



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

Выполнил:
студент группы ИУ5 – 62Б
Карягин А.Д.
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	61	3	1	145	233	1	0	150	0	2.3	0	0	1	1
1	31	2	1	130	250	0	1	187	0	3.5	0	0	2	1
2	40	1	1	130	204	0	0	172	0	1.4	2	0	2	1
3	51	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	50	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	trest b ps	chol	fbs	reste c g	thala c h	exan g	oldp e ak	slope	ca	thal	targe t
age	1.00 0 000	- 0.09 8 447	- 0.06 8 653	0.27 9 351	0.21 3 678	0.12 1 308	- 0.11 6 211	- 0.39 8 522	0.09 6 801	0.21 0 013	- 0.16 8 814	0.27 6 326	0.06 8 001	- 0.22 5 439
sex	- 0.09 8 447	1.00 0 000	- 0.04 9 353	- 0.05 6 769	- 0.19 7 912	0.04 5 032	- 0.05 8 196	- 0.04 4 020	0.14 1 664	0.09 6 093	- 0.03 0 711	0.11 8 261	0.21 0 041	- 0.28 0 937
cp	- 0.06 8 653	- 0.04 9 353	1.00 0 000	0.04 7 608	- 0.07 6 904	0.09 4 444	0.04 4 421	0.29 5 762	- 0.39 4 280	- 0.14 9 230	0.11 9 717	- 0.18 1 053	- 0.16 1 736	0.43 3 798
trest b ps	0.27 9 351	- 0.05 6 769	0.04 7 608	1.00 0 000	0.12 3 174	0.17 7 531	- 0.11 4 103	- 0.04 6 698	0.06 7 616	0.19 3 216	- 0.12 1 475	0.10 1 389	0.06 2 210	- 0.14 4 931
chol	0.21 3 678	- 0.19 7 912	- 0.07 6 904	0.12 3 174	1.00 0 000	0.01 3 294	- 0.15 1 040	- 0.00 9 940	0.06 7 023	0.05 3 952	- 0.00 4 038	0.07 0 511	0.09 8 803	- 0.08 5 239
fbs	0.12 1 308	0.04 5 032	0.09 4 444	0.17 7 531	0.01 3 294	1.00 0 000	- 0.08 4 189	- 0.00 8 567	0.02 5 665	0.00 5 747	- 0.05 9 894	0.13 7 979	- 0.03 2 019	- 0.02 8 046
reste c g	- 0.11 6 211	- 0.05 8 196	0.04 4 421	- 0.11 4 103	- 0.15 1 040	- 0.08 4 189	1.00 0 000	0.04 4 123	- 0.07 0 733	- 0.05 8 770	0.09 3 045	- 0.07 2 042	- 0.01 1 981	0.13 7 230
thal a ch	- 0.39 8 522	- 0.04 4 020	0.29 5 762	- 0.04 6 698	- 0.00 9 940	- 0.00 8 567	0.04 4 123	1.00 0 000	- 0.37 8 812	- 0.34 4 187	0.38 6 784	- 0.21 3 177	- 0.09 6 439	0.42 1 741
exan g	0.09 6 801	0.14 1 664	- 0.39 4 280	0.06 7 616	0.06 7 023	0.02 5 665	- 0.07 0 733	- 0.37 8 812	1.00 0 000	0.28 8 223	- 0.25 7 748	0.11 5 739	0.20 6 754	- 0.43 6 757
oldp e ak	0.21 0 013	0.09 6 093	- 0.14 9 230	0.19 3 216	0.05 3 952	0.00 5 747	- 0.05 8 770	- 0.34 4 187	0.28 8 223	1.00 0 000	- 0.57 7 537	0.22 2 682	0.21 0 244	- 0.43 0 696
slop e	- 0.16 8 814	- 0.03 0 711	0.11 9 717	- 0.12 1 475	- 0.00 4 038	- 0.05 9 894	0.09 3 045	0.38 6 784	- 0.25 7 748	- 0.57 7 537	1.00 0 000	- 0.08 0 155	- 0.10 4 764	0.34 5 877
ca	0.27 6 326	0.11 8 261	- 0.18 1 053	0.10 1 389	0.07 0 511	0.13 7 979	- 0.07 2 042	- 0.21 3 177	0.11 5 739	0.22 2 682	- 0.08 0 155	1.00 0 000	0.15 1 832	- 0.39 1 724

