

Московский государственный технический университет им. Н.Э. Баумана
Факультет «Информатика, искусственный интеллект и системы управления»
Кафедра «Системы обработки информации и управления»



Отчет по рубежному контролю №1
по дисциплине «Методы машинного обучения»
Вариант 7

ИСПОЛНИТЕЛИ:
Пенегина В.В.

ПРЕПОДАВАТЕЛЬ:
Гапанюк Ю.Е.

Москва, 2023

Задание 7:

Для набора данных проведите устранение пропусков для одного (произвольного) числового признака с использованием метода заполнения медианой.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# загрузить набор данных в объект pandas DataFrame
data = pd.read_csv('C:/Users/0ner/PycharmProjects/RPSgame/winequality-red.csv')

# до заполнения
print(data.isnull().sum())

# определить, какой столбец содержит числовой признак, для которого требуется заполнить пропуски
numeric_feature = 'fixed acidity'

# рассчитать медиану числового признака
median = data[numeric_feature].median()
```

```
# заменить все пропущенные значения в столбце на вычисленную медиану
data[numeric_feature].fillna(median, inplace=True)

# проверить, что все пропуски были заполнены
print(data.isnull().sum())

# сохранить изменения в исходном наборе данных
data.to_csv('C:/Users/0ner/PycharmProjects/RPSgame/winequality-red.csv', index=False)

# загрузить данные
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv', delimiter=';')

# построить парные диаграммы
sns.pairplot(df, vars=['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar', 'chlorides',
                      'free sulfur dioxide', 'total sulfur dioxide', 'density', 'pH', 'sulphates', 'alcohol', 'quality'])

# показать диаграммы
plt.show()
```

Доп. задание:

