ИУ5-65Б Погосян С.Л. РК1 Вар №16

Номер задачи - 2, номер набора данных - 8

Рубежный контроль №1 Задание Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

Набор данных:

https://www.kaggle.com/mathan/fifa-2018-match-statistics

Дополнительное требование:

Для студентов групп ИУ5-65Б — для набора данных построить «парные диаграммы»

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

0.0.1. Загружаем данные:

```
[2]: data = pd.read_csv('FIFA_2018.csv', sep=",")
[3]: #
     data.shape
[3]: (128, 27)
[4]: #
     data.dtypes
[4]: Date
                                  object
     Team
                                  object
     Opponent
                                  object
     Goal Scored
                                   int64
     Ball Possession %
                                   int64
                                   int64
     Attempts
     On-Target
                                   int64
                                   int64
     Off-Target
     Blocked
                                   int64
     Corners
                                   int64
     Offsides
                                   int64
     Free Kicks
                                   int64
     Saves
                                   int64
     Pass Accuracy %
                                   int64
     Passes
                                   int64
```

```
Distance Covered (Kms)
                                   int64
                                   int64
     Fouls Committed
     Yellow Card
                                   int64
     Yellow & Red
                                   int64
     Red
                                   int64
    Man of the Match
                                  object
     1st Goal
                                 float64
     Round
                                  object
     PS0
                                  object
     Goals in PSO
                                   int64
     Own goals
                                 float64
     Own goal Time
                                 float64
     dtype: object
[5]: #
     data.isnull().sum()
[5]: Date
                                   0
     Team
                                   0
     Opponent
                                   0
     Goal Scored
                                   0
     Ball Possession %
                                   0
     Attempts
                                   0
     On-Target
                                   0
     Off-Target
                                   0
     Blocked
                                   0
                                   0
     Corners
                                   0
     Offsides
     Free Kicks
                                   0
     Saves
                                   0
     Pass Accuracy %
                                   0
     Passes
                                   0
     Distance Covered (Kms)
                                   0
     Fouls Committed
                                   0
     Yellow Card
                                   0
     Yellow & Red
                                   0
     Red
                                   0
     Man of the Match
                                   0
     1st Goal
                                  34
     Round
                                   0
     PS0
                                   0
     Goals in PSO
                                   0
     Own goals
                                 116
     Own goal Time
                                 116
     dtype: int64
[6]: #
            10
```

data.head(10)