thal	0.068001	0.21041	-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	0.04421	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	0.04421	-0.114103	-0.151040	-0.084189	1.000000	0.04123	-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	0.04123	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	0.008747	-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

targ	-	-	0.43	-	-	-	0.13	0.42	-	-	0.34	-	-	1.00
e	0.22	0.28	3	0.14	0.08	0.02	7	1	0.43	0.43	5	0.39	0.34	0
t	5	0	798	4	5	8	230	741	6	0	877	1	4	000
	439	937		931	239	046			757	696		724	029	

In [12]:

del

data['B']

In [22]:

sns.heat

map(data

.corr(),

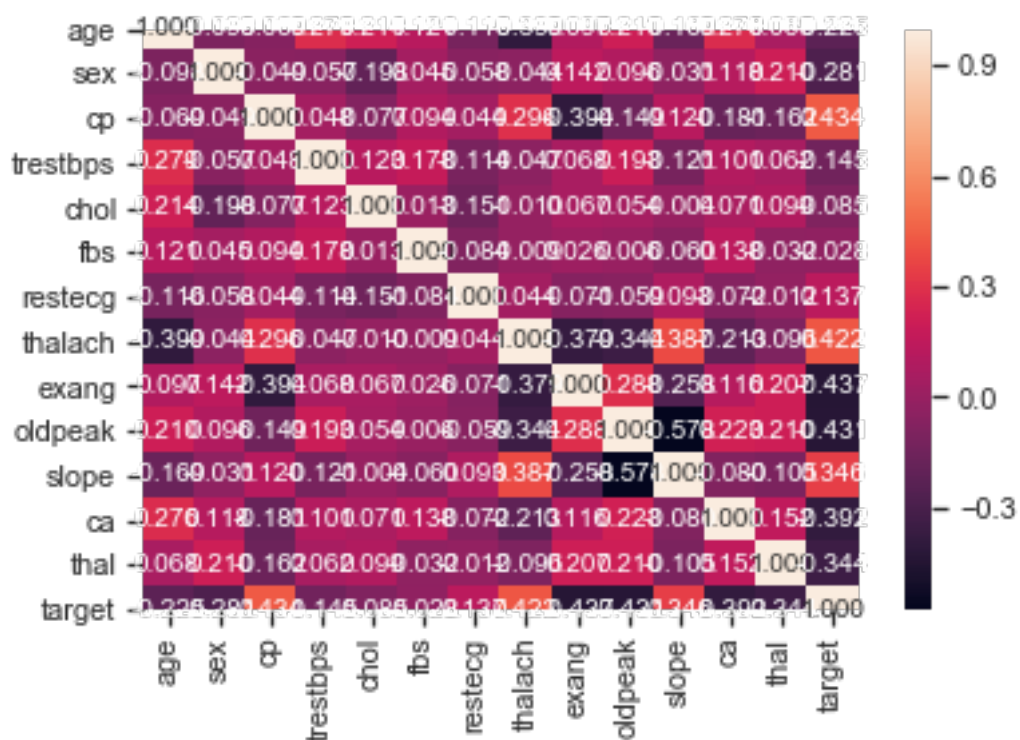
annot=Tr

ue,

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]:
```

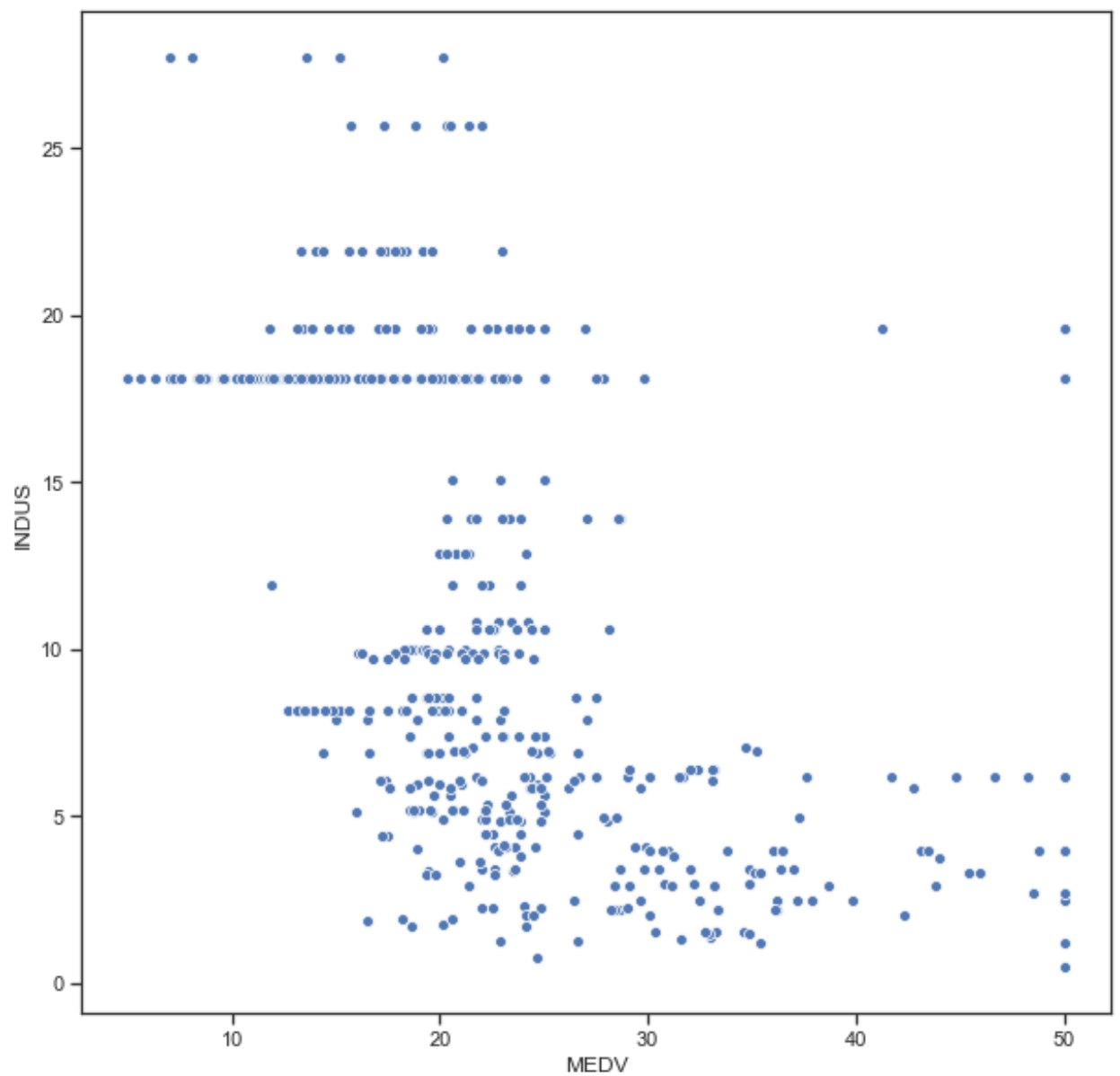
	CRIM	ZN	INDUS	NOX	RM	TAX	PTRATIO	LSTAT	MEDV
CRIM	1.000000	-0.200469	0.406583	0.420972	-0.219247	0.582764	0.289946	0.455621	-0.388305
ZN	-0.200469	1.000000	-0.533828	-0.516604	0.311991	-0.314563	-0.391679	-0.412995	0.360445
INDUS	0.406583	-0.533828	1.000000	0.763651	-0.391676	0.720760	0.383248	0.603800	-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>
```

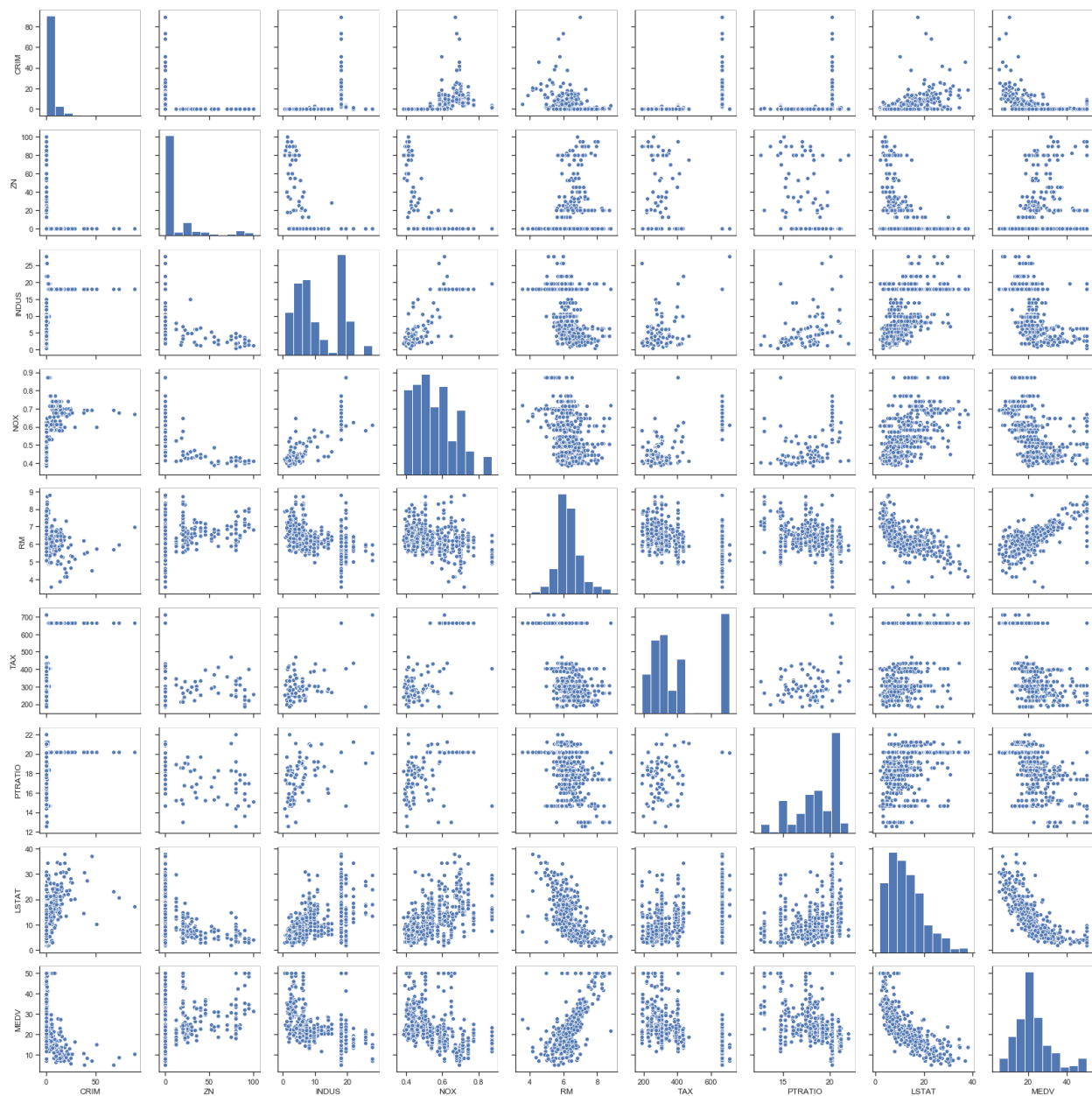


In [20]:

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

Out[20]:

<seaborn.axisgrid.PairGrid at 0x2b30647d348>



In []: