

Task3

November 9, 2023

Task 3

We need to predict the probability of scores. For example, given a match data, output the probability of 3-1 score. Describe your idea/code to solve the problem.

Importing libraries

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[2]: df_train = pd.read_csv("./train.csv")
```

```
/var/folders/r8/_hql2n5j5_wf6w5j15gn300000gn/T/ipykernel_21326/2794806720.py:1
: DtypeWarning: Columns (7) have mixed types. Specify dtype option on import or
set low_memory=False.
df_train = pd.read_csv("./train.csv")
```

```
[3]: df_train.head()
```

```
[3]:
```

	id	target	home_team_name	away_team_name	\
0	11906497	away	Newell's Old Boys	River Plate	
1	11984383	home	Real Estelí	Deportivo Las Sabanas	
2	11983301	draw	UPNFM	Marathón	
3	11983471	away	León	Morelia	
4	11883005	home	Cobán Imperial	Iztapa	

	match_date	league_name	league_id	is_cup	\
0	2019-12-01 00:45:00	Superliga	636	False	
1	2019-12-01 01:00:00	Primera Division	752	False	
2	2019-12-01 01:00:00	Liga Nacional	734	False	
3	2019-12-01 01:00:00	Liga MX	743	False	
4	2019-12-01 01:00:00	Liga Nacional	705	False	

	home_team_coach_id	away_team_coach_id	...	away_team_history_league_id_1	\
0	468196.0	468200.0	...	1122.0	
1	516788.0	22169161.0	...	752.0	
2	2510608.0	456313.0	...	734.0	
3	1552508.0	465797.0	...	743.0	
4	429958.0	426870.0	...	705.0	

	away_team_history_league_id_2	away_team_history_league_id_3 \
0	642.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	743.0
4	705.0	705.0

	away_team_history_league_id_4	away_team_history_league_id_5 \
0	636.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	743.0
4	705.0	705.0

	away_team_history_league_id_6	away_team_history_league_id_7 \
0	1122.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	743.0
4	705.0	705.0

	away_team_history_league_id_8	away_team_history_league_id_9 \
0	642.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	746.0
4	705.0	705.0

	away_team_history_league_id_10
0	1122.0
1	752.0
2	734.0
3	743.0
4	705.0

[5 rows x 190 columns]

```
[4]: df_train_ts = pd.read_csv("./train_target_and_scores.csv")
```

```
[5]: df_train_ts.head()
```

```
[5]:
```

	id	score	target
0	11906497	2-3	away
1	11984383	1-0	home
2	11983301	2-2	draw
3	11983471	1-2	away

```
4 11883005 1-0 home
```

```
[6]: df_test = pd.read_csv("test.csv")
```

```
[7]: df_test.head()
```

```
[7]:      id home_team_name away_team_name      match_date league_name \
0  17761448      team home      team away 2021-05-01 00:15:00  Division 1
1  17695487      team home      team away 2021-05-01 00:30:00      Liga MX
2  17715496      team home      team away 2021-05-01 01:00:00  Paulista A2
3  17715493      team home      team away 2021-05-01 01:00:00  Paulista A2
4  17715492      team home      team away 2021-05-01 01:00:00  Paulista A2

      league_id is_cup home_team_coach_id away_team_coach_id \
0           755  False           9605704.0           30866250.0
1           743  False           10319425.0           460370.0
2          1314  False           440905.0           220998.0
3          1314  False           21036670.0           1553238.0
4          1314  False           3618158.0           37400444.0

      home_team_history_match_date_1 ... away_team_history_league_id_1 \
0           2021-04-28 00:30:00 ...           755.0
1           2021-04-24 02:06:00 ...           743.0
2           2021-04-28 23:00:00 ...          1314.0
3           2021-04-29 01:00:00 ...          1314.0
4           2021-04-29 01:00:00 ...          1314.0

      away_team_history_league_id_2 away_team_history_league_id_3 \
0              755.0              755.0
1              743.0              743.0
2             1314.0             1314.0
3             1314.0             1314.0
4             1314.0             1314.0

      away_team_history_league_id_4 away_team_history_league_id_5 \
0              755.0              755.0
1              743.0              743.0
2             1314.0             1314.0
3             1314.0             1314.0
4             1314.0             1314.0

