



**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 вычислительная техника»
Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Технологии машинного обучения»

Отчет рубежному контролю №1
Вариант 7

Выполнил:
студент группы ИУ5-62Б

Веревкина Диана В.
Подпись и дата:

Проверил:
преподаватель каф.
ИУ5
Гапанюк Ю.Е.
Подпись и дата:

Москва, 2022 г.

Задание

Номер варианта	Номер задачи	Номер набора данных, указанного в задаче
7	1	7

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 ИУ5-62Б, ИУ5Ц-82Б - для произвольной колонки данных построить гистограмму.

<https://www.kaggle.com/mohansacharya/graduate-admissions>

Решение

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style = 'ticks')
df = pd.read_csv('Admission_Predict.csv')
df.head()
```

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65

Обработка пропусков

```
#обработка пропусков
#поиск пропусков
df.isna()
df.dropna()
```

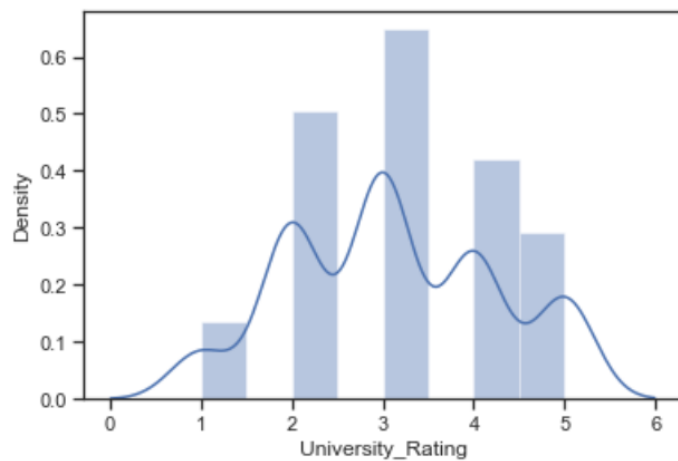
	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65
...
495	496	332	108	5	4.5	4.0	9.02	1	0.87
496	497	337	117	5	5.0	5.0	9.87	1	0.96
497	498	330	120	5	4.5	5.0	9.56	1	0.93
498	499	312	103	4	4.0	5.0	8.43	0	0.73
499	500	327	113	4	4.5	4.5	9.04	0	0.84

500 rows × 9 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Serial No.            500 non-null   int64  
1   GRE Score              500 non-null   int64  
2   TOEFL Score           500 non-null   int64  
3   University Rating     500 non-null   int64  
4   SOP                   500 non-null   float64 
5   LOR                   500 non-null   float64 
6   CGPA                  500 non-null   float64 
7   Research              500 non-null   int64  
8   Chance of Admit       500 non-null   float64 
dtypes: float64(4), int64(5)
memory usage: 35.3 KB
```

```
sns_plot = sns.distplot(df['University_Rating'])
fig = sns_plot.get_figure()
```

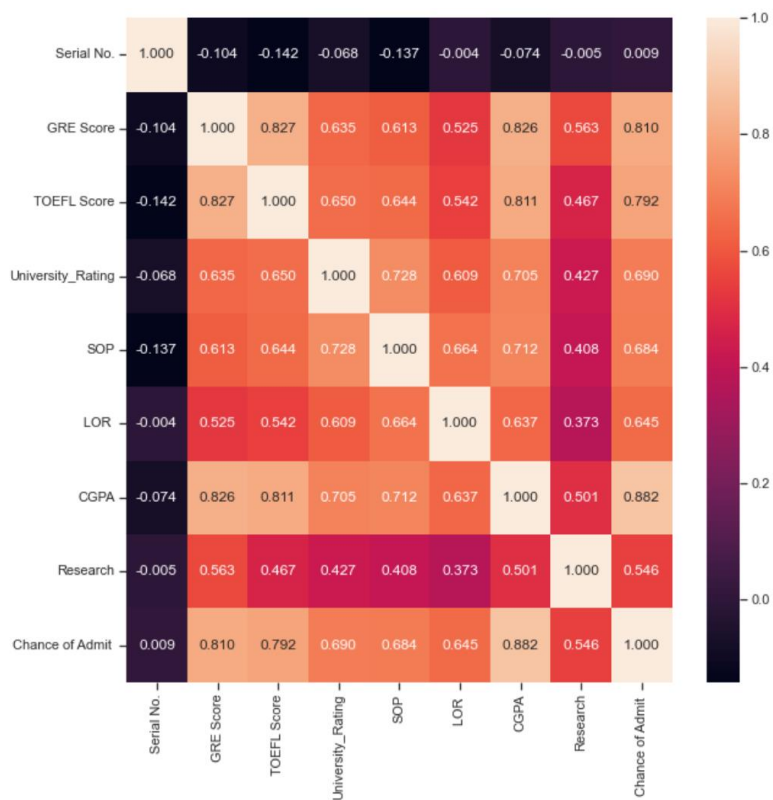


Корреляционный анализ

```
#Корреляционный анализ
df.corr()
```

	Serial No.	GRE Score	TOEFL Score	University_Rating	SOP	LOR	CGPA	Research	Chance of Admit
Serial No.	1.000000	-0.103839	-0.141696	-0.067641	-0.137352	-0.003694	-0.074289	-0.005332	0.008505
GRE Score	-0.103839	1.000000	0.827200	0.635376	0.613498	0.524679	0.825878	0.563398	0.810351
TOEFL Score	-0.141696	0.827200	1.000000	0.649799	0.644410	0.541563	0.810574	0.467012	0.792228
University_Rating	-0.067641	0.635376	0.649799	1.000000	0.728024	0.608651	0.705254	0.427047	0.690132
SOP	-0.137352	0.613498	0.644410	0.728024	1.000000	0.663707	0.712154	0.408116	0.684137
LOR	-0.003694	0.524679	0.541563	0.608651	0.663707	1.000000	0.637469	0.372526	0.645365
CGPA	-0.074289	0.825878	0.810574	0.705254	0.712154	0.637469	1.000000	0.501311	0.882413
Research	-0.005332	0.563398	0.467012	0.427047	0.408116	0.372526	0.501311	1.000000	0.545871
Chance of Admit	0.008505	0.810351	0.792228	0.690132	0.684137	0.645365	0.882413	0.545871	1.000000

```
plt.figure(figsize=(10,10))
sns.heatmap(df.corr(), annot=True, fmt='.3f')
```



Из представленной матрицы можно сделать вывод, что наибольшая корреляция наблюдается между признаками:

- GRE Score и TOEFL Score
- GRE Score и CGPA
- GRE Score и Chance of Admit
- TOEFL Score и CGPA
- CGPA и Chance of Admit