




Soccer Game Prediction

Linying, Long, Tim, Zihang





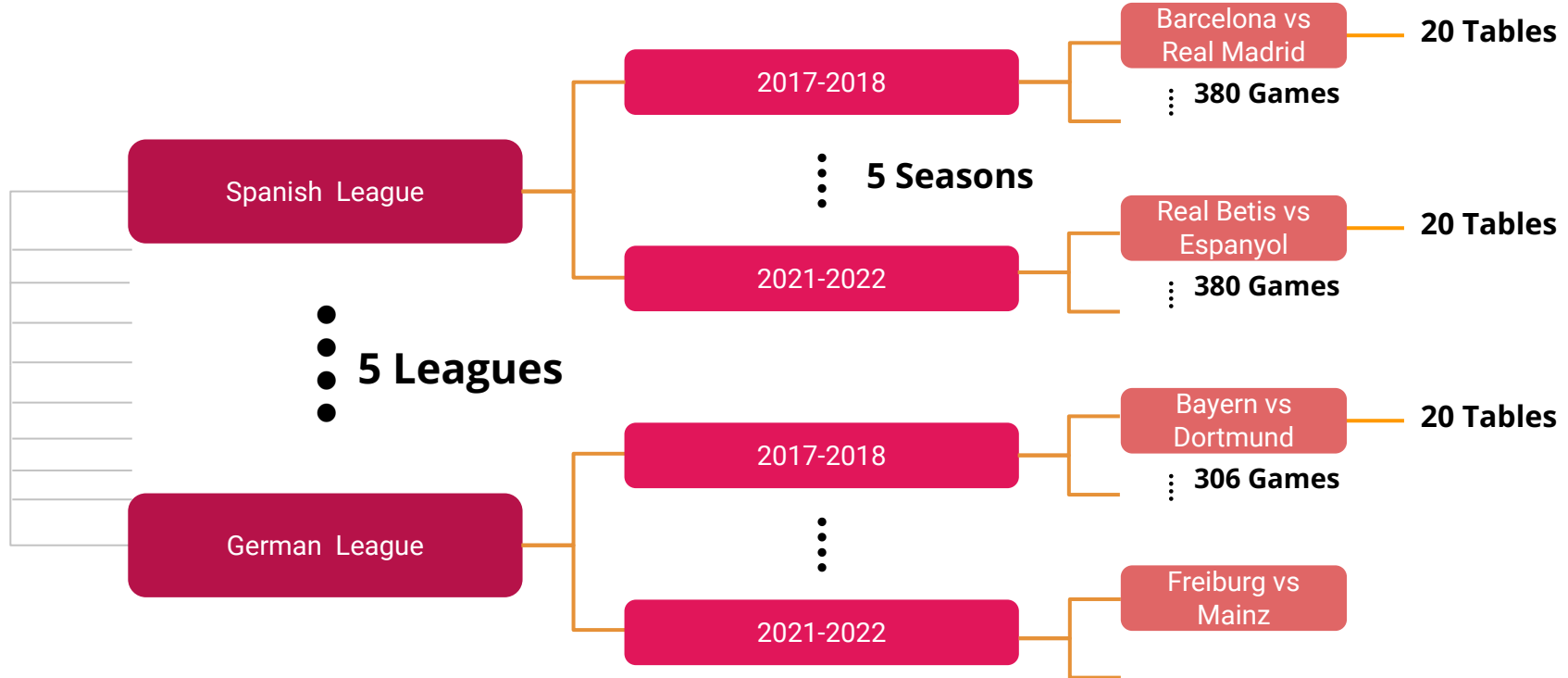
Preprocess

Model Training



Web Development

Web Hierarchy



Pre-Process

- Web Scraping
 - Get all url in the link: Request , Bs4.
 - Get stats tables directly: pandas
- Data Cleaning
 - Assign each player's data to a unique player Class
 - Model Designing at this stage
 - Feature Selection/ Engineering
 - Weights of previous match stats
 - (e.g. simply taking mean or (0.4, 0.3, 0.2, 0.1))

