model_worldcupwinner

March 28, 2023

[1]:
```
import pandas as pd
import pickle
from scipy.stats import poisson
```

[2]:
```
dict_table = pickle.load(open('dict_table', 'rb'))
df_historical_data = pd.read_csv('clean_fifa_worldcup_historical_data.csv')
df_fixture = pd.read_csv('clean_fifa_worldcup_fixture.csv')
```

[3]:
```
#Dividimos el dataframe en df_home y df_away
df_home = df_historical_data[['HomeTeam', 'HomeGoals', 'AwayGoals']]
df_away = df_historical_data[['AwayTeam', 'HomeGoals', 'AwayGoals']]

#Renombramos algunas columnas
df_home = df_home.rename(columns={'HomeTeam': 'Team', 'HomeGoals':␣
 ↪'GoalsScored', 'AwayGoals': 'GoalsConceded'})
df_away = df_away.rename(columns={'AwayTeam': 'Team', 'HomeGoals':␣
 ↪'GoalsConceded', 'AwayGoals': 'GoalsScored'})
```

[4]:
```
#Concatenamos df_home y df_away, hacer group por team y calcular el promedio
df_team_strength = pd.concat([df_home, df_away], ignore_index=True).
 ↪groupby('Team').mean()
df_team_strength
```

[4]:
```
               GoalsScored  GoalsConceded
Team
Algeria           1.000000       1.461538
Angola            0.333333       0.666667
Argentina         1.681159       1.130435
Australia         1.000000       2.000000
Austria           1.565217       1.608696
…                    …             …
United States     1.121212       1.878788
Uruguay           1.680000       1.340000
Wales             0.800000       0.800000
West Germany      2.127660       1.382979
Yugoslavia        1.592593       1.296296
```

```
[81 rows x 2 columns]
```

[5]:

```
#funcion predict_points (distribucion de Poisson)
def predict_points(home, away):
    if home in df_team_strength.index and away in df_team_strength.index:
        # goals_scored * goals_conceded
        lamb_home = df_team_strength.at[home,'GoalsScored'] * df_team_strength.
 ↪at[away,'GoalsConceded']
        lamb_away = df_team_strength.at[away,'GoalsScored'] * df_team_strength.
 ↪at[home,'GoalsConceded']
        prob_home, prob_away, prob_draw = 0, 0, 0
        for x in range(0,11): #number of goals home team
            for y in range(0, 11): #number of goals away team
                p = poisson.pmf(x, lamb_home) * poisson.pmf(y, lamb_away)
                if x == y:
                    prob_draw += p
                elif x > y:
                    prob_home += p
                else:
                    prob_away += p

        points_home = 3 * prob_home + prob_draw
        points_away = 3 * prob_away + prob_draw
        return (points_home, points_away)
    else:
        return (0, 0)
```

[6]:

```
#testeando la funcion para ver que valores arroja
print(predict_points('England', 'United States'))
print(predict_points('Argentina', 'Mexico'))
print(predict_points('Qatar (H)', 'Ecuador')) # Qatar vs Team X -> 0 points to␣
 ↪both
```

```
(2.2413414036134265e+00, 0.5894589951463173)
(2.3035783626598416e+00, 0.5439755479545922)
(0, 0)
```

[7]:

```
#separo los partidos en fase de grupos, octavos, cuartos, semis y final
df_fixture_group_48 = df_fixture[:48].copy()
df_fixture_knockout = df_fixture[48:56].copy()
df_fixture_quarter = df_fixture[56:60].copy()
df_fixture_semi = df_fixture[60:62].copy()
df_fixture_final = df_fixture[62:].copy()
```

[8]:

```
#correr todos los partidos de fase de grupos y actualizar las tablas de cada␣
 ↪grupo
for group in dict_table:
    teams_in_group = dict_table[group]['Team'].values
    df_fixture_group_6 = df_fixture_group_48[df_fixture_group_48['home'].
 ↪isin(teams_in_group)]
    for index, row in df_fixture_group_6.iterrows():
        home, away = row['home'], row['away']
        points_home, points_away = predict_points(home, away)
        dict_table[group].loc[dict_table[group]['Team'] == home, 'Pts'] +=␣
 ↪points_home
        dict_table[group].loc[dict_table[group]['Team'] == away, 'Pts'] +=␣
 ↪points_away

    dict_table[group] = dict_table[group].sort_values('Pts', ascending=False).
 ↪reset_index()
    dict_table[group] = dict_table[group][['Team', 'Pts']]
    dict_table[group] = dict_table[group].round(0)
```

[9]:
```
#mostrar tabla actualizada
dict_table['Grupo A']
```

[9]:
```
          Team   Pts
0  Netherlands  10.0
1      Senegal   9.0
2      Ecuador   6.0
3    Qatar (H)   0.0
```

