

1 Variables Dataset

Info básica del partido

Nombre	Significado	Tipo	Aclaración
Div	División de liga	Categórica	
Date	Fecha de juego	Fecha (datetime)	
HomeTeam	Equipo local	Texto	
AwayTeam	Equipo visitante	Texto	
Season	Temporada del partido	Texto	Abreviación (2005_06 para 2005-2006)

1.1 Resultados de partido

Nombre	Significado	Tipo	Aclaración
FTHG	Goles locales	Entero	
FTAG	Goles visitantes	Entero	
FTR	Resultado del partido	Categórico	(H = local,D = empate,A = visitante)
HTHG	Goles locales al descanso	Entero	
HTAG	Goles visitantes al descanso	Entero	
HTR	Resultado al descanso	Categórico	Igual que en FTR

1.2 Estadísticas del partido

Nombre	Significado	Tipo	Aclaración
HS	Tiros totales locales	Entero	
AS	Tiros totales visitantes	Entero	
HST	Tiros a puerta totales locales	Entero	
AST	Tiros a puerta totales visitantes	Entero	
HF	Faltas locales	Entero	
AF	Faltas visitantes	Entero	
HC	Corners locales	Entero	
AC	Corners visitantes	Entero	
HY	Tarjetas amarillas locales	Entero	
AY	Tarjetas amarillas visitantes	Entero	
HR	Tarjetas rojas locales	Entero	
AR	Tarjetas rojas visitantes	Entero	

1.3 Cuotas de apuestas

Nombre	Significado	Tipo	Aclaración
B365H	Cuota de victoria local	Decimal	
B365D	Cuota de empate	Decimal	
B365A	Cuota de victoria visitante	Decimal	

2 Variables calculadas

2.1 Ratings ELO

Probabilidad esperada de victoria calculada con la siguiente fórmula:

$$E_{\text{home}} = \frac{1}{1 + 10^{\frac{ELO_{\text{away}} - ELO_{\text{home}}}{400}}}$$

Actualización después de cada partido:

$$ELO_{\text{home}} = ELO_{\text{home}} + K \cdot (S_{\text{home}} - E_{\text{home}})$$

Nombre	Significado	Tipo	Aclaración
elo_home	Rating ELO del equipo local ANTES del partido	Decimal	Valor inicial 1500
elo_away	Rating ELO del equipo visitante ANTES del partido	Decimal	Valor inicial 1500
elo_diff	Diferencia ELO entre local y visitante	Decimal	

2.2 Características de forma

Características de forma de los últimos 7 partidos (sin contar el partido en el que se calculan)

Nombre	Significado	Tipo	Aclaración
home_avg_goals_scored_7	Goles locales	Decimal	
away_avg_goals_scored_7	Goles visitantes	Decimal	
home_avg_goals_conceded_7	Goles locales en contra	Decimal	
away_avg_goals_conceded_7	Goles visitantes en contra	Decimal	
goal_diff_form_home	Diferencia de goles promedio del equipo local	Decimal	
goal_diff_form_away	Diferencia de goles promedio del equipo visitante	Decimal	

2.3 Características estadísticas

Características de estadísticas de los últimos 7 partidos (sin contar el partido en el que se calculan)

Nombre	Significado	Tipo	Aclaración
home_avg_shots_7	Tiros locales	Decimal	
away_avg_shots_7	Tiros visitantes	Decimal	
home_avg_shots_on_target_7	Tiros a puerta locales	Decimal	
away_avg_shots_on_target_7	Tiros a puerta visitantes	Decimal	
home_avg_corners_7	Corners locales	Decimal	
away_avg_corners_7	Corners visitantes	Decimal	
home_avg_fouls_7	Faltas locales	Decimal	
away_avg_fouls_7	Faltas visitantes	Decimal	
home_avg_yellows_7	Tarjetas amarillas locales	Decimal	
away_avg_yellows_7	Tarjetas amarillas visitantes	Decimal	
home_avg_reds_7	Tarjetas rojas locales	Decimal	
away_avg_reds_7	Tarjetas rojas visitantes	Decimal	

2.4 Índices compuestos

2.4.1 attack_strength_home y attack_strength_away

Índice de fuerza ofensiva del equipo local y visitante, calculado tal que así:

$$(\text{homelaway})_avg_goals_scored_7 + 0.1 \cdot (\text{homelaway})_avg_shots_on_target_7$$

2.4.2 defense_strength_home y defense_strength_away

Índice de fuerza defensiva del equipo local y visitante, calculado tal que así:

$$\frac{1}{(\text{homelaway})_avg_goals_conceded_7 + 0.1 \cdot (\text{homelaway})_avg_yellows_7 + 10^{-3}}$$

2.4.3 discipline_index_home y discipline_index_away

Índice de disciplina (el n de tarjetas que te sacan) del equipo local y visitante calculado tal que así:

$$1 - \frac{0.5 \cdot (\text{homelaway})_avg_yellows_7 + 1.5 \cdot (\text{homelaway})_avg_reds_7}{10}$$

2.5 Características de mercado (probabilidades)

2.5.1 B365H_prob, B365D_prob y B365A_prob

Probabilidades implícitas de victoria local, empate o victoria visitante. Se calcula con la siguiente fórmula, para eliminar el margen de la casa de apuestas:

$$prob_raw = \frac{1}{B365(H|D|A)}$$

Normalización para eliminar el margen:

$$sum_probs = \frac{1}{B365H} + \frac{1}{B365D} + \frac{1}{B365A}$$
$$B365(H|D|A) = \frac{prob_raw}{sum_probs}$$

2.5.2 prob_diff_home_away

Diferencia entre probabilidades local y visitante

2.5.3 prob_fav_margin

Margen del favorito (diferencia entre máxima y mínima probabilidad). Se calcula así:

$$\max(B365H_prob, B365D_prob, B365A_prob) - \min(B365H_prob, B365D_prob, B365A_prob)$$

2.6 Datos de Alineaciones

Nombre	Significado	Tipo	Aclaración
Home_Lineup_List	Lista de la alineación local	String	
Away_Lineup_List	Lista de la alineación visitante	String	

2.6.1 home_team_norm y away_team_norm

Variables usadas para conectar varios datasets.

2.6.2 home_team_value y away_team_value

Valor de mercado de cada alineación local y visitante. Se calcula mediante los valores de los jugadores sacados de transfermarkt, con el valor respectivo a la temporada en la que esta datado.

Se normaliza el nombre del jugador (los datos vienen de distintas fuentes, es necesario normalizar) y si el valor del fuzzy match es mayor a 0.55, se usa el valor almacenado de ese jugador.

2.7 derby

Valor si el partido es derby/rivalidad histórica.

Datos sacados de: https://en.wikipedia.org/wiki/Spanish_football_rivalries

Valor 0 si no es derby, valor 1 si lo es.