      away_team_history_league_id_6 away_team_history_league_id_7 \
0              755.0              755.0
1              743.0              743.0
2             1314.0             1314.0
3             1314.0             1314.0
4             1314.0             1314.0
```

	away_team_history_league_id_8	away_team_history_league_id_9 \
0	755.0	755.0
1	743.0	743.0
2	1314.0	1315.0
3	1314.0	1746.0
4	1314.0	1314.0

	away_team_history_league_id_10
0	755.0
1	743.0
2	1315.0
3	1746.0
4	1314.0

[5 rows x 189 columns]

Data preprocess

Based on features such as “home_team_coach_id” from “train.csv” and “score” from “train_target_and_scores.csv”

```
[8]: update_df_train = pd.concat([df_train_ts['score'], df_train],axis=1)
      update_df_train.head()
```

	score	id	target	home_team_name	away_team_name \
0	2-3	11906497	away	Newell's Old Boys	River Plate
1	1-0	11984383	home	Real Estelí	Deportivo Las Sabanas
2	2-2	11983301	draw	UPNFM	Marathón
3	1-2	11983471	away	León	Morelia
4	1-0	11883005	home	Cobán Imperial	Iztapa

	match_date	league_name	league_id	is_cup \
0	2019-12-01 00:45:00	Superliga	636	False
1	2019-12-01 01:00:00	Primera Division	752	False
2	2019-12-01 01:00:00	Liga Nacional	734	False
3	2019-12-01 01:00:00	Liga MX	743	False
4	2019-12-01 01:00:00	Liga Nacional	705	False

	home_team_coach_id ...	away_team_history_league_id_1 \
0	468196.0 ...	1122.0
1	516788.0 ...	752.0
2	2510608.0 ...	734.0
3	1552508.0 ...	743.0
4	429958.0 ...	705.0

	away_team_history_league_id_2	away_team_history_league_id_3 \
0	642.0	636.0

1	752.0	752.0
2	734.0	734.0
3	743.0	743.0
4	705.0	705.0

	away_team_history_league_id_4	away_team_history_league_id_5 \
0	636.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	743.0
4	705.0	705.0

	away_team_history_league_id_6	away_team_history_league_id_7 \
0	1122.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	743.0
4	705.0	705.0

	away_team_history_league_id_8	away_team_history_league_id_9 \
0	642.0	636.0
1	752.0	752.0
2	734.0	734.0
3	743.0	746.0
4	705.0	705.0

	away_team_history_league_id_10
0	1122.0
1	752.0
2	734.0
3	743.0
4	705.0

[5 rows x 191 columns]

```
[52]: missval_cols= update_df_train.columns[update_df_train.isnull().any()].tolist()
      for i in range(len(missval_cols)):
          update_df_train_ = update_df_train.drop(update_df_train.index
          ↳ [update_df_train[update_df_train[missval_cols[i]].isnull()].index.tolist()])
      update_df_train_.head()
```

```
[52]: score      id target      home_team_name      away_team_name \
0    2-3  11906497  away  Newell's Old Boys      River Plate
1    1-0  11984383  home      Real Estelí  Deportivo Las Sabanas
2    2-2  11983301  draw                UPNFM      Marathón
3    1-2  11983471  away                León      Morelia
```

4 1-0 11883005 home Cobán Imperial Iztapa

	match_date	league_name	league_id	is_cup	\
0	2019-12-01 00:45:00	Superliga	636	False	
1	2019-12-01 01:00:00	Primera Division	752	False	
2	2019-12-01 01:00:00	Liga Nacional	734	False	
3	2019-12-01 01:00:00	Liga MX	743	False	
4	2019-12-01 01:00:00	Liga Nacional	705	False	

	home_team_coach_id	...	away_team_history_league_id_1	\
0	468196.0	...	1122.0	
1	516788.0	...	752.0	
2	2510608.0	...	734.0	
3	1552508.0	...	743.0	
4	429958.0	...	705.0	

	away_team_history_league_id_2	away_team_history_league_id_3	\
0	642.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	743.0	
4	705.0	705.0	

	away_team_history_league_id_4	away_team_history_league_id_5	\
0	636.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	743.0	
4	705.0	705.0	

	away_team_history_league_id_6	away_team_history_league_id_7	\
0	1122.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	743.0	
4	705.0	705.0	

	away_team_history_league_id_8	away_team_history_league_id_9	\
0	642.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	746.0	
4	705.0	705.0	

	away_team_history_league_id_10
0	1122.0
1	752.0

