



**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 ИУ5 «Системы обработки информации и управления»**

**«Технологии разведочного анализа и обработки данных»
по курсу «Технологии машинного обучения»
Лабораторная работа №1
Вариант 5**

Выполнил:
студент группы ИУ5 – 62Б
Гринин О.Е.
подпись, дата

Проверил:
преподаватель кафедры ИУ5
Гапанюк Ю.Е.
подпись, дата

2020 г.

```
import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

sns.set(style="ticks")
```

In [3]:

```
data = pd.read_csv('data/heart.csv', sep=",")
```

In [4]:

```
data.head()
```

Out[4]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

In [5]:

```
data.shape
```

Out[5]:

```
(303, 14)
```

In [14]:

```
data.columns
```

Out[14]:

```
Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
       'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
      dtype='object')
```

In [16]:

```
data.corr()
```

Out[16]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1.000000	-0.098447	-0.068653	0.279351	0.213678	0.121308	-0.116211	-0.398522	0.096801	0.210013	-0.168814	0.276326	0.068001	-0.225439
sex	-0.098447	1.000000	-0.049353	-0.056769	-0.197912	0.045032	-0.058196	-0.044020	0.141664	0.096093	-0.030711	0.118261	0.210041	-0.280937
cp	-0.068653	-0.049353	1.000000	0.047608	-0.076904	0.094444	0.044421	0.295762	-0.394280	-0.149230	0.119717	-0.181053	-0.161736	0.433798
trestbps	0.279351	-0.056769	0.047608	1.000000	0.123174	0.177531	-0.114103	-0.046698	0.067616	0.193216	-0.121475	0.101389	0.062210	-0.144931
chol	0.213678	-0.197912	-0.076904	0.123174	1.000000	0.013294	-0.151040	-0.009940	0.067023	0.053952	-0.004038	0.070511	0.098803	-0.085239
fbs	0.121308	0.045032	0.094444	0.177531	0.013294	1.000000	-0.084189	-0.008567	0.025665	0.005747	-0.059894	0.137979	-0.032019	-0.028046
restecg	-0.116211	-0.058196	0.044421	-0.114103	-0.151040	-0.084189	1.000000	0.044123	-0.070733	-0.058770	0.093045	-0.072042	-0.011981	0.137230
thalach	-0.398522	-0.044020	0.295762	-0.046698	-0.009940	-0.008567	0.044123	1.000000	-0.378812	-0.344187	0.386784	-0.213177	-0.096439	0.421741
exang	0.096801	0.141664	-0.394280	0.067616	0.067023	0.025665	-0.070733	-0.378812	1.000000	0.288223	-0.257748	0.115739	0.206754	-0.436757
oldpeak	0.210013	0.096093	-0.149230	0.193216	0.053952	0.005747	-0.058770	-0.344187	0.288223	1.000000	-0.577537	0.222682	0.210244	-0.430696
slope	-0.168814	-0.030711	0.119717	-0.121475	-0.004038	-0.059894	0.093045	0.386784	-0.257748	-0.577537	1.000000	-0.080155	-0.104764	0.345877
ca	0.276326	0.118261	-0.181053	0.101389	0.070511	0.137979	-0.072042	-0.213177	0.115739	0.222682	-0.080155	1.000000	0.151832	-0.391724
thal	0.068001	0.210041	-0.161736	0.062210	0.098803	0.032019	-0.011981	-0.096439	0.206754	0.210244	-0.104764	0.151832	1.000000	-0.344029

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
target	0.225439	0.280937	0.433798	0.144931	0.085239	0.028046	0.137230	0.421741	0.436757	0.430696	0.345877	0.391724	0.344029	1.000000

In [9]:

```
data.isnull().sum()
```

Out[9]:

```
age      0
sex      0
cp       0
trestbps 0
chol     0
fbs      0
restecg  0
thalach  0
exang    0
oldpeak  0
slope    0
ca       0
thal     0
target   0
dtype: int64
```