[6]:	Da →% \	ate	Tea	n (Opponent	Goal S	Scored	Ball Pos	ssessi	onu
	0 14-06-20	018	Russi	a Saud:	i Arabia		5			Ш
		018 8	Saudi Arabi	a	Russia		0			Ш
	→60 2 15-06-20 →43	018	Едур	t	Uruguay		0			Ш
	3 15-06-20 →57	018	Urugua	у	Egypt		1			Ш
	4 15-06-20 →64	018	Morocc	0	Iran		0			Ш
	5 15-06-20 →36	018	Ira	n	Morocco		1			Ш
	6 15-06-20 →39	018	Portuga	1	Spain		3			ш
	7 15-06-20 →61	018	Spai	n I	Portugal		3			Ш
	8 16-06-20	018	Franc	e Aı	ıstralia		2			Ш
	9 16-06-20	018	Australi	a	France		1			ш
	Attempts	s On-	-Target Of	f-Target	t Blocke	ed Corr	ners	Yellow	Card	\
	Attempts 0 13		-Target Of 7	_	t Blocke 3	ed Corr 3	ners 6	Yellow	Card 0	\
	0 13		_		3		_	Yellow	0	\
	0 13	3 6	7 0	3	3 3	3	6 2	Yellow	0 0	\
	0 13 1 6 2 8	3 6 8	7 0 3	3	3 3 3	3 3 2	6 2 0	Yellow	0 0 2	\
	0 13 1 6 2 8 3 14	3 6 8 4	7 0 3 4	3	3 3 3	3 3 2 4	6 2 0 5	Yellow	0 0 2 0	\
	0 13 1 6 2 8 3 14 4 13	3 6 8 4 3	7 0 3 4 3	3	3 3 3 5	3 3 2 4 4	6 2 0 5	Yellow	0 0 2 0 1	\
	0 13 1 6 2 8 3 14 4 13 5 8	3 6 8 4 3 8	7 0 3 4 3 2	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	3 3 3 5 5	3 2 4 4	6 2 0 5 2	Yellow	0 0 2 0 1 3	\
	0 13 1 2 3 3 14 4 13 5 6	3 6 8 4 3 8	7 0 3 4 3 2 3	;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	3 3 3 5 5 5 5	3 2 4 4 1 3	6 2 0 5 2 4	Yellow	0 0 2 0 1 3	
	0 13 1 6 2 8 3 14 4 13 5 8 6 8 7 12	3 6 8 4 3 8 8 8	7 0 3 4 3 2 3 5		3 3 3 5 5 5 2	3 2 4 4 1 3	6 2 0 5 2 4 5	Yellow	0 0 2 0 1 3 1	
	0 13 1 2 8 3 14 4 13 5 6 8 7 12 8 12	3 6 8 4 3 8 8 2 2	7 0 3 4 3 2 3 5		3 3 3 5 5 5 5 2 5	3 2 4 4 1 3 2	6 2 5 5 5 5 5	Yellow	0 0 2 0 1 3 1 1	
	0 13 1 2 8 3 14 4 13 5 6 8 7 12 8 12	3 6 8 4 3 8 8 8	7 0 3 4 3 2 3 5		3 3 3 5 5 5 2	3 2 4 4 1 3	6 2 0 5 2 4 5	Yellow	0 0 2 0 1 3 1	
	0 13 1 6 2 8 3 14 4 13 5 6 6 8 7 12 8 13 9	3 6 8 4 3 8 8 8 2 2 4	7 0 3 4 3 2 3 5 5		3 3 3 5 5 5 2 5 4 2	3 2 4 4 1 3 2 3 1	6 2 5 5 5 5 5		0 0 2 0 1 3 1 1 1 3	
	0 13 1 6 2 8 3 14 4 13 5 8 6 8 7 13 8 13 9 Yellow 8	3 6 8 4 3 8 8 2 2 4 & Red	7 0 3 4 3 2 3 5 5 1		3 3 3 5 5 5 2 5 4 2 Match 1s	3 2 4 4 1 3 2 3 1	6 2 5 5 5 1	Round I	0 0 2 0 1 3 1 1 1 3	
	0 13 1 6 2 8 3 14 4 13 5 8 6 8 7 12 8 13 9 Yellow 8	3 6 8 4 3 8 8 8 2 2 4 & Red 0	7 0 3 4 3 2 3 5 5 1 Red Man 0		3 3 3 5 5 5 2 5 4 2 Match 1s	3 2 4 4 1 3 2 3 1 st Goal 12.0	6 2 5 5 5 1 Group	Round I	0 0 2 0 1 3 1 1 1 3	
	0 13 1 6 2 8 3 14 4 13 5 6 8 7 12 8 13 9 Yellow 8	3 6 8 4 3 8 8 8 2 2 4 & Red 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0		3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN	6 2 5 5 5 1 Group Group	Round I Stage Stage	0 0 2 0 1 3 1 1 1 3 No No	
	0 13 1 6 2 8 3 14 4 13 5 6 8 7 13 8 13 9 Yellow 8	3 6 8 4 3 8 8 2 2 4 & Red 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0		3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal NaN NaN	6 2 5 5 5 1 Group Group Group	Round H Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 PSO \ No	
	0 13 1 6 2 8 3 14 4 13 5 8 6 8 7 12 8 12 9 Yellow 8 0 1 2 3	3 6 8 4 3 8 8 8 2 2 4 & Red 0 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0 0		3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN NaN 89.0	6 2 5 5 5 1 Group Group Group Group Group	Round I Stage Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 No No No	
	0 13 1 6 2 8 3 14 4 13 5 6 8 7 13 8 13 9 Yellow 8	3 6 8 4 3 8 8 2 2 4 & Red 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0		3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal NaN NaN	6 2 5 5 5 1 Group Group Group Group Group	Round H Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 PSO \ No	
	0 13 1 6 2 8 3 14 4 13 5 8 6 8 7 12 8 12 9 Yellow 8 0 1 2 3	3 6 8 4 3 8 8 8 2 2 4 & Red 0 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0 0		3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN NaN 89.0	6 2 5 5 5 6 6 6 7 7 7 6 7 6 7 7 7 7 7 7 7	Round I Stage Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 No No No	
	0 13 1 6 2 8 3 14 4 13 5 6 8 7 12 8 13 9 Yellow 8 0 1 2 3 4	3 6 8 4 4 3 8 8 2 2 4 4 & Red 0 0 0 0 0 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0 0		3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN NaN 89.0 NaN	6 2 5 5 5 1 Group Group Group Group Group Group Group Group	Round H Stage Stage Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 No No No	
	0 13 1 6 2 8 3 14 4 13 5 6 8 7 12 8 13 9 Yellow 8 0 1 2 3 4 5	3 6 8 4 4 3 8 8 8 2 2 4 4 8 Red 0 0 0 0 0 0 0 0 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0 0 0		3	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN NaN 89.0 NaN 90.0	6 2 5 5 5 1 Group Group Group Group Group Group Group Group Group	Round In Stage Stage Stage Stage Stage Stage Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 No No No No	
	0 13 1 6 2 8 3 14 4 13 5 6 7 13 8 13 9 Yellow 8 0 1 2 3 4 5 6 7	3 6 8 4 4 3 8 8 2 2 4 4 8 Red 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0 0 0		3 3 3 3 5 5 5 5 2 5 4 2 4 2 4 2 4 4 2 7 8 8 No No Yes No Yes No Yes No Yes No Yes	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN 89.0 NaN 90.0 4.0 24.0	6 2 5 5 5 7 6 6 6 7 7 7 7 7 7 8 9	Round In Stage	0 0 2 0 1 3 1 1 1 3 PSO \ No No No No No No	
	0 13 1 6 2 8 3 14 4 13 5 6 8 7 13 8 13 9 Yellow 8 0 1 2 3 4 5 6	3 6 8 4 4 3 8 8 2 2 4 4 & Red 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	7 0 3 4 3 2 3 5 5 1 Red Man 0 0 0 0 0 0		3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 2 Match 1s 4 Yes No No Yes No Yes No Yes No Yes	3 3 2 4 4 4 1 3 2 3 1 1 St Goal 12.0 NaN NaN 89.0 NaN 90.0 4.0	6 2 0 5 5 1 Group	Round I Stage Stage Stage Stage Stage Stage Stage Stage	0 0 2 0 1 3 1 1 1 3 No No No No No	

```
Goals in PSO
                   Own goals
                               Own goal Time
0
                          NaN
                                           NaN
                0
                          NaN
1
                                           NaN
2
                0
                          NaN
                                           NaN
3
                0
                          NaN
                                           NaN
4
                0
                          1.0
                                          90.0
5
                0
                          NaN
                                           NaN
6
                0
                          NaN
                                           NaN
7
                0
                          NaN
                                           NaN
8
                0
                          NaN
                                           NaN
9
                0
                          1.0
                                          81.0
```