```

2          734.0
3          743.0
4          705.0

```

[5 rows x 191 columns]

```

[21]: missval_cols= df_test.columns[df_test.isnull().any()].tolist()
      for i in range(len(missval_cols)):
          update_df_test = df_test.drop(df_test.
          ↪index[df_test[df_test[missval_cols[i]].isnull()].index.tolist()])
      update_df_test.head()

```

```

[21]:      id home_team_name away_team_name      match_date league_name \
0  17761448      team home      team away  2021-05-01 00:15:00  Division 1
1  17695487      team home      team away  2021-05-01 00:30:00      Liga MX
2  17715496      team home      team away  2021-05-01 01:00:00  Paulista A2
3  17715493      team home      team away  2021-05-01 01:00:00  Paulista A2
4  17715492      team home      team away  2021-05-01 01:00:00  Paulista A2

```

```

      league_id is_cup home_team_coach_id away_team_coach_id \
0          755  False          9605704.0          30866250.0
1          743  False          10319425.0          460370.0
2          1314  False          440905.0          220998.0
3          1314  False          21036670.0          1553238.0
4          1314  False          3618158.0          37400444.0

```

```

      home_team_history_match_date_1 ... away_team_history_league_id_1 \
0          2021-04-28 00:30:00 ...          755.0
1          2021-04-24 02:06:00 ...          743.0
2          2021-04-28 23:00:00 ...          1314.0
3          2021-04-29 01:00:00 ...          1314.0
4          2021-04-29 01:00:00 ...          1314.0

```

```

      away_team_history_league_id_2 away_team_history_league_id_3 \
0          755.0          755.0
1          743.0          743.0
2          1314.0          1314.0
3          1314.0          1314.0
4          1314.0          1314.0

```

```

      away_team_history_league_id_4 away_team_history_league_id_5 \
0          755.0          755.0
1          743.0          743.0
2          1314.0          1314.0
3          1314.0          1314.0
4          1314.0          1314.0

```

	away_team_history_league_id_6	away_team_history_league_id_7 \
0	755.0	755.0
1	743.0	743.0
2	1314.0	1314.0
3	1314.0	1314.0
4	1314.0	1314.0

	away_team_history_league_id_8	away_team_history_league_id_9 \
0	755.0	755.0
1	743.0	743.0
2	1314.0	1315.0
3	1314.0	1746.0
4	1314.0	1314.0

	away_team_history_league_id_10
0	755.0
1	743.0
2	1315.0
3	1746.0
4	1314.0

[5 rows x 189 columns]

```
[9]: update_df_train_drop = update_df_train.copy()
update_df_train_drop.drop(['home_team_name',
↪ 'away_team_name', 'target', 'score', 'id'], axis='columns', inplace=True)
update_df_train_drop.head()
```

