

ТМО РК №1

Кузьмин Александр Юрьевич ИУ5-62Б

Вариант №9

Задача №2.

Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

Дополнительно для ИУ5-62Б: для произвольной колонки данных построить гистограмму.

Набор данных №1:

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

Импорты:

```
from sklearn.datasets import load_iris
import pandas as pd
import matplotlib.pyplot as plt
```

Работа:

```
iris = load_iris()
data = pd.DataFrame(data = iris.data, columns = iris.feature_names)
data.head()

   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width
0                5.1                3.5                1.4
0.2
1                4.9                3.0                1.4
0.2
2                4.7                3.2                1.3
0.2
3                4.6                3.1                1.5
0.2
4                5.0                3.6                1.4
0.2

data.dtypes
```

```
sepal length (cm)    float64
sepal width (cm)     float64
petal length (cm)    float64
petal width (cm)     float64
dtype: object
```

```
data.isnull().sum()
```

```
sepal length (cm)    0
sepal width (cm)     0
petal length (cm)    0
petal width (cm)     0
dtype: int64
```

Пропусков в данных нет

```
data.shape
```

```
(150, 4)
```

```
data = data.dropna(subset=['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)'])
```

```
data.shape
```

```
(150, 4)
```

Ничего не изменилось

Построение гистограммы, отображающей распределение длин чашелистиков ирисов:

```
data['sepal length (cm)'].hist(edgecolor='black', grid=False)
plt.title('Распределение длин чашелистиков Ирисов (см)')
plt.xlabel('Длина чашелистика')
plt.ylabel('Частота')
plt.show()
```

Распределение длин чашелистиков Ирисов (см)

