Московский государственный технический
университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Разработка интернет-приложений»

Отчет по рубежному контролю №1

Выполнил:
Студент группы ИУ5-63Б
Балабас Анна
Руководители: Гапанюк Ю.Е.

Дата: 18.04.22

Москва, 2022 г.

Задача №1.

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

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

Текст программы

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

```
[20] from sklearn import datasets
     data = datasets.load_iris()
```

```
[21]
     # Считайте DataFrame, используя данные функции
     df = pd.DataFrame(data.data, columns=data.feature_names)
     # Добавьте столбец "target" и заполните его данными.
     df['target'] = data.target
     # Посмотрим первые пять строк
     df.head()
```

	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

```
[24] df.columns
```

```
Index(['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',
       'petal width (cm)', 'target'],
      dtype='object')
```

```
[25] # Список колонок с типами данных
     df.dtypes
```

```
sepal length (cm)    float64
sepal width (cm)     float64
petal length (cm)    float64
petal width (cm)     float64
target                 int64
dtype: object
```

```
[26] # Проверим наличие пустых значений
     # Цикл по колонкам датасета
     for col in df.columns:
         # Количество пустых значений - все значения заполнены
         temp_null_count = df[df[col].isnull()].shape[0]
         print('{} - {}'.format(col, temp_null_count))
```

```
sepal length (cm) - 0
sepal width (cm) - 0
petal length (cm) - 0
petal width (cm) - 0
target - 0
```

```
[27] # Основные статистические характеристки набора данных
     df.describe()
```

	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

```
[28] df.corr()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941	0.782561
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126	-0.426658
petal length (cm)	0.871754	-0.428440	1.000000	0.962865	0.949035
petal width (cm)	0.817941	-0.366126	0.962865	1.000000	0.956547
target	0.782561	-0.426658	0.949035	0.956547	1.000000

```
[29] df.corr(method='spearman')
```

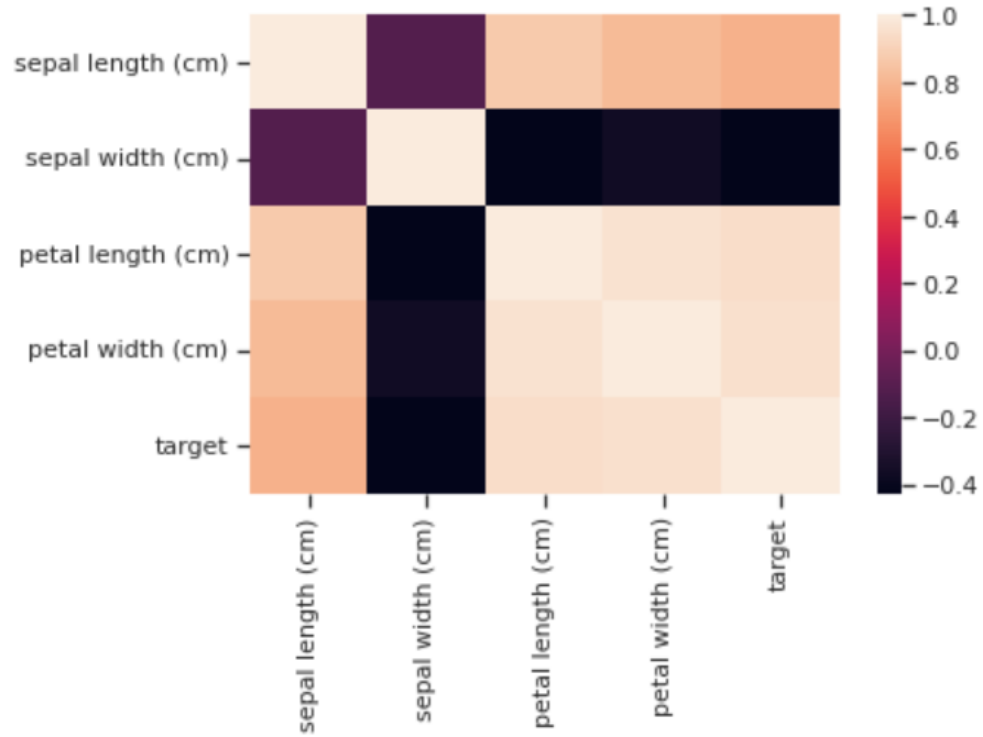
	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
sepal length (cm)	1.000000	-0.166778	0.881898	0.834289	0.798078
sepal width (cm)	-0.166778	1.000000	-0.309635	-0.289032	-0.440290
petal length (cm)	0.881898	-0.309635	1.000000	0.937667	0.935431
petal width (cm)	0.834289	-0.289032	0.937667	1.000000	0.938179
target	0.798078	-0.440290	0.935431	0.938179	1.000000