```
[9]:      match_date      league_name  league_id is_cup \
0  2019-12-01 00:45:00      Superliga      636  False
1  2019-12-01 01:00:00  Primera Division      752  False
2  2019-12-01 01:00:00      Liga Nacional      734  False
3  2019-12-01 01:00:00        Liga MX      743  False
4  2019-12-01 01:00:00      Liga Nacional      705  False
```

	home_team_coach_id	away_team_coach_id	home_team_history_match_date_1 \
0	468196.0	468200.0	2019-11-26 00:10:00
1	516788.0	22169161.0	2019-11-27 21:00:00
2	2510608.0	456313.0	2019-11-28 01:15:00
3	1552508.0	465797.0	2019-11-28 01:00:00
4	429958.0	426870.0	2019-11-27 18:00:00

	home_team_history_match_date_2	home_team_history_match_date_3 \
0	2019-11-10 18:30:00	2019-11-03 18:30:00
1	2019-11-24 01:00:00	2019-11-10 01:00:00
2	2019-11-21 01:30:00	2019-11-03 21:30:00
3	2019-11-23 03:10:00	2019-11-10 03:00:00

4	2019-11-24 01:00:00	2019-11-10 02:00:00
---	---------------------	---------------------

	home_team_history_match_date_4	...	away_team_history_league_id_1	\
0	2019-10-29 22:10:00	...	1122.0	
1	2019-11-02 21:30:00	...	752.0	
2	2019-10-27 01:00:00	...	734.0	
3	2019-11-04 02:00:00	...	743.0	
4	2019-11-03 01:00:00	...	705.0	

	away_team_history_league_id_2	away_team_history_league_id_3	\
0	642.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	743.0	
4	705.0	705.0	

	away_team_history_league_id_4	away_team_history_league_id_5	\
0	636.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	743.0	
4	705.0	705.0	

	away_team_history_league_id_6	away_team_history_league_id_7	\
0	1122.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	743.0	
4	705.0	705.0	

	away_team_history_league_id_8	away_team_history_league_id_9	\
0	642.0	636.0	
1	752.0	752.0	
2	734.0	734.0	
3	743.0	746.0	
4	705.0	705.0	

	away_team_history_league_id_10
0	1122.0
1	752.0
2	734.0
3	743.0
4	705.0

[5 rows x 186 columns]

```
[11]: update_df_test_ = df_test.copy()
      update_df_test_.drop(['home_team_name', 'away_team_name', 'id'],
      ↪axis='columns', inplace=True)
      update_df_test_.head()
```

```
[11]:      match_date  league_name  league_id  is_cup  home_team_coach_id  \
0   2021-05-01 00:15:00  Division 1      755   False      9605704.0
1   2021-05-01 00:30:00      Liga MX      743   False      10319425.0
2   2021-05-01 01:00:00  Paulista A2     1314   False       440905.0
3   2021-05-01 01:00:00  Paulista A2     1314   False      21036670.0
4   2021-05-01 01:00:00  Paulista A2     1314   False      3618158.0

      away_team_coach_id  home_team_history_match_date_1  \
0      30866250.0      2021-04-28 00:30:00
1      460370.0      2021-04-24 02:06:00
2      220998.0      2021-04-28 23:00:00
3      1553238.0      2021-04-29 01:00:00
4      37400444.0      2021-04-29 01:00:00

      home_team_history_match_date_2  home_team_history_match_date_3  \
0      2021-04-23 23:00:00      2021-04-21 00:30:00
1      2021-04-17 00:30:00      2021-04-10 00:30:00
2      2021-04-24 23:00:00      2021-04-22 23:00:00
3      2021-04-24 23:00:00      2021-04-23 01:00:00
4      2021-04-24 23:00:00      2021-04-22 23:00:00

      home_team_history_match_date_4  ...  away_team_history_league_id_1  \
0      2021-04-17 00:15:00  ...      755.0
1      2021-04-04 01:00:00  ...      743.0
2      2021-04-21 01:00:00  ...     1314.0
3      2021-04-21 01:00:00  ...     1314.0
4      2021-04-20 23:00:00  ...     1314.0

      away_team_history_league_id_2  away_team_history_league_id_3  \
0      755.0      755.0
1      743.0      743.0
2     1314.0     1314.0
3     1314.0     1314.0
4     1314.0     1314.0

      away_team_history_league_id_4  away_team_history_league_id_5  \
0      755.0      755.0
1      743.0      743.0
2     1314.0     1314.0
3     1314.0     1314.0
4     1314.0     1314.0
```

	away_team_history_league_id_6	away_team_history_league_id_7	\
0	755.0	755.0	
1	743.0	743.0	
2	1314.0	1314.0	
3	1314.0	1314.0	
4	1314.0	1314.0	

	away_team_history_league_id_8	away_team_history_league_id_9	\
0	755.0	755.0	
1	743.0	743.0	
2	1314.0	1315.0	
3	1314.0	1746.0	
4	1314.0	1314.0	

	away_team_history_league_id_10
0	755.0
1	743.0
2	1315.0
3	1746.0
4	1314.0

[5 rows x 186 columns]

```
[12]: df_train_ts.head()
```

```
[12]:      id score target
0  11906497  2-3   away
1  11984383  1-0   home
2  11983301  2-2   draw
3  11983471  1-2   away
4  11883005  1-0   home
```

```
[14]: X = update_df_train_drop.copy()
      y = pd.DataFrame(update_df_train.score)
```

Handling Categorical values

```
[15]: import random
      from sklearn import preprocessing, metrics
      from sklearn.model_selection import train_test_split
      from sklearn.metrics import accuracy_score , precision_score ,
      ↪recall_score,confusion_matrix,classification_report
```

```
[16]: le = preprocessing.LabelEncoder()
      y_le= le.fit(y)
      y_trans = y_le.transform(y)
      y_trans
```

```
/Users/quangtn/opt/anaconda3/envs/mc_challenge1/lib/python3.10/site-  
packages/sklearn/preprocessing/_label.py:97: DataConversionWarning: A column-  
vector y was passed when a 1d array was expected. Please change the shape of y  
to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
/Users/quangtn/opt/anaconda3/envs/mc_challenge1/lib/python3.10/site-  
packages/sklearn/preprocessing/_label.py:132: DataConversionWarning: A column-  
vector y was passed when a 1d array was expected. Please change the shape of y  
to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, dtype=self.classes_.dtype, warn=True)
```

```
[16]: array([48, 15, 47, ..., 16, 9, 42])
```

```
[17]: schema_features= list(X.columns)  
for i in schema_features:  
    X_le = le.fit(X[i])  
    X[i] = X_le.transform(X[i])
```

```
[18]: X_test = update_df_test_.copy()
```

```
[19]: schema_features= list(X_test.columns)  
for i in schema_features:  
    X_le = le.fit(X_test[i])  
    X_test[i] = X_le.transform(X_test[i])
```

Splitting train.csv data into train/val data

```
[20]: X_train, X_val, y_train, y_val = train_test_split(X, y_trans, train_size=0.7,  
↳ random_state=42)  
X_train.shape, X_val.shape
```

```
[20]: ((77656, 186), (33282, 186))
```

Building Machine learning model

```
[22]: from sklearn.ensemble import RandomForestRegressor  
  
regressor = RandomForestRegressor(n_estimators=100,  
                                random_state=0)  
  
# fit the regressor with x and y data  
regressor.fit(X, y_trans)
```

```
[22]: RandomForestRegressor(random_state=0)
```

```
[23]: score = regressor.score(X_train, y_train)  
print("R-squared:", score)
```

```
R-squared: 0.8703375196514357
```

Prediction with test.csv data

```
[24]: y_pred = regressor.predict(X_test)
      y_pred
```

```
[24]: array([23.17, 18.33, 24.45, ..., 26.72, 25.91, 19.94])
```

```
[25]: y_score = regressor.apply(X_test)
```

```
[26]: y_score.shape
```

```
[26]: (72711, 100)
```

Because I have code a method manually to find a score of a match instead of direct inverse_tranfrom function from LabelEncoder.

```
[27]: y_train_ = pd.DataFrame(y_train)
      y_train_.head()
```

```
[27]:      0
0      0
1     64
2     87
3     19
4     15
```

```
[28]: y_pred_ = np.round(y_pred)
      y_pred_ = list(y_pred_)
      y_pred_
```

```
[28]: [23.0,
      18.0,
      24.0,
      19.0,
      22.0,
      21.0,
      25.0,
      34.0,
      32.0,
      28.0,
      30.0,
      36.0,
      23.0,
      36.0,
      33.0,
      26.0,
      27.0,
      24.0,
      32.0,
```

28.0,
17.0,
28.0,
22.0,
28.0,
29.0,
23.0,
26.0,
29.0,
27.0,
25.0,
35.0,
30.0,
25.0,
23.0,
21.0,
23.0,
18.0,
16.0,
22.0,
25.0,
29.0,
23.0,
14.0,
29.0,
21.0,
24.0,
29.0,
33.0,
35.0,
30.0,
25.0,
27.0,
24.0,
28.0,
29.0,
27.0,
29.0,
30.0,
28.0,
32.0,
26.0,
24.0,
32.0,
33.0,
36.0,
27.0,

16.0,
23.0,
23.0,
32.0,
28.0,
26.0,
36.0,
31.0,
31.0,
30.0,
28.0,
25.0,
26.0,
31.0,
30.0,
26.0,
27.0,
29.0,
25.0,
27.0,
22.0,
26.0,
24.0,
32.0,
25.0,
20.0,
28.0,
29.0,
31.0,
39.0,
28.0,
29.0,
26.0,
28.0,
30.0,
26.0,
23.0,
27.0,
29.0,
31.0,
28.0,
26.0,
32.0,
26.0,
22.0,
29.0,
20.0,

21.0,
35.0,
28.0,
23.0,
30.0,
30.0,
32.0,
22.0,
27.0,
27.0,
24.0,
27.0,
28.0,
28.0,
25.0,
31.0,
31.0,
23.0,
25.0,
31.0,
26.0,
22.0,
25.0,
29.0,
30.0,
30.0,
27.0,
30.0,
30.0,
25.0,
32.0,
21.0,
26.0,
29.0,
27.0,
29.0,
20.0,
22.0,
28.0,
23.0,
28.0,
29.0,
29.0,
29.0,
30.0,
28.0,
29.0,

28.0,
32.0,
25.0,
21.0,
30.0,
36.0,
22.0,
34.0,
27.0,
25.0,
28.0,
28.0,
28.0,
27.0,
29.0,
32.0,
28.0,
19.0,
24.0,
20.0,
26.0,
30.0,
18.0,
21.0,
26.0,
24.0,
31.0,
28.0,
24.0,
18.0,
34.0,
27.0,
37.0,
28.0,
27.0,
20.0,
20.0,
31.0,
25.0,
26.0,
16.0,
26.0,
29.0,
36.0,
28.0,
24.0,
30.0,

28.0,
25.0,
21.0,
25.0,
23.0,
31.0,
24.0,
33.0,
26.0,
29.0,
24.0,
30.0,
30.0,
26.0,
22.0,
27.0,
25.0,
20.0,
24.0,
29.0,
26.0,
20.0,
24.0,
30.0,
26.0,
27.0,
25.0,
19.0,
24.0,
36.0,
27.0,
24.0,
24.0,
22.0,
25.0,
26.0,
28.0,
19.0,
21.0,
29.0,
23.0,
27.0,
26.0,
24.0,
26.0,
27.0,
25.0,

27.0,
20.0,
26.0,
25.0,
25.0,
25.0,
31.0,
33.0,
35.0,
25.0,
25.0,
21.0,
18.0,
27.0,
27.0,
29.0,
32.0,
27.0,
31.0,
27.0,
30.0,
26.0,
32.0,
23.0,
29.0,
32.0,
25.0,
30.0,
23.0,
33.0,
21.0,
30.0,
23.0,
24.0,
30.0,
17.0,
21.0,
22.0,
30.0,
28.0,
29.0,
25.0,
33.0,
25.0,
25.0,
32.0,
22.0,

24.0,
26.0,
27.0,
26.0,
30.0,
29.0,
25.0,
31.0,
32.0,
29.0,
22.0,
30.0,
25.0,
27.0,
23.0,
34.0,
26.0,
35.0,
26.0,
28.0,
26.0,
26.0,
29.0,
30.0,
23.0,
25.0,
24.0,
26.0,
25.0,
26.0,
25.0,
25.0,
21.0,
25.0,
25.0,
26.0,
28.0,
23.0,
30.0,
26.0,
21.0,
28.0,
29.0,
27.0,
26.0,
27.0,
25.0,

25.0,
26.0,
29.0,
31.0,
30.0,
25.0,
27.0,
29.0,
23.0,
23.0,
20.0,
24.0,
29.0,
29.0,
27.0,
23.0,
27.0,
25.0,
27.0,
28.0,
20.0,
30.0,
29.0,
23.0,
22.0,
22.0,
22.0,
24.0,
25.0,
23.0,
25.0,
24.0,
27.0,
31.0,
22.0,
30.0,
35.0,
21.0,
22.0,
30.0,
21.0,
28.0,
27.0,
18.0,
20.0,
26.0,
25.0,

32.0,
28.0,
21.0,
32.0,
29.0,
31.0,
27.0,
22.0,
26.0,
27.0,
33.0,
20.0,
29.0,
23.0,
25.0,
24.0,
26.0,
28.0,
23.0,
26.0,
23.0,
30.0,
23.0,
29.0,
22.0,
22.0,
30.0,
27.0,
19.0,
26.0,
26.0,
24.0,
30.0,
26.0,
32.0,
24.0,
31.0,
29.0,
22.0,
25.0,
27.0,
28.0,
34.0,
24.0,
19.0,
26.0,
25.0,

41.0,
28.0,
25.0,
19.0,
37.0,
20.0,
24.0,
27.0,
23.0,
21.0,
26.0,
21.0,
33.0,
23.0,
26.0,
26.0,
28.0,
28.0,
29.0,
22.0,
31.0,
37.0,
28.0,
25.0,
35.0,
30.0,
33.0,
24.0,
34.0,
28.0,
27.0,
28.0,
22.0,
25.0,
23.0,
27.0,
22.0,
27.0,
25.0,
32.0,
20.0,
24.0,
25.0,
27.0,
34.0,
23.0,
23.0,

22.0,
20.0,
34.0,
25.0,
26.0,
30.0,
24.0,
23.0,
25.0,
23.0,
24.0,
25.0,
28.0,
22.0,
31.0,
31.0,
24.0,
25.0,
25.0,
21.0,
19.0,
22.0,
31.0,
20.0,
19.0,
24.0,
25.0,
34.0,
32.0,
27.0,
29.0,
24.0,
22.0,
31.0,
25.0,
24.0,
26.0,
29.0,
24.0,
24.0,
26.0,
27.0,
30.0,
20.0,
25.0,
27.0,
18.0,

19.0,
25.0,
32.0,
22.0,
26.0,
24.0,
31.0,
25.0,
27.0,
28.0,
33.0,
28.0,
29.0,
27.0,
26.0,
21.0,
18.0,
34.0,
30.0,
22.0,
26.0,
30.0,
25.0,
19.0,
27.0,
26.0,
19.0,
28.0,
23.0,
28.0,
21.0,
19.0,
18.0,
19.0,
25.0,
21.0,
23.0,
19.0,
25.0,
31.0,
23.0,
29.0,
29.0,
25.0,
25.0,
24.0,
23.0,

23.0,
24.0,
30.0,
22.0,
30.0,
30.0,
24.0,
25.0,
21.0,
26.0,
29.0,
34.0,
24.0,
26.0,
25.0,
24.0,
32.0,
27.0,
29.0,
28.0,
31.0,
22.0,
25.0,
28.0,
27.0,
12.0,
22.0,
24.0,
29.0,
25.0,
23.0,
32.0,
32.0,
33.0,
23.0,
25.0,
20.0,
25.0,
27.0,
27.0,
19.0,
28.0,
32.0,
36.0,
27.0,
30.0,
31.0,

23.0,
26.0,
27.0,
29.0,
28.0,
23.0,
20.0,
19.0,
24.0,
30.0,
19.0,
30.0,
25.0,
26.0,
30.0,
24.0,
21.0,
19.0,
21.0,
20.0,
23.0,
26.0,
21.0,
26.0,
27.0,
30.0,
38.0,
18.0,
24.0,
22.0,
24.0,
27.0,
24.0,
24.0,
27.0,
23.0,
29.0,
26.0,
32.0,
22.0,
33.0,
31.0,
21.0,
24.0,
32.0,
20.0,
30.0,

29.0,
21.0,
20.0,
26.0,
27.0,
31.0,
31.0,
20.0,
28.0,
34.0,
24.0,
29.0,
29.0,
24.0,
23.0,
29.0,
30.0,
34.0,
30.0,
25.0,
21.0,
21.0,
19.0,
33.0,
29.0,
30.0,
19.0,
25.0,
20.0,
20.0,
19.0,
30.0,
22.0,
24.0,
29.0,
30.0,
25.0,
20.0,
24.0,
22.0,
22.0,
32.0,
26.0,
32.0,
37.0,
23.0,
31.0,

23.0,
21.0,
26.0,
33.0,
19.0,
28.0,
31.0,
32.0,
32.0,
35.0,
36.0,
33.0,
29.0,
25.0,
29.0,
27.0,
26.0,
28.0,
26.0,
25.0,
23.0,
22.0,
28.0,
31.0,
25.0,
29.0,
30.0,
30.0,
24.0,
26.0,
27.0,
25.0,
24.0,
26.0,
28.0,
25.0,
33.0,
28.0,
27.0,
28.0,
25.0,
28.0,
26.0,
23.0,
23.0,
23.0,
28.0,

26.0,
19.0,
31.0,
23.0,
23.0,
33.0,
24.0,
23.0,
26.0,
19.0,
26.0,
24.0,
30.0,
24.0,
27.0,
21.0,
20.0,
25.0,
29.0,
27.0,
21.0,
27.0,
22.0,
19.0,
25.0,
23.0,
34.0,
25.0,
24.0,
23.0,
18.0,
25.0,
21.0,
26.0,
24.0,
21.0,
27.0,
20.0,
27.0,
25.0,
27.0,
20.0,
19.0,
31.0,
18.0,
27.0,
25.0,

36.0,
34.0,
24.0,
28.0,
24.0,
25.0,
27.0,
22.0,
25.0,
22.0,
24.0,
28.0,
19.0,
23.0,
27.0,
30.0,
25.0,
28.0,
20.0,
25.0,
31.0,
25.0,
34.0,
26.0,
26.0,
24.0,
33.0,
26.0,
29.0,
27.0,
27.0,
29.0,
25.0,
26.0,
28.0,
26.0,
22.0,
24.0,
27.0,
32.0,
28.0,
32.0,
28.0,
22.0,
22.0,
23.0,
22.0,

16.0,
24.0,
25.0,
20.0,
20.0,
21.0,
33.0,
28.0,
22.0,
31.0,
26.0,
23.0,
22.0,
28.0,
27.0,
21.0,
22.0,
27.0,
28.0,
25.0,
31.0,
21.0,
26.0,
34.0,
28.0,
27.0,
32.0,
31.0,
22.0,
20.0,
20.0,
23.0,
30.0,
28.0,
27.0,
27.0,
31.0,
27.0,
28.0,
30.0,
23.0,
29.0,
17.0,
18.0,
28.0,
21.0,
25.0,

28.0,
23.0,
23.0,
24.0,
22.0,
25.0,
34.0,
32.0,
20.0,
26.0,
23.0,
23.0,
32.0,
22.0,
22.0,
26.0,
29.0,
32.0,
25.0,
19.0,
21.0,
30.0,
28.0,
30.0,
23.0,
26.0,
23.0,
28.0,
23.0,
26.0,
23.0,
28.0,
23.0,
24.0,
25.0,
30.0,
22.0,
20.0,
25.0,
24.0,
26.0,
29.0,
19.0,
23.0,
26.0,
28.0,
28.0,

25.0,
29.0,
23.0,
24.0,
28.0,
29.0,
29.0,
34.0,
28.0,
25.0,
23.0,
29.0,
20.0,
29.0,
21.0,
22.0,
25.0,
25.0,
25.0,
23.0,
23.0,
20.0,
23.0,
19.0,
21.0,
22.0,
24.0,
21.0,
32.0,
39.0,
34.0,
31.0,
37.0,
23.0,
24.0,
32.0,
29.0,
21.0,
24.0,
19.0,
35.0,
...]

```
[29]: y_trans_df = pd.DataFrame(y_trans.astype(int), columns=["label_encoder"])  
y_merge = pd.concat([y, y_trans_df],axis=1)  
y_merge.head()
```

```
[29]:
```

	score	label_encoder
0	2-3	48
1	1-0	15
2	2-2	47
3	1-2	19
4	1-0	15

```
[30]: score_pred = []
for i in range(len(y_pred_)):
    score_pred.append(y_merge.
↳loc[(y_merge['label_encoder']==y_pred_[i]), 'score'].iloc[0])
```

```
[40]: score_csv = pd.concat([df_test.id, df_test.home_team_name, df_test.
↳away_team_name,
                                df_test.match_date, pd.DataFrame(score_pred,
↳columns=["score"])],
                                axis=1).to_csv("./Task3_RandomForestResgessor_submission.
↳csv")
```

```
[ ]:
```