

Barrero - Martínez - Bondi

Football Prediction

English Premier League

Reto

¿Quién será el siguiente
ganador de la English
Premier League?



EPL

- 20 equipos por temporada
- 38 jornadas
- 1.5 goles por partido
- +£2,200M (+\$59,682M MXN)



Fuentes de datos

Historial de partidos desde 1980

- github.com/jalapic/engsoccerdata

Valor (USD) por equipo desde 2005

- Transfermarkt

Generación de datos

- Adición de columnas de utilidad

Dataset Final

season	week_day	year	month	day	home	visitor	result
hpoints	vpoints	hltposition	vltposition	hptposition			
vptposition	hvalue	vvalue					

~ 16,000 partidos desde 1980
~ 6,000 partidos desde 2005

Dataset Final

	season	week_day	year	month	day	home	visitor	hpoints	vpoints
0	2005	1	2005	8	13	2		7	1
1	2005	1	2005	8	13	17		27	0
2	2005	1	2005	8	13	18		4	1
3	2005	1	2005	8	13	26		48	1
4	2005	1	2005	8	13	28		24	1

hltposition	vltposition	hptposition	vptposition	hvalue	vvalue	result
10	6	5		7 1.008300e+08	6.590000e+07	1
4	3	17		3 1.097800e+08	2.864300e+08	3
14	12	9		10 7.090000e+07	9.538000e+07	1
9	17	16		21 7.640000e+07	5.505000e+07	1
7	5	11		4 9.463000e+07	1.874000e+08	1

Modelo Inicial

1. Predecir juegos ganados por jornada
2. Generar tabla de posiciones
3. Métricas basadas en exactitud de victorias

Pipeline

Multinomial Logistic
Regression

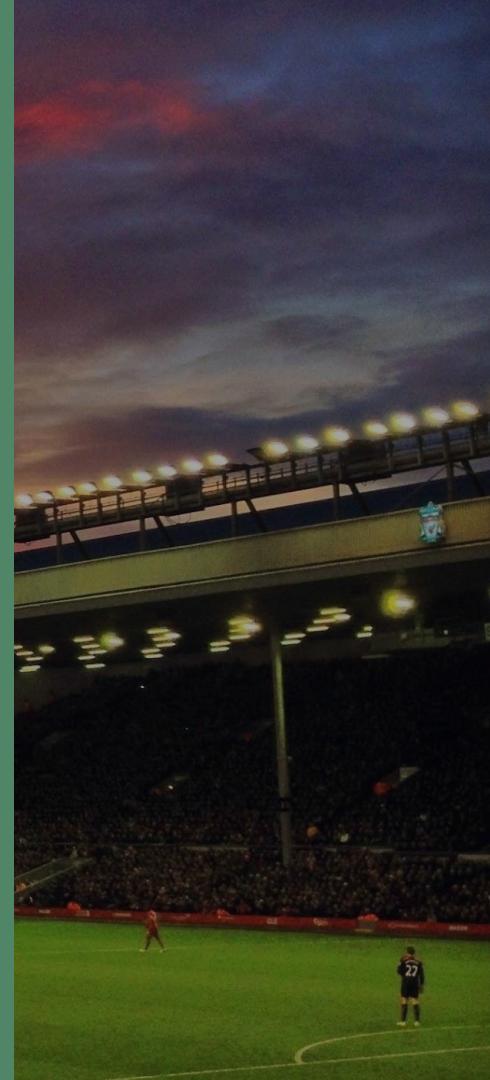
Dataset

StandardScaler

TimeSeriesSplit

KNeighborClassifier

GaussianNB



Métricas

Multinomial Logistic
Regression

KNeighborClassifier

GaussianNB

	precision	recall	f1-score	support
1	0.00	0.00	0.00	114
2	0.56	0.77	0.65	242
3	0.51	0.58	0.54	162
accuracy			0.54	518
macro avg	0.36	0.45	0.40	518
weighted avg	0.42	0.54	0.47	518

	precision	recall	f1-score	support
1	0.40	0.05	0.09	114
2	0.55	0.81	0.66	242
3	0.52	0.48	0.50	162
accuracy			0.54	518
macro avg	0.49	0.45	0.42	518
weighted avg	0.51	0.54	0.48	518

	precision	recall	f1-score	support
1	0.00	0.00	0.00	114
2	0.58	0.71	0.64	242
3	0.46	0.64	0.54	162
accuracy			0.53	518
macro avg	0.35	0.45	0.39	518
weighted avg	0.42	0.53	0.47	518

54 %

Modelo Final

1. Predecir juegos ganados por jornada
2. Generar tabla de posiciones
3. Métricas basadas en tablas de posiciones



Métricas

2018

80%

2019

	precision	recall	f1-score	support
1	0.80	0.80	0.80	5
2	0.85	0.85	0.85	13
3	0.50	0.50	0.50	2
accuracy			0.80	20
macro avg	0.72	0.72	0.72	20
weighted avg	0.80	0.80	0.80	20

	precision	recall	f1-score	support
1	0.80	0.80	0.80	5
2	0.85	0.85	0.85	13
3	0.50	0.50	0.50	2
accuracy			0.80	20
macro avg	0.72	0.72	0.72	20
weighted avg	0.80	0.80	0.80	20



Conclusiones



Conclusiones

- Mejoras en el modelo
 - a. Incorporar índice compuesto calificación por equipo
 - b. Simulación Monte Carlo
- Añadir más datos por partido
 - a. Tiros a puerta
 - b. Oportunidades de gol
 - c. Importancia de partido
- Complejidad debida a los pocos goles anotados