



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

---

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

## Отчёт по рубежному контролю №1

По дисциплине:  
«Технологии машинного обучения»

Выполнил:

Студент группы ИУ5Ц-83Б **Донченко М.А.**

Проверил:

\_\_\_\_\_

(Подпись, дата)

**Гапанюк Ю. Е.**

(Фамилия И.О.)

Москва, 2021

## Задание

Для заданного набора данных постройте основные графики, входящие в этап разведочного анализа данных. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Какие графики Вы построили и почему?

Какие выводы о наборе данных Вы можете сделать на основании построенных графиков?**Дополнительные требования по группам:**

Для студентов групп ИУ5-63Б, ИУ5Ц-83Б - для произвольной колонки данных построить график "Ящик с усами (boxplot)".

Набор данных:

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_boston.html#sklearn.datasets.load\\_boston](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html#sklearn.datasets.load_boston)

# РК ИУ5Ц-83Б

## Импорт библиотек

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
import warnings
from sklearn import datasets
from sklearn.datasets import load_iris
from sklearn import linear_model
from sklearn.cluster import KMeans
from sklearn import metrics
from pandas import DataFrame
%pylab inline
```

Populating the interactive namespace from numpy and matplotlib

```
In [2]: boston = load_iris()
data = pd.DataFrame(boston.data, columns=boston.feature_names)
data['TARGET'] = boston.target
```

```
In [3]: # Первые пять строк датасета
data.head()
```

```
Out[3]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	TARGET
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

```
In [4]: # Описание датасета
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   sepal length (cm)      150 non-null    float64
1   sepal width (cm)       150 non-null    float64
2   petal length (cm)      150 non-null    float64
3   petal width (cm)       150 non-null    float64
4   TARGET                 150 non-null    int64
dtypes: float64(4), int64(1)
memory usage: 6.0 KB
```

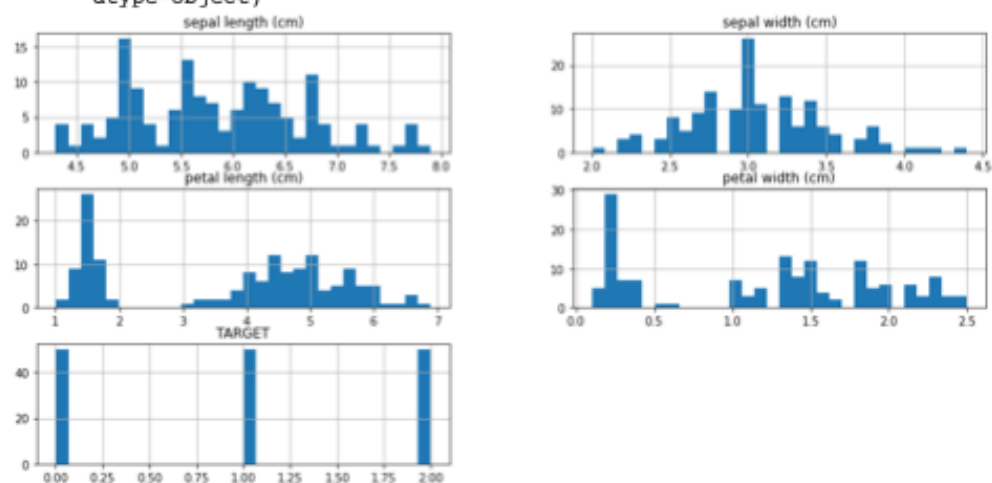
```
In [5]: # Статистические данные
data.describe()
```

```
Out[5]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	TARGET
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333	1.000000
std	0.828066	0.435866	1.765298	0.762238	0.819232
min	4.300000	2.000000	1.000000	0.100000	0.000000
25%	5.100000	2.800000	1.600000	0.300000	0.000000
50%	5.800000	3.000000	4.350000	1.300000	1.000000
75%	6.400000	3.300000	5.100000	1.800000	2.000000
max	7.900000	4.400000	6.900000	2.500000	2.000000

```
In [6]: # Гистограммы для всех признаков
data.hist(bins=30, figsize = (15,7))
```

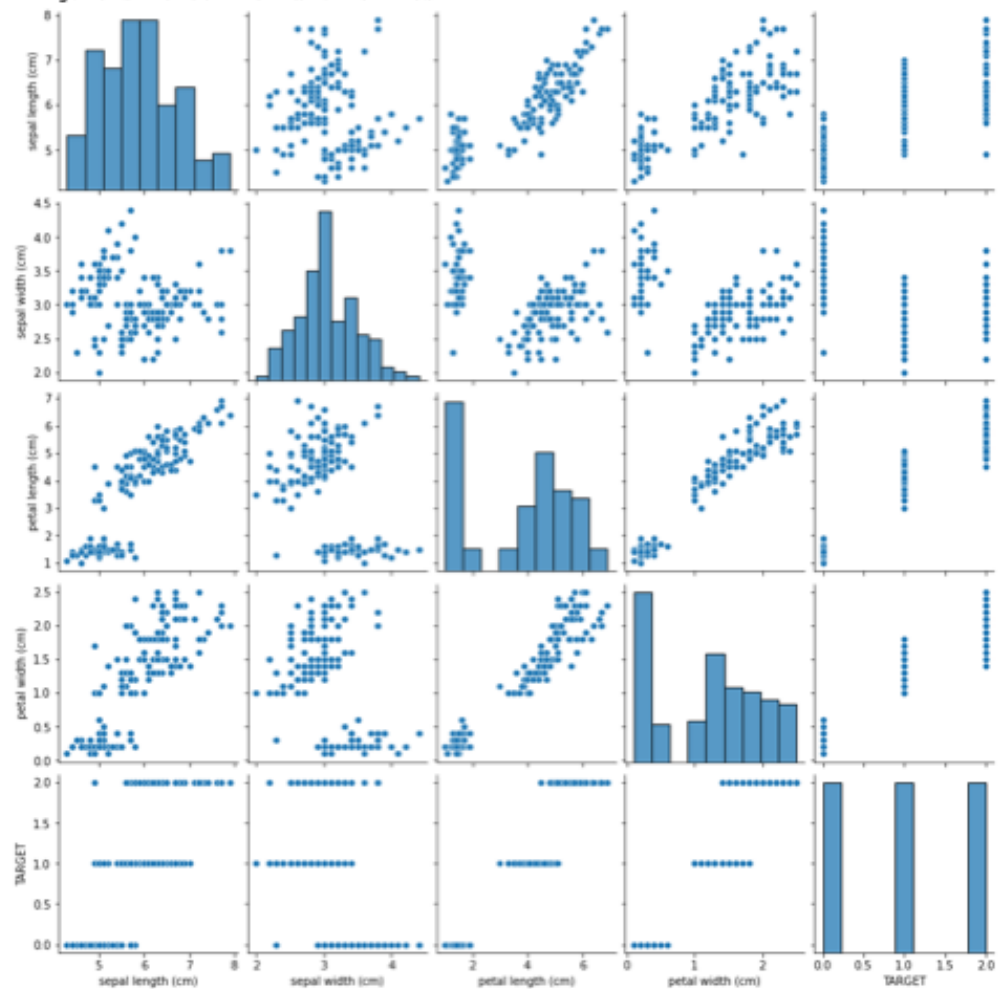
```
Out[6]: array([[<AxesSubplot:title={'center':'sepal length (cm)'}>,
<AxesSubplot:title={'center':'sepal width (cm)'}>],
[<AxesSubplot:title={'center':'petal length (cm)'}>,
<AxesSubplot:title={'center':'petal width (cm)'}>],
[<AxesSubplot:title={'center':'TARGET'}>, <AxesSubplot:>]],
dtype=object)
```



```
In [7]: # Диаграммы рассеяния для всех признаков
plt.figure(figsize=(12,6))
sns.pairplot(data)
```

Out[7]: <seaborn.axisgrid.PairGrid at 0x7ffcbe26bb50>

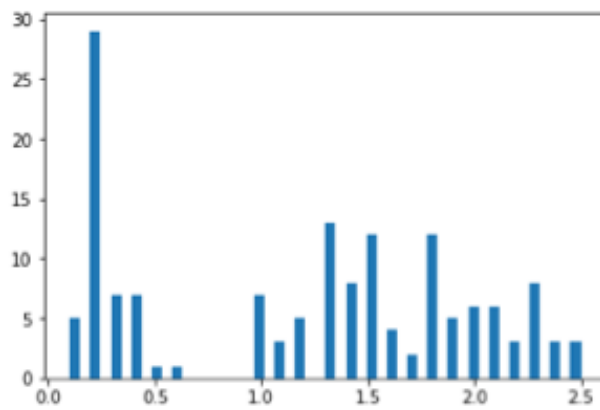
<Figure size 864x432 with 0 Axes>



```
In [8]: from sklearn.preprocessing import StandardScaler, MinMaxScaler, StandardSc
```

```
In [9]: scl = MinMaxScaler()  
scl_data = scl.fit_transform(data[['petal width (cm)']])
```

```
In [10]: plt.hist(data['petal width (cm)'], 50)  
plt.show()
```



```
In [11]: # Заполняем отсутствующие значения
data['TARGET'] = data['TARGET'].replace(0,np.nan)
data['TARGET'] = data['TARGET'].fillna(data['TARGET'].mean())
```

```
In [12]: data.head()
```

```
Out[12]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	TARGET
0	5.1	3.5	1.4	0.2	1.5
1	4.9	3.0	1.4	0.2	1.5
2	4.7	3.2	1.3	0.2	1.5
3	4.6	3.1	1.5	0.2	1.5
4	5.0	3.6	1.4	0.2	1.5

## Ящик с усами

```
In [13]: sns.boxplot(data['sepal width (cm)'])
```

```
/Users/kirsan/opt/anaconda3/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From m version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.  
warnings.warn(
```

```
Out[13]: <AxesSubplot:xlabel='sepal width (cm)'\>
```

