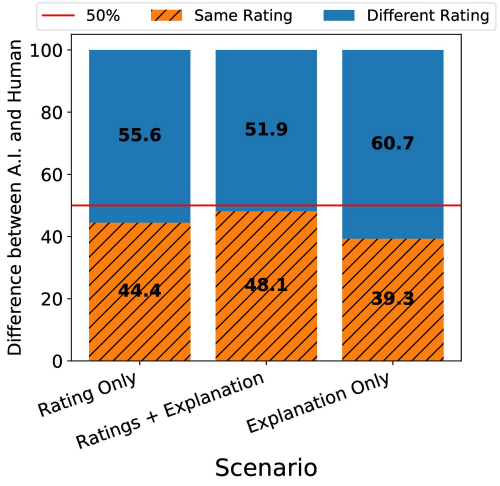
Recognition Test:

We asked the participants to recognize, for each player, what is the $AJ_{(P)}$ prediction between the artificial journalist and real journalist ratings.



Treatment Test:

We asked the participants to express, for each player, what is their rating, based on their evaluation and on information we provided to them. The possible scenarios are 3: only the $AJ_{(P)}$ prediction, only the explanation of a prediction or the $AJ_{(P)}$ prediction and the relative explanation.



Conclusions

- The Artificial Intelligence developed is able to capture the criteria behind human evaluation
- We are able to unveil the features that influence the most the evaluations of the artificial journalist.
- It would be interesting to include more sophisticated features.
- Replicate the experiments with sports journalist; in particular, redesign the whole process of experiment.
- May be a valuable support to decision of a journalist.

Consiglio Nazionale delle Ricerche

The Imitation Game: Reproducing and Explaining Human Evaluations of Soccer Performance with Artificial Intelligence

https://github.com/jonpappalord/soccer_ratings_prediction