Проверили пустые значения

In [15]:

```
data.corr()
```

Out[15]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1.000000	-0.098447	-0.068653	0.279351	0.213678	0.121308	-0.116211	0.398522	0.096801	0.210013	-0.168814	0.276326	0.068001	-0.225439
sex	-0.098447	1.000000	0.049353	0.056769	0.197912	0.045032	0.058196	0.044020	0.141664	0.096093	-0.030711	0.118261	0.210041	-0.280937

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
cp	-0.068653	-0.049353	1.000000	0.047608	-0.076904	0.094444	0.044421	0.295762	-0.394280	-0.149230	0.119717	-0.181053	-0.161736	0.433798
trestbps	0.279351	-0.056769	0.047608	1.000000	0.123174	0.177531	-0.114103	-0.046698	0.067616	0.193216	-0.121475	0.101389	0.062210	-0.144931
chol	0.213678	-0.197912	-0.076904	0.123174	1.000000	0.013294	-0.151040	-0.009940	0.067023	0.053952	-0.004038	0.070511	0.098803	-0.085239
fbs	0.121308	0.045032	0.094444	0.177531	0.013294	1.000000	-0.084189	-0.008567	0.025665	0.005747	-0.059894	0.137979	0.032019	-0.028046
restecg	-0.116211	-0.058196	0.044421	-0.114103	-0.151040	-0.084189	1.000000	0.044123	-0.070733	-0.058770	0.093045	-0.072042	-0.011981	0.137230
thalach	-0.398522	-0.044020	0.295762	-0.046698	-0.009940	-0.008567	0.044123	1.000000	-0.378812	-0.344187	0.386784	-0.213177	-0.096439	0.421741
exang	0.096801	0.141664	-0.394280	0.067616	0.067023	0.025665	-0.070733	-0.378812	1.000000	0.288223	-0.257748	0.115739	0.206754	-0.436757
oldpeak	0.210013	0.096093	-0.149230	0.193216	0.053952	0.005747	-0.058770	-0.344187	0.288223	1.000000	-0.577537	0.222682	0.210244	-0.430696
slope	-0.168814	-0.030711	0.119717	-0.121475	-0.004038	-0.059894	0.093045	0.386784	-0.257748	-0.577537	1.000000	-0.080155	-0.104764	0.345877
ca	0.276326	0.118261	-0.181053	0.101389	0.070511	0.137979	-0.072042	-0.213177	0.115739	0.222682	-0.080155	1.000000	0.151832	-0.391724
thal	0.068001	0.210041	-0.161736	0.062210	0.098803	-0.032019	-0.011981	-0.096439	0.206754	0.210244	-0.104764	0.151832	1.000000	-0.344029
target	-0.225439	-0.280937	0.433798	-0.144931	-0.085239	-0.028046	0.137230	0.421741	-0.436757	-0.430696	0.345877	-0.391724	-0.344029	1.000000

In [12]:

