МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М.Є. Жуковського

“Харківський авіаційний інститут”

Кафедра комп’ютерних систем, мереж і кібербезпеки

Лабораторна робота № 5

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 Вилучення ознак

2.1 TF-IDF (від англ. TF – term frequency, IDF – inverse document frequency)

Контрольний приклад

Зображення, що містить текст, знімок екрана, Шрифт, документ

Автоматично згенерований опис

Власні вектори

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

2.2 Word2Vec

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власний приклад

Зображення, що містить текст, знімок екрана, меню, Шрифт

Автоматично згенерований опис

3. Перетворення ознак

3.1 Tokenizer

Зображення, що містить текст, знімок екрана

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

3.2 StopWordsRemover

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власний приклад

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

3.3 n-gram

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Власний приклад

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

3.4 Binarizer

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

3.5 StringIndexer

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

3.6 IndexToString

Зображення, що містить текст, знімок екрана, Шрифт, програмне забезпечення

Автоматично згенерований опис





Зображення, що містить текст, знімок екрана, Шрифт, ряд

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт, число

Автоматично згенерований опис

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.