[11]:
```
#octavos de final
df_fixture_knockout
```

[11]:
```
           home         score           away  year
48  Netherlands           3-1  United States  2022
49    Argentina           2-1      Australia  2022
50       France           3-1         Poland  2022
51      England           3-0        Senegal  2022
52        Japan  1-1 (a.e.t.)        Croatia  2022
53       Brazil           4-1    South Korea  2022
54      Morocco  0-0 (a.e.t.)          Spain  2022
55     Portugal           6-1    Switzerland  2022
```

[14]:
```
for group in dict_table:
    group_winner = dict_table[group].loc[0, 'Team']
    runners_up = dict_table[group].loc[1, 'Team']
    df_fixture_knockout.replace({f'Winners {group}':group_winner,
                                 f'Runners-up {group}':runners_up},␣
 ↪inplace=True)

df_fixture_knockout['winner'] = '?'
df_fixture_knockout
```

[14]:

```
           home         score           away  year winner
48  Netherlands           3-1  United States  2022      ?
49    Argentina           2-1      Australia  2022      ?
50       France           3-1         Poland  2022      ?
51      England           3-0        Senegal  2022      ?
52        Japan  1-1 (a.e.t.)        Croatia  2022      ?
53       Brazil           4-1    South Korea  2022      ?
54      Morocco  0-0 (a.e.t.)          Spain  2022      ?
55     Portugal           6-1    Switzerland  2022      ?
```

[15]:

```
def get_winner(df_fixture_updated):
    for index, row in df_fixture_updated.iterrows():
        home, away = row['home'], row['away']
        points_home, points_away = predict_points(home, away)
        if points_home > points_away:
            winner = home
        else:
            winner = away
        df_fixture_updated.loc[index, 'winner'] = winner
    return df_fixture_updated
```

[16]:

```
get_winner(df_fixture_knockout)
```

[16]:

```
           home         score           away  year       winner
48  Netherlands           3-1  United States  2022  Netherlands
49    Argentina           2-1      Australia  2022    Argentina
50       France           3-1         Poland  2022       France
51      England           3-0        Senegal  2022      England
52        Japan  1-1 (a.e.t.)        Croatia  2022      Croatia
53       Brazil           4-1    South Korea  2022       Brazil
54      Morocco  0-0 (a.e.t.)          Spain  2022        Spain
55     Portugal           6-1    Switzerland  2022     Portugal
```

[17]:

```
#Cuartos de final
def update_table(df_fixture_round_1, df_fixture_round_2):
    for index, row in df_fixture_round_1.iterrows():
        winner = df_fixture_round_1.loc[index, 'winner']
        match = df_fixture_round_1.loc[index, 'score']
        df_fixture_round_2.replace({f'Winners {match}':winner}, inplace=True)
    df_fixture_round_2['winner'] = '?'
    return df_fixture_round_2
```

[18]:

```
update_table(df_fixture_knockout, df_fixture_quarter)
```

[18]:

```
           home         score       away  year winner
56      Croatia  1-1 (a.e.t.)     Brazil  2022      ?
57  Netherlands  2-2 (a.e.t.)  Argentina  2022      ?
```

```
      58      Morocco           1-0   Portugal  2022      ?
      59      England           1-2     France  2022      ?
```

```
[19]: get_winner(df_fixture_quarter)
```

```
[19]:            home         score       away  year     winner
      56      Croatia  1-1 (a.e.t.)     Brazil  2022     Brazil
      57  Netherlands  2-2 (a.e.t.)  Argentina  2022  Argentina
      58      Morocco           1-0   Portugal  2022   Portugal
      59      England           1-2     France  2022     France
```

```
[20]: #Semifinal
      update_table(df_fixture_quarter, df_fixture_semi)
```

```
[20]:          home score     away  year winner
      60  Argentina    3-0  Croatia  2022      ?
      61     France    2-0  Morocco  2022      ?
```

```
[21]: get_winner(df_fixture_semi)
```

```
[21]:          home score     away  year     winner
      60  Argentina    3-0  Croatia  2022  Argentina
      61     France    2-0  Morocco  2022     France
```

```
[22]: update_table(df_fixture_semi, df_fixture_final)
```

```
[22]:          home         score     away  year winner
      62    Croatia           2-1  Morocco  2022      ?
      63  Argentina  3-3 (a.e.t.)   France  2022      ?
```

```
[23]: get_winner(df_fixture_final)
```

```
[23]:          home         score     away  year   winner
      62    Croatia           2-1  Morocco  2022  Croatia
      63  Argentina  3-3 (a.e.t.)   France  2022   France
```