Scores & Fixtures 2021-2022 La Liga

Share & Export ▼ Glossary

N games

Web i

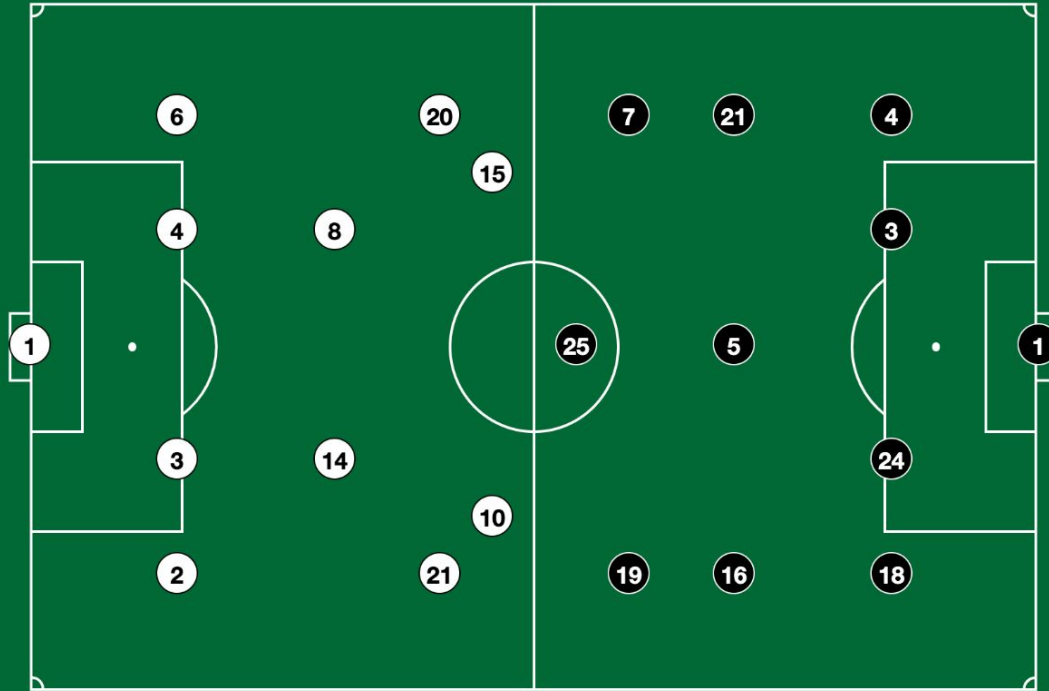
Wk	Day	Date	Time	Home	xG	Score	xG	Away	Attendance	Venue	Referee	Match Report	Notes
1	Fri	2021-08-13	21:00 (12:00)	Valencia	1.5	1-0	1.4	Getafe	9,868	Estadio de Mestalla	Jesús Gil	Match Report	
	Sat	2021-08-14	19:30 (10:30)	Mallorca	0.5	1-1	0.6	Betis	6,279	Iberostar Estadi	Antonio Matéu Lahoz	Match Report	
			19:30 (10:30)	Cádiz	0.6	1-1	0.9	Levante	6,941	Estadio Nuevo Mirandilla	Santiago Jaime	Match Report	
			22:00 (13:00)	Alavés	1.7	1-4	1.7	Real Madrid	3,968	Estadio de Mendizorroza	César Soto	Match Report	
			22:00 (13:00)	Osasuna	0.9	0-0	0.5	Espanyol	6,715	Estadio El Sadar	Miguel Ángel Ortiz Arias	Match Report	
	Sun	2021-08-15	17:30 (08:30)	Celta Vigo	2.3	1-2	1.5	Atlético Madrid	5,401	Estadio de Balaídos	José Luis Munuera	Match Report	
			20:00 (11:00)	Barcelona	2.8	4-2	0.8	Real Sociedad	20,384	Camp Nou	Alejandro Hernández	Match Report	
			22:15 (13:15)	Sevilla	3.7	3-0	0.4	Rayo Vallecano	13,962	Estadio Ramón Sánchez Pizjuán	Isidro Díaz de Mera	Match Report	
	Mon	2021-08-16	20:00 (11:00)	Villarreal	1.1	0-0	0.3	Granada	7,837	Estadio de la Cerámica	Pablo González	Match Report	
			22:00 (13:00)	Elche	1.1	0-0	0.6	Athletic Club	9,145	Estadio Manuel Martínez Valero	Adrián Cordero	Match Report	
2	Fri	2021-08-20	21:00 (12:00)	Betis	1.1	1-1	1.6	Cádiz	23,745	Estadio Benito Villamarín	Valentín Pizarro	Match Report	
	Sat	2021-08-21	17:00 (08:00)	Alavés	0.5	0-1	0.7	Mallorca	2,896	Estadio de Mendizorroza	Alejandro Muñoz	Match Report	
			19:30 (10:30)	Espanyol	1.3	0-0	0.6	Villarreal	11,095	RCDE Stadium	Jorge Figueroa	Match Report	
			19:30 (10:30)	Granada	1.1	1-1	1.5	Valencia	6,267	Estadio Nuevo Los Cármenes	Carlos del Cerro	Match Report	
			22:00 (13:00)	Athletic Club	1.4	1-1	1.0	Barcelona	9,394	San Mamés	Juan Martínez	Match Report	