```
[39] df.corr(method='kendall')
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
sepal length (cm)	1.000000	-0.076997	0.718516	0.655309	0.670444
sepal width (cm)	-0.076997	1.000000	-0.185994	-0.157126	-0.337614
petal length (cm)	0.718516	-0.185994	1.000000	0.806891	0.822911
petal width (cm)	0.655309	-0.157126	0.806891	1.000000	0.839687
target	0.670444	-0.337614	0.822911	0.839687	1.000000

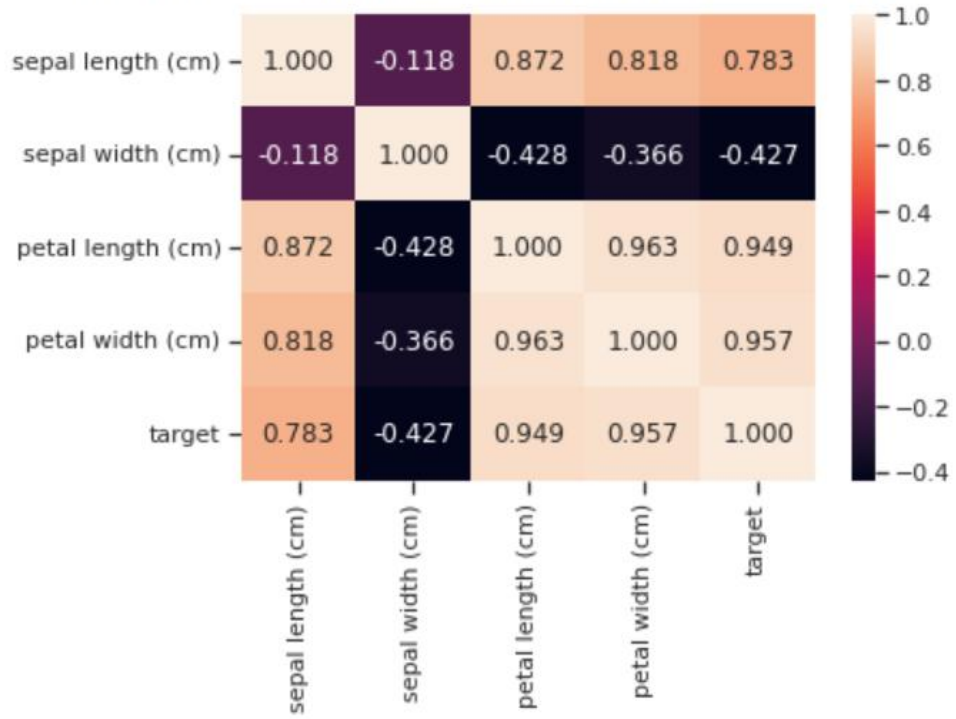
```
[30] sns.heatmap(df.corr())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f4f4f99a750>
```



```
[31] sns.heatmap(df.corr(), annot=True, fmt='.3f')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f4f502c6c50>
```




```
[32] # Треугольный вариант матрицы
     mask = np.zeros_like(df.corr(), dtype=np.bool)
     # чтобы оставить нижнюю часть матрицы
     # mask[np.triu_indices_from(mask)] = True
     # чтобы оставить верхнюю часть матрицы
     mask[np.tril_indices_from(mask)] = True
     sns.heatmap(df.corr(), mask=mask, annot=True, fmt='.3f')
```

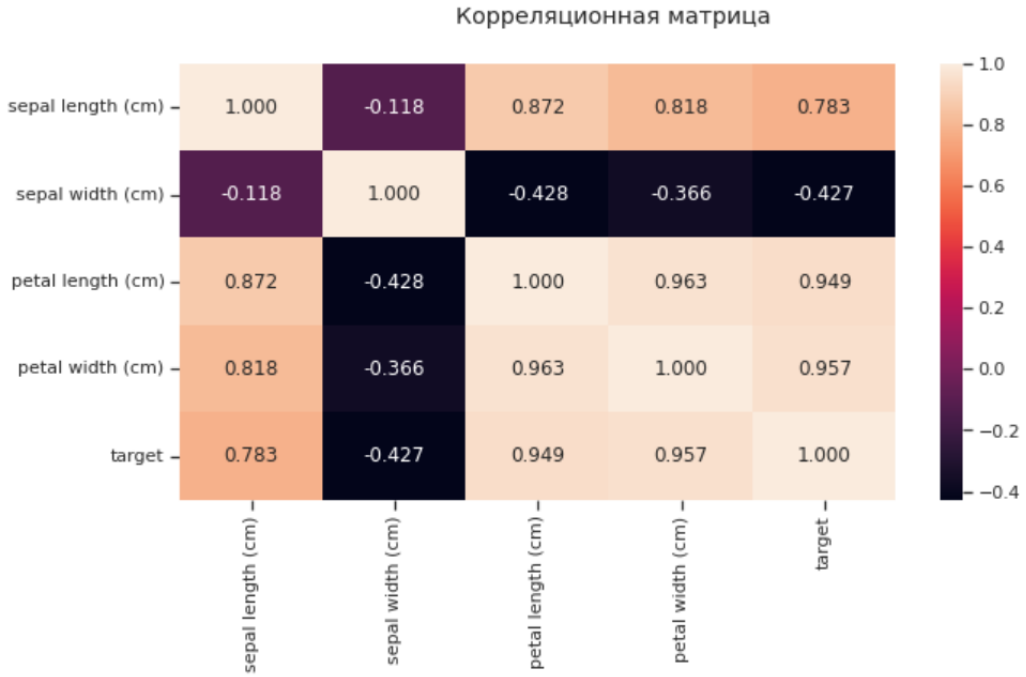
```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: DeprecationWarning: `np.bool` is a deprecated alias for the builtin `bool`.
Deprecated in NumPy 1.20; for more details and guidance: https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations

<matplotlib.axes._subplots.AxesSubplot at 0x7f4f5005cd10>
```

```
[33] fig, ax = plt.subplots(1, 3, sharex='col', sharey='row', figsize=(15,5))
     sns.heatmap(df.corr(method='pearson'), ax=ax[0], annot=True, fmt='.2f')
     sns.heatmap(df.corr(method='kendall'), ax=ax[1], annot=True, fmt='.2f')
     sns.heatmap(df.corr(method='spearman'), ax=ax[2], annot=True, fmt='.2f')
     fig.suptitle('Корреляционные матрицы, построенные различными методами')
     ax[0].title.set_text('Pearson')
     ax[1].title.set_text('Kendall')
     ax[2].title.set_text('Spearman')
```


```
[34] fig, ax = plt.subplots(1, 1, sharex='col', sharey='row', figsize=(10,5))
     fig.suptitle('Корреляционная матрица')
     sns.heatmap(df.corr(), ax=ax, annot=True, fmt='.3f')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f4f5043c290>
```



Ящик с усами

```
[38] sns.boxplot(x=df['sepal length (cm)'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f4f4ffdb1d0>
```