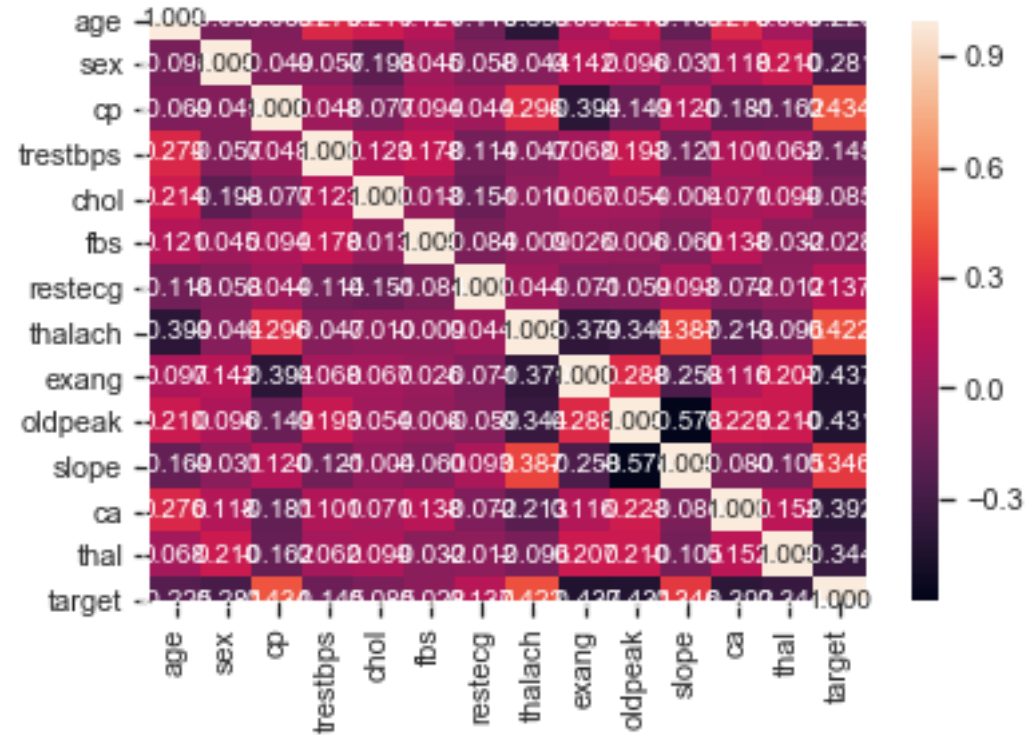
```
del data['B']
```

In [22]:

```
sns.heatmap(data.corr(), annot=True, fmt='.3f')

Out[22]:

<matplotlib.axes._subplots.AxesSubplot at 0x261d2c4a188>
```



```
In [14]:

del data['DIS']

In [15]:

del data['CHAS']

In [16]:

data.corr()

Out[16]:
```

	CRIM	ZN	INDUS	NOX	RM	TAX	PTRATIO	LSTAT	MEDV
CRIM	1.000000	-0.200469	0.406583	0.420972	-0.219247	0.582764	0.289946	0.455621	-0.388305
ZN	-0.200469	1.000000	-0.533828	-0.516604	0.311991	-0.314563	-0.391679	-0.412995	0.360445
INDUS	0.406583	-0.533828	1.000000	0.763651	-0.391676	0.720760	0.383248	0.603800	-0.483725
NOX	0.420972	-0.516604	0.763651	1.000000	-0.302188	0.668023	0.188933	0.590879	-0.427321
RM	-0.219247	0.311991	-0.391676	-0.302188	1.000000	-0.292048	-0.355501	-0.613808	0.695360
TAX	0.582764	-0.314563	0.720760	0.668023	-0.292048	1.000000	0.460853	0.543993	-0.468536
PTRATIO	0.289946	-0.391679	0.383248	0.188933	-0.355501	0.460853	1.000000	0.374044	-0.507787

	CRIM	ZN	INDUS	NOX	RM	TAX	PTRATIO	LSTAT	MEDV
LSTAT	0.455621	-0.412995	0.603800	0.590879	-0.613808	0.543993	0.374044	1.000000	-0.737663
MEDV	-0.388305	0.360445	-0.483725	-0.427321	0.695360	-0.468536	-0.507787	-0.737663	1.000000

