## Якубов Артём ИУ5-63Б

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

**Team Opponent** 

Russia

Saudi

Arabia

Egypt

Uruguay

Morocco

Saudi

Arabia

Russia

Uruguay

Egypt

Iran

data = pd.read csv('FIFA 2018 Statistics.csv', sep=",")

Goal

5

0

0

1

0

object

object

object

int64

int64

int64

int64

int64

int64

int64

int64

int64 int64

int64

int64

int64 int64

int64

int64 int64

object float64

object

object

float64

float64

0

0

0

0

0

0

0

0 0

0

116

116

sc1 data = sc1.fit transform(data[['Goal Scored']])

from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer

Преобразование категориальных признаков в

количественные двумя способами

Метод не расширяет признаковое пространство набора данных

from sklearn.preprocessing import LabelEncoder, OneHotEncoder

Out[12]: array(['Russia', 'Saudi Arabia', 'Egypt', 'Uruguay', 'Morocco', 'Iran',

temp\_data[temp\_data['Team'].isnull()].shape

cat enc le = le.fit transform(temp data['Team'])

Out[16]: array(['Argentina', 'Australia', 'Belgium'], dtype=object)

cat\_enc\_ohe = ohe.fit\_transform(temp\_data)

Out[20]: <128x32 sparse matrix of type '<class 'numpy.float64'>'

le.inverse\_transform([0, 1, 2])

'Germany', 'Mexico', 'Brazil', 'Switzerland', 'Sweden', 'Korea Republic', 'Belgium', 'Panama', 'Tunisia', 'England', 'Colombia', 'Japan', 'Poland', 'Senegal'], dtype=object)

'Portugal', 'Spain', 'France', 'Australia', 'Argentina', 'Iceland', 'Peru', 'Denmark', 'Croatia', 'Nigeria', 'Costa Rica', 'Serbia',

'Russia', 'Saudi Arabia', 'Senegal', 'Serbia', 'Spain', 'Sweden', 'Switzerland', 'Tunisia', 'Uruguay'], dtype=object)

Использование данного метода не задает отношение порядка между значениями признака

0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0.], [0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.] 

print('Bcero ctpok: {}'.format(total\_count))

Масштабирование данных

int64

Scored

Ball

%

40

60

43

57

64

**Attempts** 

13

6

8

14

13

Possession

On-

Target Target

7

0

3

4

3

Off-

3

6

Blocked Corners ...

6

0 ...

5

5

3

3

2

4

4

Yellow

Card

0

0

2

0

1

Yel

&

import seaborn as sns

%matplotlib inline sns.set(style="ticks")

data.head()

**Date** 

14-

06-

2018

14-

06-

2018

15-

06-

2018

15-

06-

15-

06-

2018

5 rows × 27 columns

data.dtypes

Date

Team

Opponent

Attempts

Blocked

Corners

Saves

Passes

Offsides

Free Kicks

Pass Accuracy %

Fouls Committed Yellow Card

Man of the Match

Yellow & Red

Goals in PSO

Own goal Time

dtype: object

In [5]: data.isnull().sum()

1st Goal Round

Own goals

PSO

Team

Opponent

Attempts On-Target Off-Target

Blocked

Corners Offsides

Passes

1st Goal Round

Goals in PSO

Own goal Time

In [6]: total\_count = data.shape[0]

sc1 = MinMaxScaler()

plt.show()

40

30

20

10

0

2

3

4

Out[13]: (0, 1)

In [14]:

Label encoding

temp\_data.head()

Team

Russia

Egypt

Uruguay

Morocco

Saudi Arabia

temp\_data = data[['Team']]

temp data['Team'].unique()

np.unique(temp data)

One-hot encoding

ohe = OneHotEncoder()

temp\_data.shape

cat\_enc\_ohe.shape

cat\_enc\_ohe.todense()[0:10]

temp\_data.head(10)

**Team** 

Russia

Egypt

Uruguay

Morocco

Portugal

Spain

France

Australia

Ящик с усами

5

8

6

2.

Saves 4

In [24]:

10

sns.boxplot(y=data['Saves'])

<AxesSubplot:ylabel='Saves'>

15 Attempts 20

25

Out[23]: <AxesSubplot:xlabel='Attempts'>

Отображает одномерное распределение вероятности

sns.boxplot(x=data['Attempts'])

Iran

Saudi Arabia

0

2

4

5

6

8

(128, 1)

cat\_enc\_ohe

Out[19]: (128, 32)

Out[18]:

In [15]: le = LabelEncoder()

In [9]: plt.hist(data['Goal Scored'], 50)

Всего строк: 128

dtype: int64

Own goals

PSO

Free Kicks Saves

Pass Accuracy %

Fouls Committed Yellow Card Yellow & Red

Man of the Match

Distance Covered (Kms)

Goal Scored Ball Possession %

Out[5]: Date

Distance Covered (Kms)

On-Target

Off-Target

Goal Scored

Ball Possession %

2018

0

1

3

In [4]:

Out[4]:

hot encoding) для одного признака. Какие методы Вы использовали для решения задачи и почему?

Для заданного набора данных произведите масштабирование данных (для одного признака) и преобразование категориальных признаков в количественные двумя способами (label encoding, one

6 Вариант, задача №3