

TITOL DEL PROJECTE

Predicció de resultats de futbol (1a divisió espanyola) amb un anàlisis segons si juguen com a local o visitant i històric

És possible predir el resultat (1-X-2)* d'un partit de futbol i en quin % d'acuraccy. I el número de gols?

1. Presentació del conjunt de dades escollit.

- El conjunt de dades escollit son els resultats de la primera lliga espanyola des de la temporada 1928-1929. També un altre dataset de la classificació dels equips per crear noves columnes.
- Aquest conjunt el transformaré amb noves columnes que descriu en el següent apartat.
- També d'entrada el reduiré perquè la primera temporada sigui la del 1980-1981 (la primera després que jo vaig néixer) i més futbol modern. De totes maneres he utilitzat totes les dades per poder anar creant les columnes.
- Les variables escollides son les que jo crec que poden ser influents segons el meu coneixement o el que vull observar a l'hora de definir un resultat. La gran majoria les crearé jo a partir de la classificació i els últims resultats com a locals i visitants. M'hagués agradat agafar més variables influents o estudiar més opcions però de moment treballaré amb les que tinc. És una de les opcions d'ampliació del projecte si tinc més temps.
- Les dades no serán iguals per tots els equips ja que alguns tindran més o menys participació.
- El target pot ser confós perquè en un primer moment pot semblar que sigui predir 1-X-2 però realment el que volem predir és el número de gols de cada equip que això acabarà devirant en el resultat de la quiniela 1-X-2. Per fer-ho ens trobem que tenim dos columnes targets (Gols locals i gols visitants) i per tant ho haurem de separar en dos datasets. Locals i visitants. Per tenir només un target.
- Pot sortir el dubte si realment es tracta d'un problema supervisat o un problema de classificació. Amb 1-X-2 sembla clar que pot ser de classificació però a l'hora de calcular els gols és veritat que més o menys va del 0 al 6 però podria ser superior... Faré un estudi dels dos tipus de ML.

Tipologia les dades: Categòriques, numèriques, ordinals, dummies, dates.

Sector: Enteniment, esport, futbol / apostes.

Fonts: bdfutbol.com

Context: Dades històriques per veure com afecten els diferents equips i resultats en futures prediccions. Intentar predir jornada 35-36 de la temporada 2022-2023 de la primera divisió espanyola.

2. Definició de les variables.

Marco les originals, la resta les crearé jo.

1. DATASET RESULTATS

Nom Variable	Definició
Temp	Identificación de la temporada
Jor	Jornada
Data	Dia, mes i any en que es va jugar el partit
Local	Equip que juga com a local
Visitant	Equip que juga com a visitant
L_Goal	Gols de l'equip local
V_Goal	Gols de l'equip visitant
1-X-2	Resutat del partit (1= guanyador del local, 3= Empat, 2 = Guanyador visitant)
LP	Punts resultants d'aquest partit per l'equip local
VP	Punts resultants d'aquest partit per l'equip local
Period	Agrupar les jornades en diferents períodes ja que crec que segons el moment de la temporada afecta alguns partits
L_Han	Handicap segons la posició de l'equip local en les 10 últimes temporades
L_10r	Mitja de punts dels últims 10 resultats de l'equip que juga com a local
L_10gf_cL	Mitja de gols a favor de l'equip local en els ultims 10 partits com a local
L_10ga_cL	Mitja de gols en contra de l'equip local en els ultims 10 partits com a local
LV_10cL	Victories com a local de l'equip local en els últims 10 partits
LE_10cL	Maxim de partits consecutius en que l'equip local no ha fet un gol
LD_10cL	Maxim de partits consecutis en que l'equip local no ha rebut un gol
V_Han	Handicap segons la posició de l'equip local en les 10 últimes temporades
V_10r	Mitja de punts dels últims 10 resultats de l'equip que juga com a local
V_10gf_cV	Mitja de gols a favor de l'equip local en els ultims 10 partits com a local
V_10ga_cV	Mitja de gols en contra de l'equip local en els ultims 10 partits com a local
VV_10cV	Victories com a local de l'equip local en els últims 10 partits
LE_10cV	Maxim de partits consecutius en que l'equip local no ha fet un gol
LD_10cV	Maxim de partits consecutis en que l'equip local no ha rebut un gol

2. DATASET CLASSIFICACIÓ

Nom variable	Definició
Temporada	Identificació de la temporada
Posición	Posició de cada equip
Equipo	Equip

PT	Punts aconseguits en aquella temporada
10_Clas	Mitja de la posició en les darreres 10 temporades
Handicap	Handicap de l'1 al 4 segons la columna anterior.

3. Presentació dels objectius.

1. ML amb indiferència del rival -> Veure si es possible classificar en 1-X-2 el resultat d'un partit de futbol i també calcular els gols.
2. ML basant-me amb les dades només amb els mateixos partits .
3. Que passa si treiem els partits del Barça i Madrid (els que més guanyen i en el cas d'apostes menys diners donen).
4. Extreure conclusions a partir d'aquí.
5. Comprovar els resultats que ens dona el model amb el que ha passat a la vida.

3.1 Possibles ampliacions futures:

1. Crear una Api Local amb el programa
2. Afegir variables no numèriques.
3. Investigar més models i millor els hiperparametres.