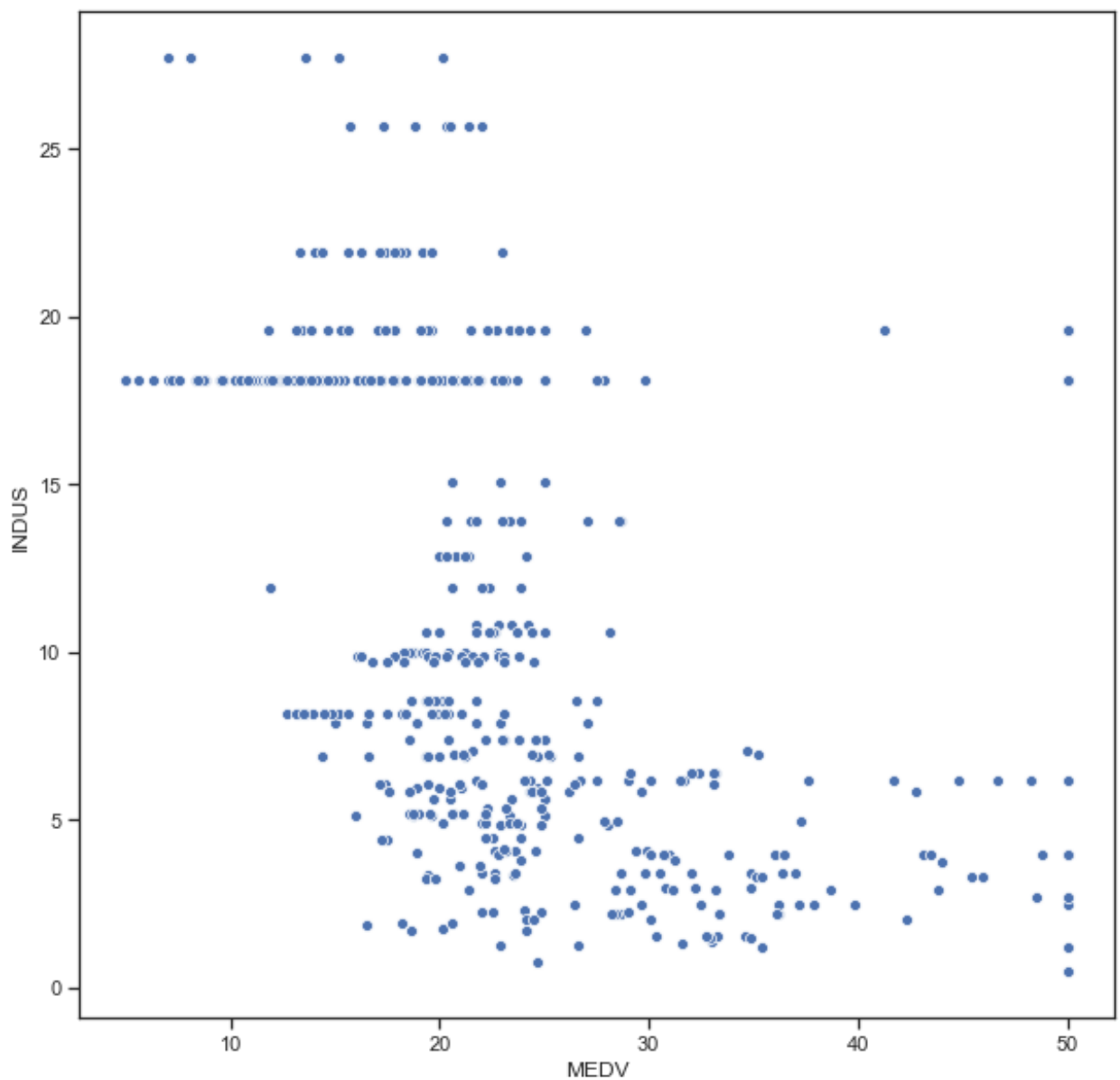
In [19]:

```
fig, ax = plt.subplots(figsize = (10, 10))
```

```
sns.scatterplot(ax = ax, x = 'MEDV', y = 'INDUS', data = data)
```

Out[19]:

<matplotlib.axes._subplots.AxesSubplot at 0x2b30641f808>

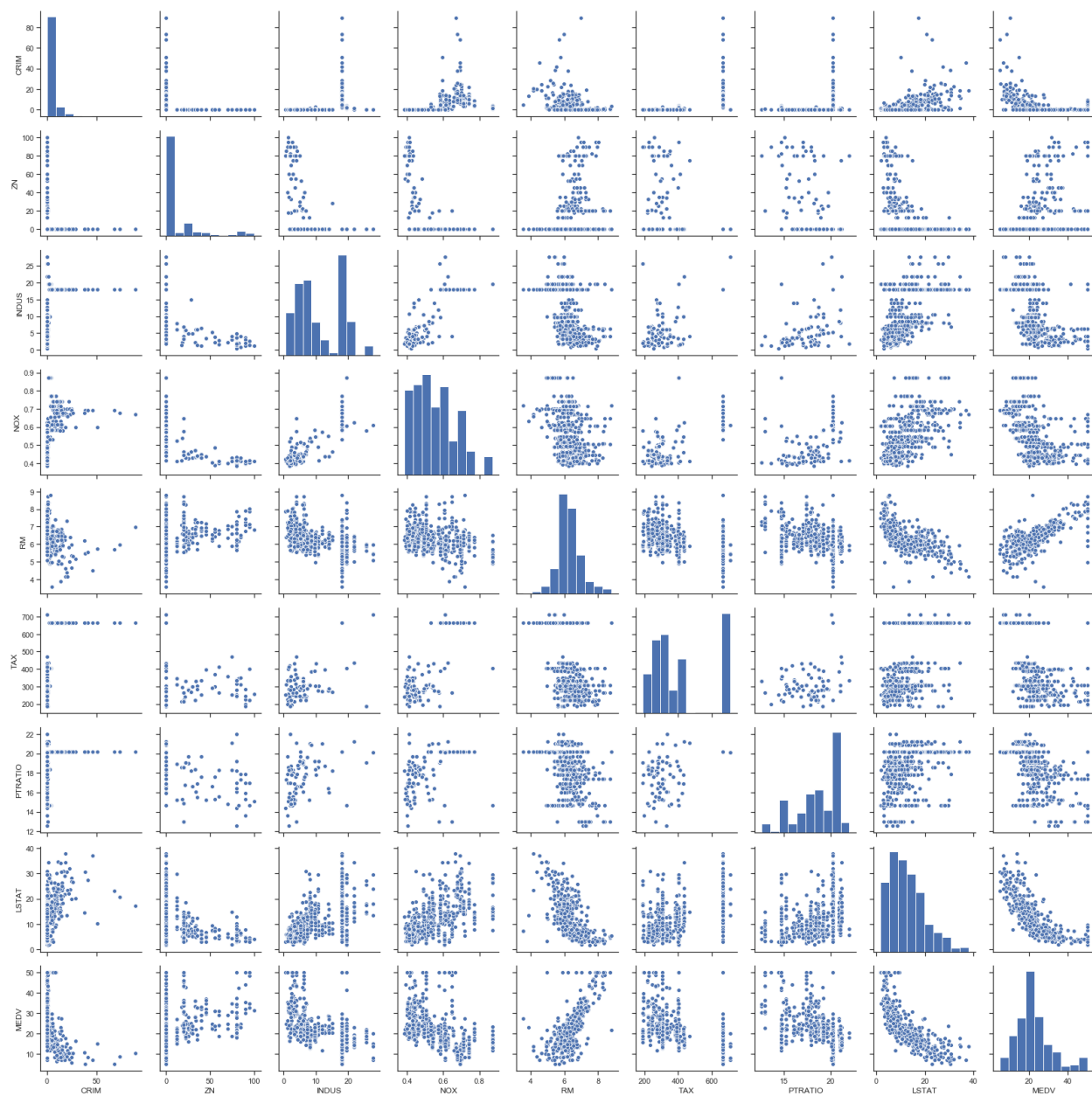


In [20]:

```
sns.pairplot(data)
```

Out[20]:

<seaborn.axisgrid.PairGrid at 0x2b30647d348>



In []: