



**Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное 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»**

**Курс «Технологии машинного обучения»
Отчёт по рубежному контролю №1
«Технологии разведочного анализа и обработки данных.»
Вариант № 2**

Выполнила:
студентка группы ИУ5-62Б Вешторт Е.С.

Дата:
13.04.2025

Подпись:


Проверил:
Гапанюк Ю.Е.

Дата:

Подпись:

2025 г.

РК1, вариант 2, Вешторт Ева, ИУ5-62Б

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.

Набор данных: load_wine

Для произвольной колонки данных построить гистограмму.

```
In [1]: from sklearn.datasets import load_wine
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Загрузка данных

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

Проверка наличия пропусков

```
In [3]: df.isnull().sum()
```

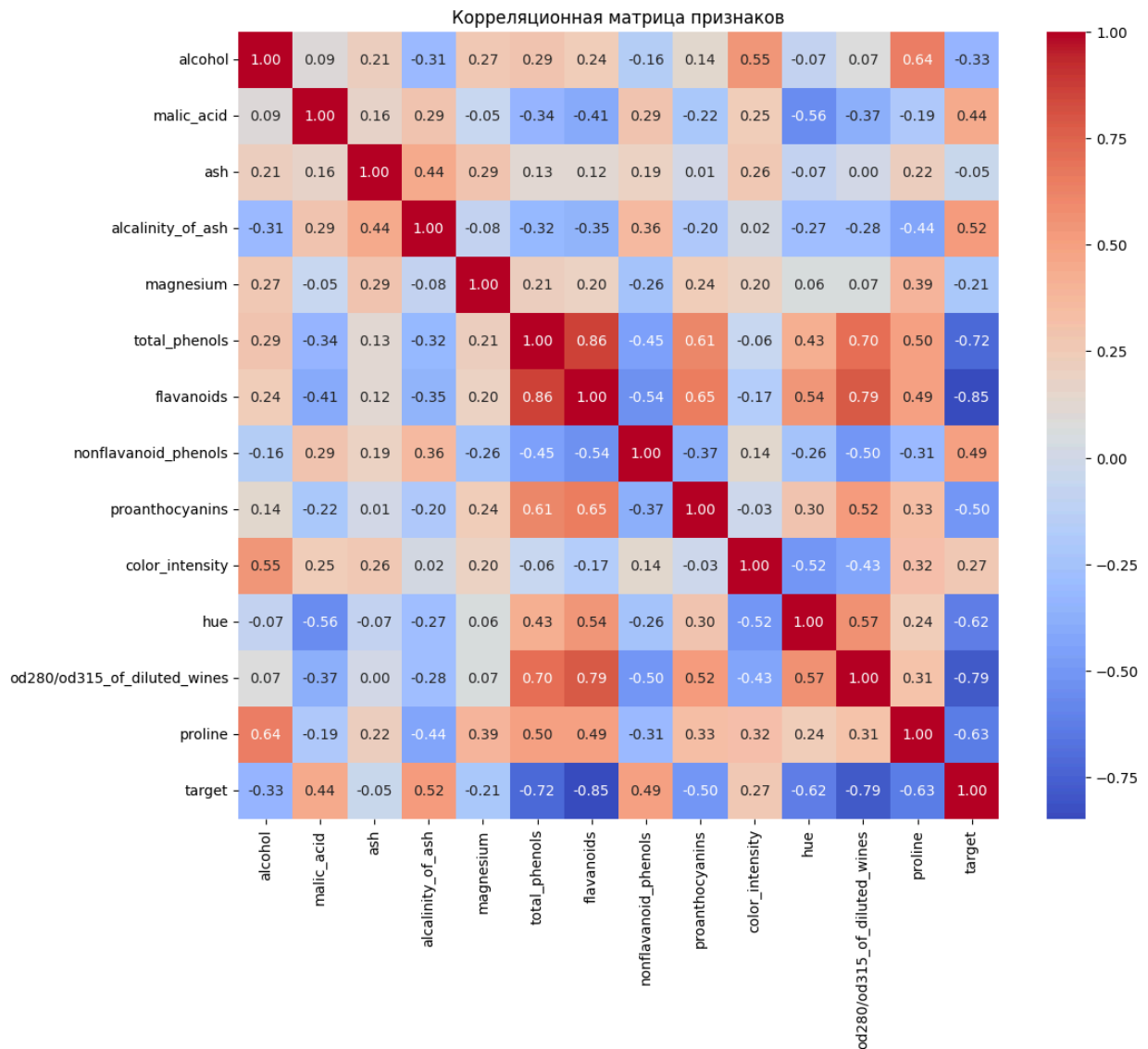
```
Out[3]: alcohol                0
malic_acid                    0
ash                           0
alcalinity_of_ash             0
magnesium                     0
total_phenols                 0
flavanoids                    0
nonflavanoid_phenols          0
proanthocyanins               0
color_intensity               0
hue                           0
od280/od315_of_diluted_wines  0
proline                       0
target                        0
dtype: int64
```

Данные не содержат пропусков

Построение корреляционной матрицы

```
In [4]: corr_matrix = df.corr(numeric_only=True)

plt.figure(figsize=(12, 10))
sns.heatmap(corr_matrix, annot=True, fmt='.2f', cmap='coolwarm')
plt.title('Корреляционная матрица признаков')
plt.show()
```



Анализ корреляций

- Целевая переменная имеет сильную отрицательную корреляцию с признаками flavonoids, od280/od315_of_diluted_wines и total_phenols, среднюю отрицательную корреляцию с признаками proline, hue и proanthocyanins, а также среднюю положительную корреляцию с признаками alcalinity_of_ash и nonflavanoid_phenols.
- Признаки flavonoids, od280/od315_of_diluted_wines и total_phenols имеют сильную положительную корреляцию между собой, а также среднюю положительную корреляцию с признаками proline, hue и proanthocyanins.

Вывод

- Корреляции между признаками и целевой переменной позволяют ожидать хорошую точность от классификационных моделей.
- Самый высокий вклад в модель будут вносить признаки flavonoids, od280/od315_of_diluted_wines и total_phenols, средний - proline, hue, proanthocyanin, alcalinity_of_ash и nonflavanoid_phenols, влияние остальных признаков будет меньше.
- Некоторые признаки можно отбросить или использовать для снижения размерности.

Дополнительное задание, построение гистограммы

```
In [5]: plt.figure(figsize=(8, 5))
plt.hist(df['alcohol'], bins=20, color='skyblue', edgecolor='black')
plt.title('Гистограмма признака "alcohol"')
plt.xlabel('Значение')
plt.ylabel('Частота')
plt.grid(True)
plt.show()
```

