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 Системы обработки информации и управления

Отчёт по рубежному контролю №1

По дисциплине:
«Технологии машинного обучения»

Выполнил:

Студент группы ИУ5 _____ **Ахвердиев В.И**

(Подпись, дата) (Фамилия И.О.)

Проверил:

_____ **Гапанюк Ю. Е.**

(Подпись, дата) (Фамилия И.О.)

Москва, 2021

Задание

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 ИУ5-61Б - для пары произвольных колонок данных построить график "Диаграмма рассеяния".

Набор данных:

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html#sklearn.datasets.load_wine

РК

Импорт библиотек

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
import warnings
from sklearn import datasets
from sklearn.datasets import load_wine
from sklearn import linear_model
from sklearn.cluster import KMeans
from sklearn import metrics
from pandas import DataFrame
%pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

In [2]:
```
boston = load_wine()
data = pd.DataFrame(boston.data, columns=boston.feature_names)
data['TARGET'] = boston.target
```

In [3]:
```
data.head()
```

Out[3]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonfl
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06	
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76	
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24	
3	14.37	1.95	2.50	16.8	113.0	3.85	3.49	
4	13.24	2.59	2.87	21.0	118.0	2.80	2.69	

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   alcohol                       178 non-null    float64
 1   malic_acid                    178 non-null    float64
 2   ash                           178 non-null    float64
 3   alcalinity_of_ash             178 non-null    float64
 4   magnesium                     178 non-null    float64
 5   total_phenols                 178 non-null    float64
 6   flavanoids                    178 non-null    float64
 7   nonflavanoid_phenols          178 non-null    float64
 8   proanthocyanins               178 non-null    float64
 9   color_intensity               178 non-null    float64
 10  hue                           178 non-null    float64
 11  od280/od315_of_diluted_wines  178 non-null    float64
 12  proline                       178 non-null    float64
 13  TARGET                        178 non-null    int64  
dtypes: float64(13), int64(1)
memory usage: 19.6 KB
```

In [5]:
```
data.describe()
```

Out[5]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	f
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	17
mean	13.000618	2.336348	2.366517	19.494944	99.741573	2.295112	
std	0.811827	1.117146	0.274344	3.339564	14.282484	0.625851	
min	11.030000	0.740000	1.360000	10.600000	70.000000	0.980000	
25%	12.362500	1.602500	2.210000	17.200000	88.000000	1.742500	
50%	13.050000	1.865000	2.360000	19.500000	98.000000	2.355000	
75%	13.677500	3.082500	2.557500	21.500000	107.000000	2.800000	
max	14.830000	5.800000	3.230000	30.000000	162.000000	3.880000	

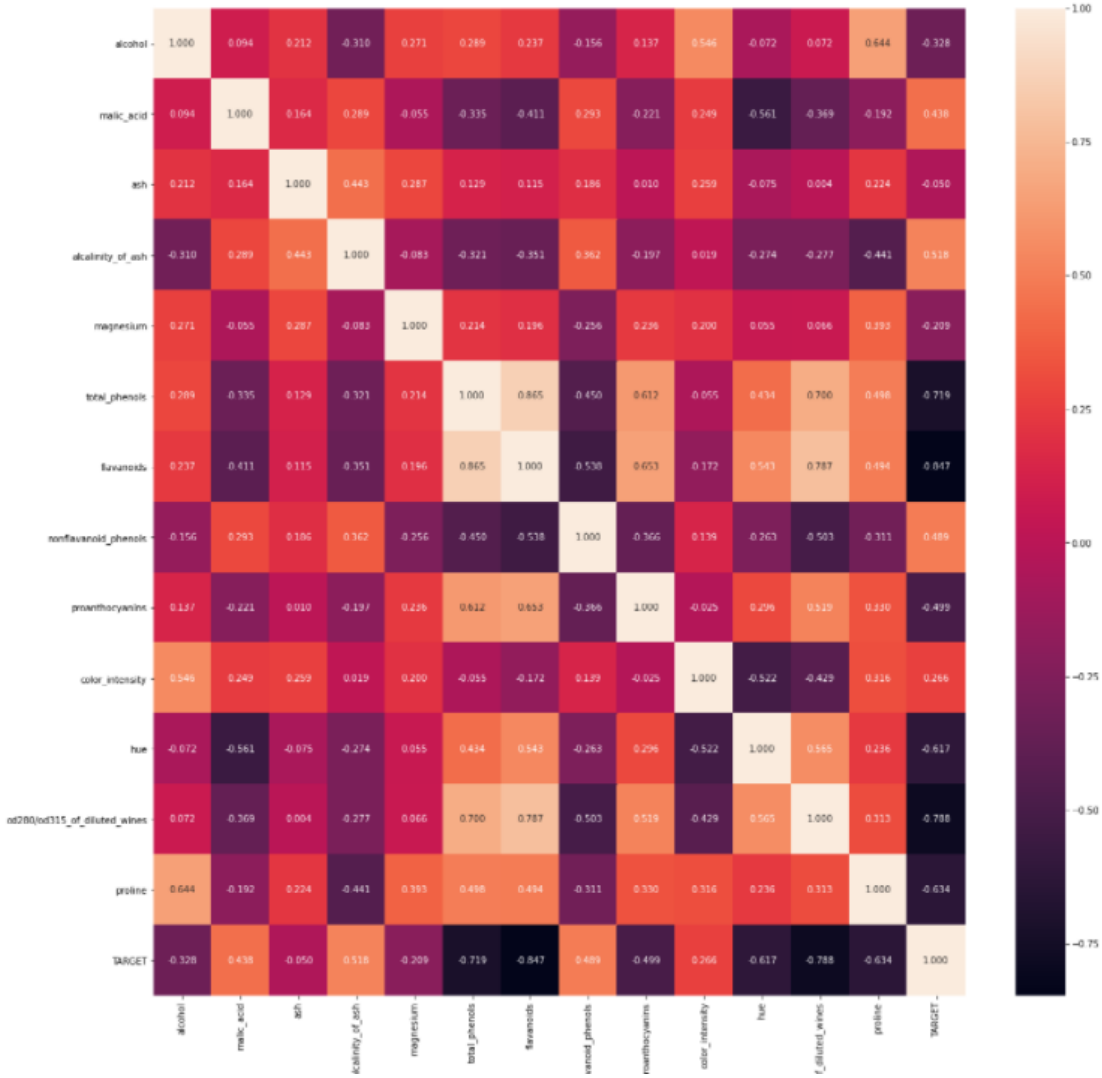
In [6]:
```
## Корр. анализ
corr_matrix = data.corr()
```

In [7]:
```
corr_matrix['TARGET']
```

Out[7]:
```
alcohol                        -0.328222
malic_acid                      0.437776
ash                            -0.049643
alcalinity_of_ash               0.517859
magnesium                      -0.209179
total_phenols                  -0.719163
flavanoids                     -0.847498
nonflavanoid_phenols            0.489109
proanthocyanins                -0.499130
color_intensity                 0.265668
hue                            -0.617369
od280/od315_of_diluted_wines   -0.788230
proline                        -0.633717
TARGET                          1.000000
Name: TARGET, dtype: float64
```

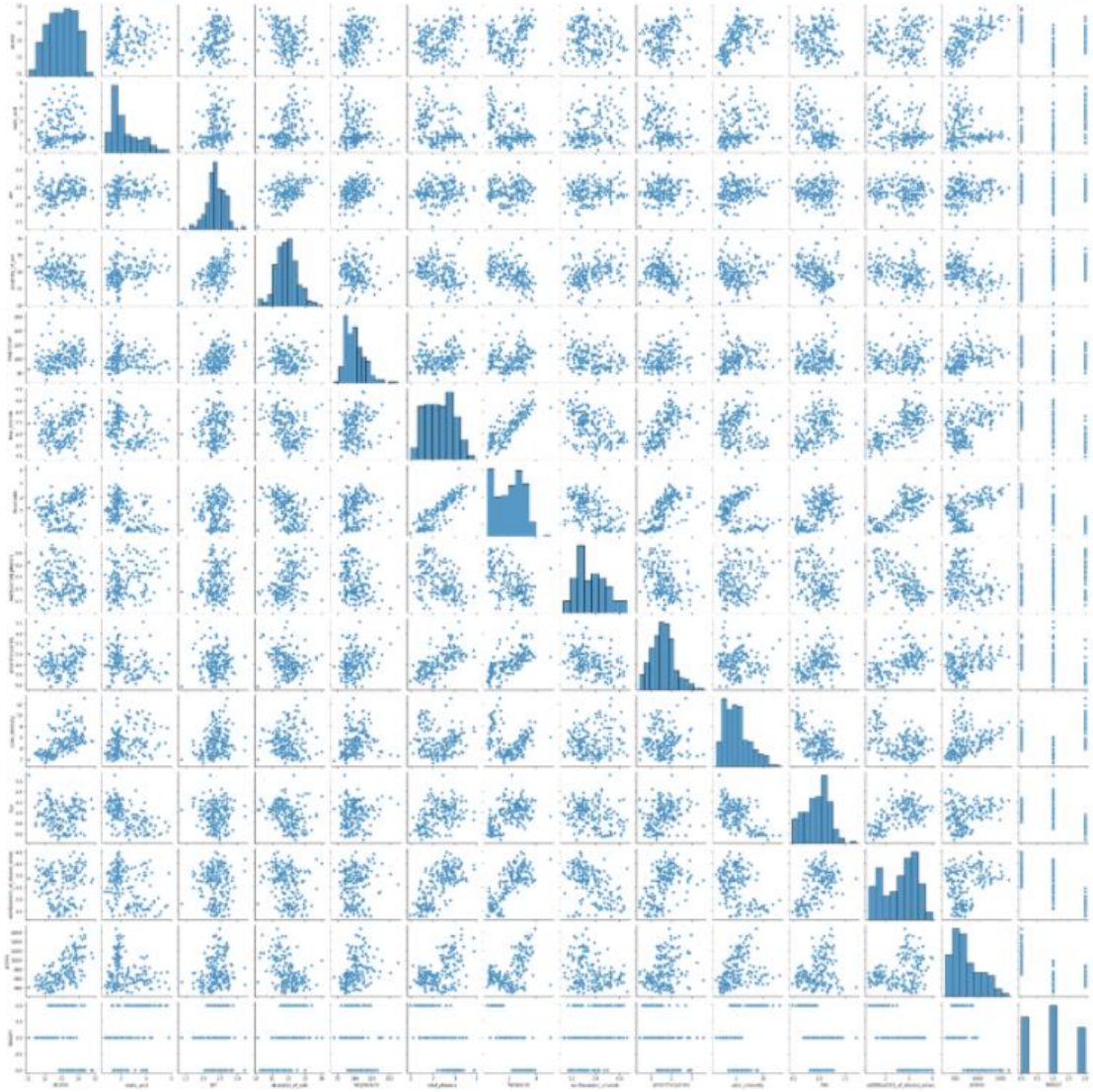
In [8]:
```
plt.figure(figsize=(20,20))
sns.heatmap(corr_matrix, annot=True, fmt='.3f')
```

Out[8]: <AxesSubplot:>



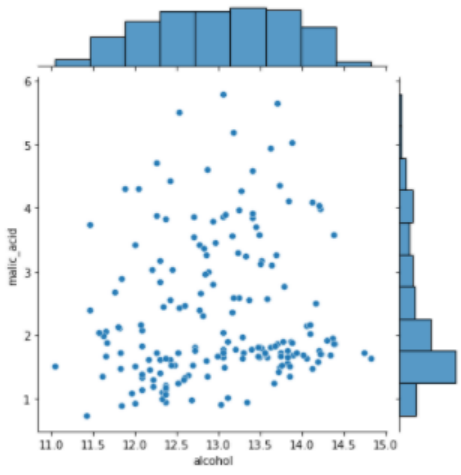
Out[9]: <seaborn.axisgrid.PairGrid at 0x1f8e34943a0>

<Figure size 864x432 with 0 Axes>



In [10]:
```
# Увеличенные диаграммы рассеяния
sns.jointplot(x = "alcohol", y = "malic_acid", kind="scatter", data = data)
```

Out[10]: <seaborn.axisgrid.JointGrid at 0x1f8ebe04fd0>



In [11]:
```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(ax=ax, x='alcohol', y='malic_acid', data=data, hue='proline')
```

Out[11]:
```
<AxesSubplot:xlabel='alcohol', ylabel='malic_acid'>
```