МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ



Дудкин Даниил Сергеевич

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1

IoT и Искусственный интеллект

Группа SE-231M

Задание

Изучив данные о пассажирах Титаника из датасета titanic.scv (ссылка: https://github.com/tpo9hbi4/MachineDeepLearning) и используя описанные выше методы, необходимо дать ответы на вопросы:

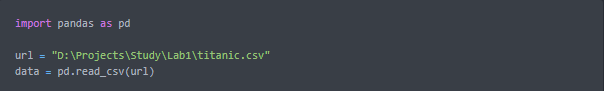
1. Сколько мужчин было на корабле?
2. Какая доля пассажиров (в %) выжила?
3. Какая доля пассажиров (в %) от общего количества путешествовала во 2-ом классе?
4. Посчитайте среднее и медиану возраста всех людей на корабле.
5. Коррелируют ли число братьев/сестер с числом родителей/детей? Посчитайте корреляцию по Пирсону между признаками SibSp и Parch датасета

Решение:

**Шаг 1: Загрузка данных**

Сначала мы загружаем набор данных с помощью функции read\_csv() из библиотеки Pandas.

**Код:**



**Объяснение**:

* pd.read\_csv() загружает данные из CSV-файла в объект DataFrame.
* Набор данных содержит информацию о пассажирах Титаника: их пол, возраст, класс, количество братьев и сестер (SibSp), родителей и детей (Parch), выжили ли они и т.д.

**Шаг 2: Определение количества мужчин на корабле**

Для этого фильтруем набор данных по признаку пола и считаем количество строк, где пол равен 'male'.

**Код:**



**Объяснение**:

* data['Sex'] == 'male' возвращает логический массив (True для мужчин).
* data[data['Sex'] == 'male'] фильтрует набор данных, оставляя только строки с мужчинами.
* .shape[0] возвращает количество строк, т.е. количество мужчин.

**Шаг 3: Подсчет доли выживших пассажиров**

Для подсчета доли выживших используем столбец Survived, где 1 обозначает выживших. Среднее значение этого столбца даст процент выживших.

**Код:**



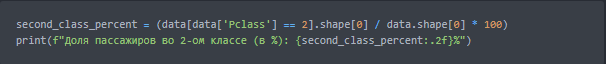
**Объяснение**:

* data['Survived'] содержит значения 0 (не выжил) и 1 (выжил).
* .mean() вычисляет среднее значение, что соответствует доле выживших.

**Шаг 4: Определение доли пассажиров второго класса**

Фильтруем пассажиров по классу (Pclass == 2) и считаем процент этих пассажиров от общего числа.

**Код:**



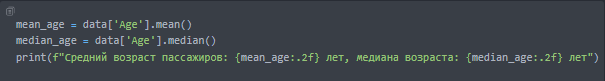
**Объяснение**:

* data['Pclass'] == 2 фильтрует пассажиров второго класса.
* .shape[0] возвращает количество пассажиров второго класса.
* Делим это количество на общее число пассажиров data.shape[0], чтобы получить процент.

**Шаг 5: Подсчет среднего и медианного возраста пассажиров**

Для этого используем методы .mean() и .median() для столбца Age.

**Код:**



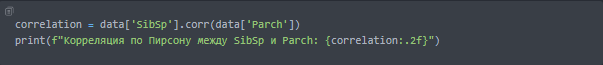
**Объяснение**:

* data['Age'].mean() вычисляет средний возраст.
* data['Age'].median() вычисляет медианный возраст, который менее чувствителен к выбросам.

**Шаг 6: Вычисление корреляции Пирсона между признаками SibSp и Parch**

Для оценки взаимосвязи между количеством братьев/сестер (SibSp) и количеством родителей/детей (Parch) используем метод корреляции Пирсона.

**Код:**



**Объяснение**:

* data['SibSp'].corr(data['Parch']) вычисляет коэффициент корреляции Пирсона, показывающий степень линейной зависимости между двумя признаками.
* Коэффициент варьируется от -1 до 1: значения ближе к 1 указывают на положительную корреляцию, ближе к -1 — на отрицательную, а близкие к 0 означают отсутствие линейной зависимости.

**Шаг 7: Результаты**

1. Количество мужчин на корабле: 577
2. Доля выживших пассажиров (в %): 38.38%
3. Доля пассажиров во 2-ом классе (в %): 20.65%
4. Средний возраст пассажиров: 29.70 лет, медиана возраста: 28.00 лет
5. Корреляция по Пирсону между SibSp и Parch: 0.41