	Sun	2021-08-22	17:00 (08:00)	Real Sociedad	1.9	1-0	0.5	Rayo Vallecano	7,652	Estadio Municipal de Anoeta	Santiago Jaime	Match Report	
			19:30 (10:30)	Atlético Madrid	1.2	1-0	0.2	Elche	24,926	Estadio Wanda Metropolitano	Ricardo de Burgos	Match Report	
			22:00 (13:00)	Levante	1.9	3-3	1.6	Real Madrid	9,838	Estadio Ciudad de Valencia	Guillermo Cuadra	Match Report	
	Mon	2021-08-23	20:00 (11:00)	Getafe	0.5	0-1	1.2	Sevilla	4,810	Coliseum Alfonso Pérez	José Sánchez	Match Report	
			22:00 (13:00)	Osasuna	2.0	0-0	0.1	Celta Vigo	7,189	Estadio El Sadar	Alberola Rojas	Match Report	
3	Fri	2021-08-27	20:00 (11:00)	Mallorca	1.0	1-0	0.4	Espanyol	7,529	Iberostar Estadi	Isidro Díaz de Mera	Match Report	
			22:15 (13:15)	Valencia	1.6	3-0	0.4	Alavés	10,627	Estadio de Mestalla	Miguel Ángel Ortiz Arias	Match Report	
	Sat	2021-08-28	17:00 (08:00)	Celta Vigo	1.1	0-1	1.4	Athletic Club	5,492	Estadio de Balaídos	Antonio Matéu Lahoz	Match Report	
			19:30 (10:30)	Real Sociedad	1.0	1-0	0.4	Levante	7,758	Estadio Municipal de Anoeta	José Luis Munuera	Match Report	
			19:30 (10:30)	Elche	0.5	1-1	1.3	Sevilla	10,385	Estadio Manuel Martínez Valero	Valentín Pizarro	Match Report	
			22:00 (13:00)	Betis	0.6	0-1	0.9	Real Madrid	22,590	Estadio Benito Villamarín	Alejandro Hernández	Match Report	
	Sun	2021-08-29	17:00 (08:00)	Barcelona	1.1	2-1	0.4	Getafe	26,543	Camp Nou	Pablo González	Match Report	
			19:30 (10:30)	Rayo Vallecano	2.2	4-0	0.5	Granada	583	Estadio del Rayo Vallecano	Jesús Gil	Match Report	
			19:30 (10:30)	Cádiz	1.6	2-3	2.4	Osasuna	7,452	Estadio Nuevo Mirandilla	Adrián Cordero	Match Report	
			22:00 (13:00)	Atlético Madrid	1.7	2-2	0.3	Villarreal	26,840	Estadio Wanda Metropolitano	César Soto	Match Report	
4	Sat	2021-09-11	18:30 (09:30)	Levante	1.0	1-1	2.2	Rayo Vallecano	12,974	Estadio Ciudad de Valencia	Alberola Rojas	Match Report	

Real Madrid (4-2-4-0)

- 1 [Thibaut Courtois](#)
- 2 [Dani Carvajal](#) ↻
- 3 [Éder Militão](#)
- 4 [David Alaba](#)
- 6 [Nacho](#) ↻
- 8 [Toni Kroos](#) ■ ↻
- 10 [Luka Modrić](#) ■
- 14 [Casemiro](#)
- 15 [Federico Valverde](#)
- 20 [Vinicius Júnior](#)
- 21 [Rodrygo](#) ↻

Bench

- 13 [Andriy Lunin](#)
- 45 [Diego Piñeiro](#)
- 5 [Jesús Vallejo](#)
- 7 [Eden Hazard](#)
- 11 [Marco Asensio](#) ↻
- 12 [Marcelo](#)
- 16 [Luka Jović](#)
- 17 [Lucas Vázquez](#) ↻ ■
- 19 [Dani Ceballos](#)
- 22 [Isco](#)
- 24 [Mariano](#) ↻
- 25 [Eduardo Camavinga](#) ↻ ■



Barcelona (4-3-3)

- 1 [Marc-André ter Stegen](#)
- 3 [Gerard Piqué](#)
- 4 [Ronald Araújo](#) ✳
- 5 [Sergio Busquets](#) ■
- 7 [Ousmane Dembélé](#) ↻
- 16 [Pedri](#) ↻
- 18 [Jordi Alba](#) ■ ↻
- 19 [Ferrán Torres](#) ✳
- 21 [Frenkie de Jong](#) ■ ↻
- 24 [Eric García](#)
- 25 [Pierre-Emerick Aubameyang](#)

Bench

- 13 [Neto](#)
- 36 [Arnau Tenas](#)
- 6 [Riqui Puig](#)
- 8 [Dani Alves](#) ↻
- 9 [Memphis](#) ↻
- 11 [Adama Traoré](#) ↻
- 12 [Martin Braithwaite](#)
- 14 [Nicolás González](#) ↻ ■
- 15 [Clément Lenglet](#)
- 17 [Luuk de Jong](#)
- 22 [Óscar Mingueza](#)
- 30 [Gavi](#) ↻

Web i_j , Table 0&1

$j \in \{1, 2, 3, \dots, N\}$

Team Stats

Real Madrid   Barcelona

Possession

40% 60%



Passing Accuracy

419 of 488 — 86% 89% — 639 of 720



Shots on Target

4 of 14 — 29% 56% — 10 of 18



Saves

6 of 10 — 60% 100% — 4 of 4



Cards



Real Madrid

9 Fouls
4 Corners
8 Crosses
589 Touches

Barcelona

12
3
6
802

Real Madrid

7 Tackles
13 Interceptions
5 Aerials Won
10 Clearances

Barcelona

14
7
8
9

Real Madrid

0 Offsides
3 Goal Kicks
6 Throw Ins
45 Long Balls

Barcelona

0
7
13
71


Web i_j , Table 2

$j \in \{1, 2, 3, \dots, N\}$

Real Madrid Goalkeeper Stats

Share & Export ▼

















[Glossary](#)

				Shot Stopping					Launched			Passes				Goal Kicks			Crosses			Sweeper	
Player	Nation	Age	Min	SoTA	GA	Saves	Save%	PSxG	Cmp	Att	Cmp%	Att	Thr	Launch%	AvgLen	Att	Launch%	AvgLen	Opp	Stp	Stp%	#OPA	AvgDist
Thibaut Courtois	 BEL	29-313	90	10	4	6	60.0	3.2	1	6	16.7	28	3	21.4	31.5	3	0.0	31.7	4	0	0.0	0	

Barcelona Player Stats

Share & Export ▼

[Glossary](#)

Summary	Passing	Pass Types	Defensive Actions			Possession			Miscellaneous Stats																						
						Performance										Expected			SCA		Passes				Carries		Take-Ons				
Player	#	Nation	Pos	Age	Min	Gls	Ast	PK	PKatt	Sh	SoT	CrdY	CrdR	Touches	Tkl	Int	Blocks	xG	np	xG	xAG	SCA	GCA	Cmp	Att	Cmp%	PrgP	Carries	PrgC	Att	Succ
Pierre-Emerick Aubameyang	25	 GAB	FW	32-275	70	2	1	0	0	6	5	0	0	27	0	0	1	2.1	2.1	0.6		4	1	13	18	72.2	3	18	0	0	0
Memphis	9	 NED	FW	28-035	20	0	0	0	0	1	1	0	0	19	0	0	0	0.0	0.0	0.0		0	0	10	15	66.7	1	9	1	0	0
Ferrán Torres	19	 ESP	LW	22-019	90	1	1	0	0	6	2	0	0	50	2	0	0	0.8	0.8	1.1		5	1	35	41	85.4	3	44	3	0	0
Ousmane Dembélé	7	 FRA	RW	24-309	79	0	2	0	0	3	1	0	0	59	2	0	0	0.3	0.3	0.6		5	3	40	50	80.0	3	53	7	4	2
Adama Traoré	11	 ESP	RW	26-054	11	0	0	0	0	1	0	0	0	4	0	1	0	0.0	0.0	0.0		0	0	1	2	50.0	0	1	0	0	0
Pedri	16	 ESP	LM	19-115	85	0	0	0	0	0	0	0	0	71	3	0	1	0.0	0.0	0.1		3	0	56	64	87.5	6	62	2	5	4
Nicolás González	14	 ESP	LM	20-076	5	0	0	0	0	0	0	1	0	6	0	0	0	0.0	0.0	0.0		1	0	4	4	100.0	0	4	1	0	0
Sergio Busquets	5	 ESP	CM	33-247	90	0	0	0	0	0	0	1	0	92	4	0	2	0.0	0.0	0.1		3	1	78	84	92.9	3	69	0	0	0
Frenkie de Jong	21	 NED	RM	24-312	70	0	0	0	0	0	0	1	0	58	0	1	1	0.0	0.0	0.2		3	0	53	57	93.0	9	52	1	2	2
Gavi	30	 ESP	RM	17-227	20	0	0	0	0	0	0	0	0	11	0	0	0	0.0	0.0	0.0		0	0	8	9	88.9	0	9	1	2	2
Jordi Alba	18	 ESP	LB	32-364	85	0	0	0	0	0	0	1	0	104	1	0	0	0.0	0.0	0.5		4	0	91	103	88.3	4	78	1	0	0
Dani Alves	8	 BRA	LB	38-318	5	0	0	0	0	0	0	0	0	15	0	0	0	0.0	0.0	0.0		0	0	14	15	93.3	1	10	0	0	0
Eric García	24	 ESP	CB	21-070	90	0	0	0	0	0	0	0	0	89	1	4	2	0.0	0.0	0.0		2	0	79	81	97.5	2	70	0	0	0
Gerard Piqué	3	 ESP	CB	35-046	90	0	0	0	0	0	0	0	0	91	1	1	2	0.0	0.0	0.0		3	2	71	80	88.8	7	72	1	0	0
Ronald Araújo	4	 URU	RB	23-013	90	1	0	0	0	1	1	0	0	66	0	0	3	0.1	0.1	0.0		0	0	51	59	86.4	1	46	2	0	0
Marc-André ter Stegen	1	 GER	GK	29-324	90	0	0	0	0	0	0	0	0	40	0	0	0	0.0	0.0	0.0		0	0	35	38	92.1	0	30	0	0	0
16 Players					990	4	4	0	0	18	10	4	0	802	14	7	12	3.5	3.5	3.1		33	8	639	720	88.8	43	627	20	13	10

Web i_j, Table 10, 11

$j \in \{1, 2, 3, \dots, N\}$

Current Form

Soccer, like any other sport, is all about performances

- current form of each players matters significantly to the match outcome.





Player Class

```
nameli["Barcelona_Lionel Messi"].stat
```

✓ 0.0s

Python

	Player	Nation	#	Pos	Age	Min	Gls	Ast	PK	PKatt	...	TklW.1	PKwon	PKcon	OG	Recov	Won	Lost.1	Won%	Date	Team
0	Lionel Messi	ar ARG	10.0	FW	30-057	90	0	0	0	0	...	0	0	0	0	1	0	0	NaN	2017-08-20	Barcelona
0	Lionel Messi	ar ARG	10.0	FW	30-063	90	2	0	0	1	...	0	0	0	0	2	0	0	NaN	2017-08-26	Barcelona
1	Lionel Messi	ar ARG	10.0	FW	30-077	90	3	0	0	0	...	1	0	0	0	2	0	0	NaN	2017-09-09	Barcelona
1	Lionel Messi	ar ARG	10.0	LW	30-084	90	0	1	0	0	...	0	0	0	0	0	0	0	NaN	2017-09-16	Barcelona
0	Lionel Messi	ar ARG	10.0	FW	30-087	90	4	0	1	1	...	0	0	0	0	1	0	0	NaN	2017-09-19	Barcelona
...
1	Lionel Messi	ar ARG	10.0	FW,RW	33-309	90	1	0	0	0	...	0	0	0	0	3	0	0	NaN	2021-04-29	Barcelona
1	Lionel Messi	ar ARG	10.0	FW	33-312	90	2	0	0	1	...	0	0	0	0	1	0	0	NaN	2021-05-02	Barcelona
1	Lionel Messi	ar ARG	10.0	FW	33-318	90	0	0	0	0	...	1	0	0	0	2	0	0	NaN	2021-05-08	Barcelona
0	Lionel Messi	ar ARG	10.0	FW	33-321	90	1	0	0	0	...	0	0	0	0	0	0	1	0.0	2021-05-11	Barcelona
0	Lionel Messi	ar ARG	10.0	FW,LW	33-326	90	1	0	0	0	...	1	0	0	0	3	0	1	0.0	2021-05-16	Barcelona

138 rows x 124 columns



Model-Training

Models	Accuracy (Rounding to Int)
Linear Regression (Elastic Net)	61.11%
MLP Regression	57.34%
Regression Tree	41.00%

Classification

- Given scores are within the range of $\{0,1,2,\dots,8,9\}$
- We make it a classification problem with 10 classes.



THE BIGGEST PREMIER LEAGUE WINS

2 Feb 2021		Man Utd	9-0	Saints	
4 Mar 1995		Man Utd	9-0	Ipswich	
25 Oct 2019		Saints	0-9	Leicester	
22 Nov 2009		Spurs	9-1	Wigan	
19 Sep 1999		Newcastle	8-0	Sheff Wed	
9 May 2010		Chelsea	8-0	Wigan	
23 Dec 2012		Chelsea	8-0	Aston Villa	
18 Oct 2014		Saints	8-0	Sunderland	

```
(tensor([0., 1., 2., 3., 4., 5., 6., 7., 8., 9.]),  
 tensor([2045, 2893, 2156, 1049, 435, 172, 54, 11, 6, 2]))
```

Classification

Models	Accuracy
Logistic Regression	41.53%
SVM - linear	35.77%
SVM - RBF kernel	36.59%
XGBoost	42.97%
KNN	45.09%
MLP Classification	57.04%

Future Work

1. **Expand Dataset**
2. Predict on Not only score, but **Team overall statistics** for each match(i.e., Possession, Pass Completion)
3. Try implementing **CNN** by transforming data into pseudo-picture format
 - a. Take advantage of current state of the art techniques that fine-tuned on MNIST dataset
4. Consider **Opponent overall Strength**, but not just their past performances
5. Interpret and Combine **xG** algorithms

Web Development

- Flask
 - HTML, python connector
- HTML
 - Club selection (dropdown boxes)
 - Logo
 - Result button



Users can go through our webpage, pick two teams, click on the button and the predicted score will appear.

Resources and Git Repo:

https://github.com/timc1325/Premier_League_Prediction

<https://fbref.com/en/>

https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. Retrieved from <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>



Thank you

