МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М.Є. Жуковського

“Харківський авіаційний інститут”

Кафедра комп’ютерних систем, мереж і кібербезпеки

Лабораторна робота № 4

з дисципліни “ Великі дані для кібербезпеки”

Варіант № \_21\_

|  |  |  |
| --- | --- | --- |
| Виконав студент | гр. 555M | Орлов С.В. |
|  | (№ групи) | (П.І.Б.) |
| (підпис, дата) |  |  |
| Перевірив | д.т.н., професор, професор каф. 503 | |
|  | (науковий ступінь, вчене звання, посада) | |
|  | Фесенко Г. В. | |
| (підпис, дата) | (П.І.Б.) | |

2023

**Тема роботи**

Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: статистичні обчислення і перевірка гіпотез.

**Мета роботи**

Отримати навички статистичних обчислень і перевірок гіпотез з використанням Spark MLlib і пакета spark.ml.

**Завдання**

2.1 Створення локального вектора

Контрольний приклад

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власні вектори

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

2.2 Помічена точка

Контрольний приклад

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власні вектори із 5 елементів та двох нульових елементів

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

3. Створення матриць

Створимо щільну матрицю ((1.0, 2.0), (3.0, 4.0), (5.0, 6.0)):

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власний приклад матриці із 3-х рядків та 4-х стовпців

Зображення, що містить текст, Шрифт, знімок екрана

Автоматично згенерований опис

4. Статистичні обчислення

4.1 Зведені статистичні дані

Зображення, що містить текст, знімок екрана

Автоматично згенерований опис

Власний приклад із 6 векторів по 5 елементів

Зображення, що містить текст, знімок екрана, програмне забезпечення

Автоматично згенерований опис

4.2 Кореляція

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власний приклад, створивши свої серії даних:

RDD [Double] з 10-ти елементів кожної з двох серій;

RDD [Vector], що складаються з 4-х векторів по 5-ть елементів кожен.

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

4.3 Перевірка гіпотез

4.3.1 Використання критерію хі-квадрат Пірсона

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

4.3.2 Реалізація тесту Колмогорова-Смирнова

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власний приклад:

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

**Висновки**

Було отримано навички статистичних обчислень і перевірок гіпотез з використанням Spark MLlib і пакета spark.ml. Відпрацьовані завдання за прикладами із виконанням власних прикладів.