

**Московский государственный технический
университет им. Н.Э. Баумана**

Факультет «Информатика и системы управления»
Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Теория машинного обучения»

Отчет по рубежному контролю №1

Выполнил:

студент группы ИУ5-63Б

Ветошкин Артём

Подпись и дата:

14.06.22

Проверил:

Юрий Евгеньевич Гапанюк

Подпись и дата:

Москва, 2022 г.

Рубежный контроль №1

Загружаем библиотеки

```
#Загружаем все библиотеки
import numpy as np
import pandas as pd
from sklearn.datasets import *
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

Загружаем данные

```
df = pd.read_csv('heart.csv')
```

```
df
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows × 14 columns

Изучаем набор данных

```
df.dtypes
```

```
age          int64
sex          int64
cp           int64
trestbps     int64
chol         int64
fbs          int64
restecg      int64
thalach      int64
exang        int64
oldpeak      float64
```

```
slope      int64
ca         int64
thal       int64
target     int64
dtype: object
```

```
# Посмотрим наличие пустых значений
df.isnull().sum()
```

```
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
```

Корреляционный анализ

```
df.corr()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
age	1.000000	-0.103240	-0.071966	0.271121	0.219823	0.121243	-0.132696	-0.390227	0.088163	0.208137	-0.169105	0.271551
sex	-0.103240	1.000000	-0.041119	-0.078974	-0.198258	0.027200	-0.055117	-0.049365	0.139157	0.084687	-0.026666	0.111729
cp	-0.071966	-0.041119	1.000000	0.038177	-0.081641	0.079294	0.043581	0.306839	-0.401513	-0.174733	0.131633	-0.176206
trestbps	0.271121	-0.078974	0.038177	1.000000	0.127977	0.181767	-0.123794	-0.039264	0.061197	0.187434	-0.120445	0.104554
chol	0.219823	-0.198258	-0.081641	0.127977	1.000000	0.026917	-0.147410	-0.021772	0.067382	0.064880	-0.014248	0.074259
fbs	0.121243	0.027200	0.079294	0.181767	0.026917	1.000000	-0.104051	-0.008866	0.049261	0.010859	-0.061902	0.137156
restecg	-0.132696	-0.055117	0.043581	-0.123794	-0.147410	-0.104051	1.000000	0.048411	-0.065606	-0.050114	0.086086	-0.078072
thalach	-0.390227	-0.049365	0.306839	-0.039264	-0.021772	-0.008866	0.048411	1.000000	-0.380281	-0.349796	0.395308	-0.207888
exang	0.088163	0.139157	-0.401513	0.061197	0.067382	0.049261	-0.065606	-0.380281	1.000000	0.310844	-0.267335	0.107849
oldpeak	0.208137	0.084687	-0.174733	0.187434	0.064880	0.010859	-0.050114	-0.349796	0.310844	1.000000	-0.575189	0.221816
slope	-0.169105	-0.026666	0.131633	-0.120445	-0.014248	-0.061902	0.086086	0.395308	-0.267335	-0.575189	1.000000	-0.073440
ca	0.271551	0.111729	-0.176206	0.104554	0.074259	0.137156	-0.078072	-0.207888	0.107849	0.221816	-0.073440	1.000000
thal	0.072297	0.198424	-0.163341	0.059276	0.100244	-0.042177	-0.020504	-0.098068	0.197201	0.202672	-0.094090	0.149014
target	-0.229324	-0.279501	0.434854	-0.138772	-0.099966	-0.041164	0.134468	0.422895	-0.438029	-0.438441	0.345512	-0.382085

```
df.corr(method='spearman')
```

```

.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}

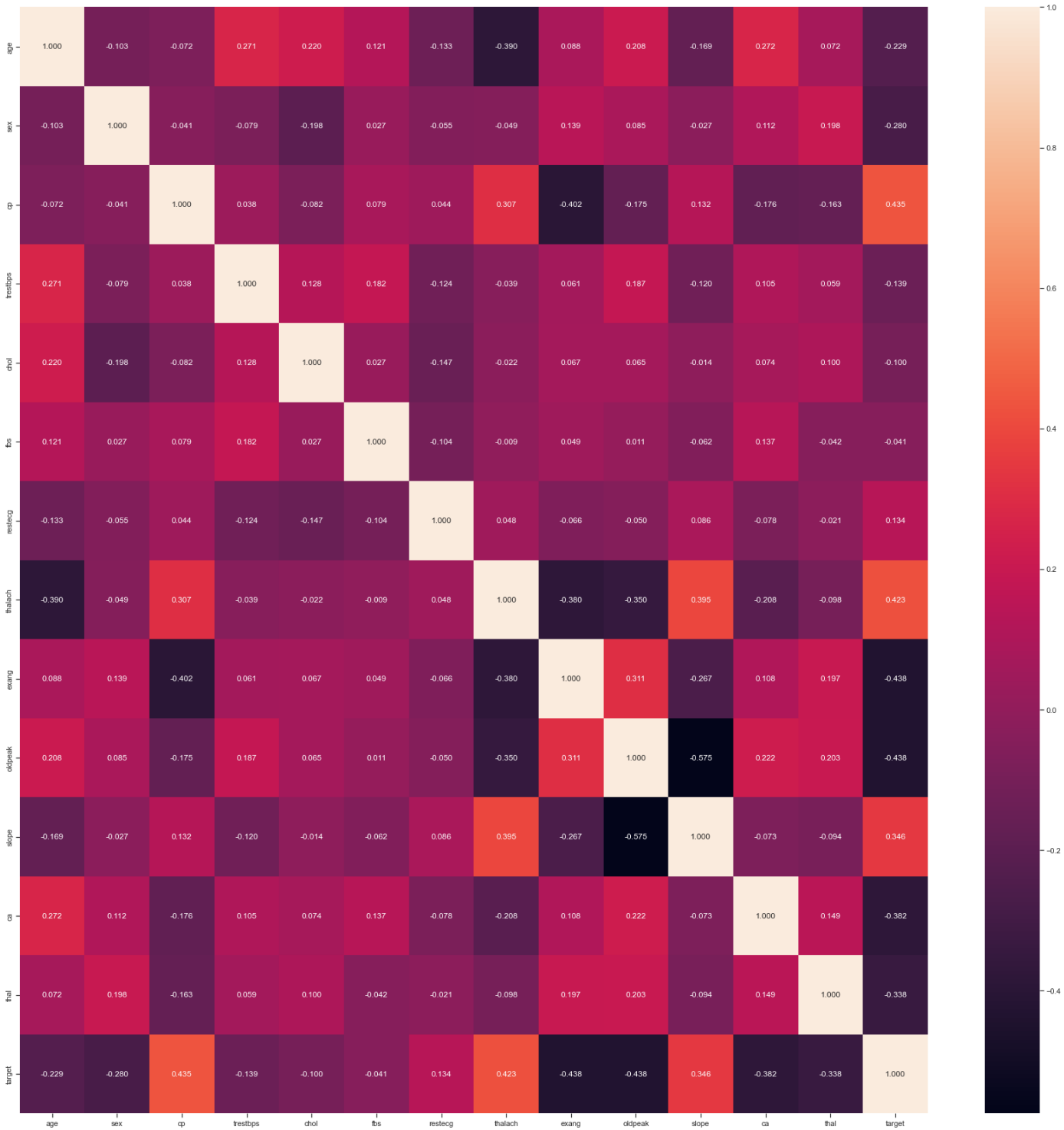
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
age	1.000000	-0.105479	-0.092148	0.280189	0.203253	0.110958	-0.149732	-0.382724	0.077594	0.264500	-0.180627	0.339808
sex	-0.105479	1.000000	-0.053450	-0.071923	-0.152997	0.027200	-0.044169	-0.047053	0.139157	0.090896	-0.020843	0.113834
cp	-0.092148	-0.053450	1.000000	0.025526	-0.092239	0.073738	0.066715	0.336116	-0.427933	-0.195285	0.169307	-0.217002
trestbps	0.280189	-0.071923	0.025526	1.000000	0.127010	0.152272	-0.137273	-0.028880	0.043791	0.146722	-0.081828	0.091323
chol	0.203253	-0.152997	-0.092239	0.127010	1.000000	0.034139	-0.163354	-0.054794	0.088895	0.057102	-0.026660	0.114584
fbs	0.110958	0.027200	0.073738	0.152272	0.034139	1.000000	-0.101709	-0.016292	0.049261	0.032542	-0.048340	0.138197
restecg	-0.149732	-0.044169	0.066715	-0.137273	-0.163354	-0.101709	1.000000	0.091548	-0.073428	-0.075213	0.109506	-0.108580
thalach	-0.382724	-0.047053	0.336116	-0.028880	-0.054794	-0.016292	0.091548	1.000000	-0.405000	-0.439987	0.443417	-0.254775
exang	0.077594	0.139157	-0.427933	0.043791	0.088895	0.049261	-0.073428	-0.405000	1.000000	0.322149	-0.282593	0.158985
oldpeak	0.264500	0.090896	-0.195285	0.146722	0.057102	0.032542	-0.075213	-0.439987	0.322149	1.000000	-0.594413	0.230183
slope	-0.180627	-0.020843	0.169307	-0.081828	-0.026660	-0.048340	0.109506	0.443417	-0.282593	-0.594413	1.000000	-0.095393
ca	0.339808	0.113834	-0.217002	0.091323	0.114584	0.138197	-0.108580	-0.254775	0.158985	0.230183	-0.095393	1.000000
thal	0.088586	0.241115	-0.212186	0.058995	0.084941	-0.013372	-0.017285	-0.159190	0.240512	0.253489	-0.146059	0.186155
target	-0.240326	-0.279501	0.464894	-0.115009	-0.132926	-0.041164	0.147402	0.429832	-0.438029	-0.437669	0.368808	-0.453329

```

fg, ax = plt.subplots(figsize=(30,30))
sns.heatmap(df.corr(), annot=True, fmt='.3f', ax=ax)
plt.show()

```



По полученной корреляционной матрице, можно сделать вывод, что модель может получиться хорошей, т.к. целевой параметр имеет заметную корреляцию с большим числом признаков. Самый большой вклад в целевой параметр вносят признаки **cp, exang, thalach, oldpeak**.