

## Технологии искусственного интеллекта

### Задание 1. Описательный анализ данных

**Содержание задания:** необходимо выполнить описательный анализ выбранного набора данных – найти числовые характеристики, визуально представить признаки (переменные) данных для сравнения, выявить корреляционные зависимости.

#### **Замечания по выполнению:**

1. Задание выполняется в среде программирования, поддерживающей язык python.
2. При выполнении задания необходимо писать комментарии к используемым функциям/командам, например, «Вычисляем характеристики данных» или «Строим линейный график для переменной (указать имя переменной)».

#### **Некоторые команды, которые могут понадобиться в процессе выполнения задания:**

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

.read_csv()
.head()
.describe()

.pairplot()      (из seaborn)

.plot()          (из matplotlib.pyplot)
.scatter()       (из matplotlib.pyplot)
.bar()           (из matplotlib.pyplot)
.pie()           (из matplotlib.pyplot)

.corr()
.groupby()
```

#### **Этапы выполнения задания:**

**Этап 1.** Выбрать набор данных, содержащий:

- числовые и категориальные переменные;
- число переменных (атрибутов) не менее 6;
- число объектов (наблюдений) не менее 100.

Источник данных:

Сайт «Machine Learning Repository»

<http://archive.ics.uci.edu/ml/datasets.php>

*Указать ссылку на выбранный набор данных в рабочем файле.*

**Этап 2.** По выбранному набору данных провести описательный анализ:

**2.1.** Найти характеристики данных:

- для числовых переменных:
  - минимальное;
  - максимальное;
  - среднее;
  - дисперсия;
  - стандартное отклонение;
  - квартили: 25%, 50% (медиана), 75%;
- для категориальных переменных:
  - минимальное;
  - максимальное;
  - мода.

**2.2.** Построить диаграмму размаха (ящик с усами) для двух выбранных переменных.

**2.3.** По полученным результатам анализа данных этапа 2 сделать выводы (не менее пяти).

**Этап 3.** По выбранному набору данных провести визуальный анализ:

**3.1.** Построить таблицу с графиками зависимостей между числовыми переменными. Сделать вывод о виде распределения переменных и наличия взаимосвязи между переменными.

**3.2.** Построить для отдельных переменных графики (с подписями осей, заголовком (подзаголовком) графика, легендой и прочим оформлением):

- линейный;
- точечный;
- столбиковый;
- круговой.

**3.3.** Сгруппировать по категориальной переменной данные, отобразить на графике (столбиковом) различия в средних значениях произвольной метрической переменной для разных категорий.

**3.4.** По полученным результатам анализа данных этапа 3 сделать выводы (не менее пяти).

**Этап 4.** По выбранному набору данных провести корреляционный анализ:

**4.1.** Построить таблицу с числовыми зависимостями (матрицу корреляционных коэффициентов).

**4.2.** По полученным результатам анализа данных этапа 4 сделать выводы (не менее пяти).