[10 rows x 27 columns]

```
[7]: total_count = data.shape[0]
print(' : {}'.format(total_count))
```

: 128

0.0.2. Обработка пропуков в числовых данных

```
1st Goal. float64. 34, 26.56%.

Own goals. float64. 116, 90.62%.

Own goal Time. float64. 116, 90.62%.
```

В колонках Own goals и Own goals содержится информация о наличии забитых голов в свои ворота и времени, когда это было сделано. Будем считать, что отсутствие информации говорит о том, что не было мячей забитых в свои ворота. Поэтому заполним пропуски в этих колонках нулями.

```
[9]: # "Own goals"
data_new_3 = data[['Own goals']].fillna(0)
```

```
[10]: data[['Own goals']]=data_new_3
[11]: #
                                  "Own goal Time"
      data_new_4 = data[['Own goal Time']].fillna(0)
      data[['Own goal Time']]=data_new_4
[13]: data.head()
[13]:
               Date
                              Team
                                         Opponent Goal Scored Ball Possession
       →%
      0 14-06-2018
                            Russia Saudi Arabia
                                                                                 Ш
       →40
      1 14-06-2018 Saudi Arabia
                                           Russia
                                                                                 Ш
       →60
      2 15-06-2018
                             Egypt
                                          Uruguay
                                                                                 Ш
       ⊸43
      3 15-06-2018
                           Uruguay
                                            Egypt
                                                                                 ш
       <u></u>57
      4 15-06-2018
                           Morocco
                                             Iran
                                                                                 Ш
       ⊸64
                               Off-Target Blocked Corners
                    On-Target
                                                                   Yellow Card
         Attempts
      0
               13
                                                   3
      1
                6
                            0
                                         3
                                                   3
                                                            2
                                                                              0
                                                               •••
      2
                8
                            3
                                         3
                                                   2
                                                            0
                                                                              2
      3
               14
                            4
                                         6
                                                   4
                                                            5
                                                                              0
      4
               13
                            3
                                         6
                                                   4
                                                            5
                                                                              1
         Yellow & Red Red Man of the Match 1st Goal
                                                                  Round PSO
                                                     12.0
                                                           Group Stage
      0
                     0
                          0
                                           Yes
                                                                          No
                                                           Group Stage
      1
                     0
                          0
                                            No
                                                      NaN
                                                                          No
      2
                     0
                          0
                                            No
                                                      {\tt NaN}
                                                           Group Stage
                                                                          No
      3
                     0
                          0
                                           Yes
                                                     89.0
                                                           Group Stage
                                                                          No
      4
                     0
                          0
                                                           Group Stage
                                            No
                                                      {\tt NaN}
                                                                          No
         Goals in PSO
                        Own goals Own goal Time
                              0.0
                     0
                              0.0
      1
                     0
                                              0.0
      2
                     0
                              0.0
                                              0.0
      3
                     0
                              0.0
                                              0.0
      4
                              1.0
                                             90.0
                     0
```

[5 rows x 27 columns]

В колонке 1st goal содержится информация о времени от начала игры, в которое был забит первый гол. Поэтому заполним пропуски нулями в тех строках, где значение

```
"Goal Scored" ранво нулю.
[14]: data.loc[(data['Goal Scored'] == 0), '1st Goal'] = 0
     Заполним оставшиеся пропуски в столбце "1st Goal" средним значением по столбцу.
[15]: data[['1st Goal']].describe()
[15]:
               1st Goal
      count
             127.000000
      mean
              29.204724
              27.289552
      std
      min
               0.000000
      25%
               0.000000
      50%
              24.000000
      75%
              51.000000
      max
              90.000000
[16]: res = np.where(np.isnan(data['1st Goal']), np.ma.array(data['1st Goal'],
                     mask = np.isnan(data['1st Goal'])).mean(axis = 0),__
       →data['1st Goal'])
[17]: data['1st Goal']=res
[18]: data.head()
[18]:
                                        Opponent Goal Scored Ball Possession
               Date
                              Team
       ∽%
      0 14-06-2018
                            Russia Saudi Arabia
                                                             5
                                                                                Ш
       →40
      1 14-06-2018 Saudi Arabia
                                          Russia
                                                                                Ш
       <del>-</del>60
      2 15-06-2018
                             Egypt
                                         Uruguay
                                                                               ш
       →43
      3 15-06-2018
                           Uruguay
                                           Egypt
                                                              1
                                                                               Ш
       →57
      4 15-06-2018
                           Morocco
                                             Iran
                                                                                Ш
                               Off-Target Blocked Corners ...
                                                                  Yellow Card
         Attempts
                   On-Target
      0
               13
      1
                            0
                                        3
                                                  3
                                                            2
                                                                            0
                6
                                                              ...
      2
                8
                            3
                                        3
                                                  2
                                                            0
                                                                            2
      3
               14
                            4
                                        6
                                                  4
                                                           5
                                                                            0
      4
               13
                            3
                                        6
                                                  4
                                                                            1
```

Round PSO \

Yellow & Red Red Man of the Match 1st Goal

0	0	0	Yes	12.0	Group Stage	No
1	0	0	No	0.0	Group Stage	No
2	0	0	No	0.0	Group Stage	No
3	0	0	Yes	89.0	Group Stage	No
4	0	0	No	0.0	Group Stage	No
	Goals in PSO	Own goals	Own goal Time			

	Goals	in PSO	Own goals	Own goal	Time
0		0	0.0		0.0
1		0	0.0		0.0
2		0	0.0		0.0
3		0	0.0		0.0
4		0	1.0		90.0

[5 rows x 27 columns]

[19]: # data.isnull().sum()

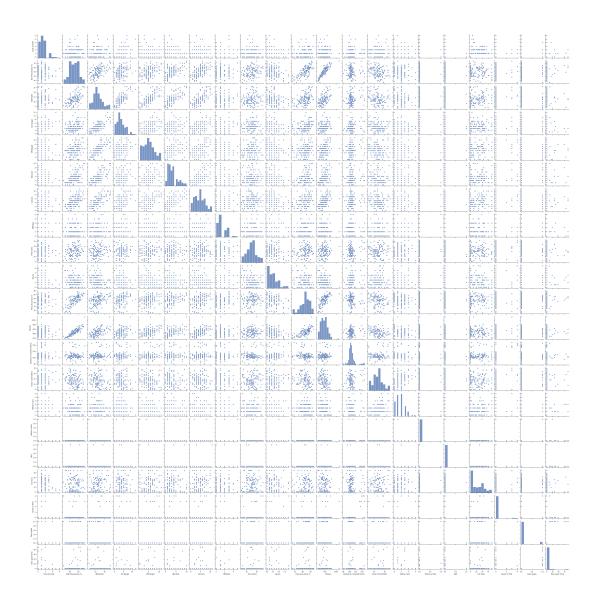
[19]: Date 0 ${\tt Team}$ 0 Opponent 0 Goal Scored 0 Ball Possession % 0 Attempts 0 On-Target 0 Off-Target 0 Blocked 0 Corners 0 Offsides 0 Free Kicks 0 Saves 0 Pass Accuracy % 0 Passes 0 Distance Covered (Kms) 0 Fouls Committed 0 Yellow Card 0 Yellow & Red 0 Red 0 Man of the Match 0 1st Goal 0 Round 0 PS0 0 Goals in PSO 0 Own goals 0 Own goal Time 0

dtype: int64

1. Парные диаграммы

[20]: sns.pairplot(data)

[20]: <seaborn.axisgrid.PairGrid at 0x7f658236d0a0>



[]: