



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01 Информатика и вычислительная техника

МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/12 Интеллектуальный анализ больших
данных в системах поддержки принятия решений.

О Т Ч Е Т

по лабораторной работе № 10

Вариант № 4

Название: Spark

Дисциплина: языки программирования для работы с большими данными

Студент

ИУ6-23М

(Группа)

(Подпись, дата)

А.А.Клушина

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

П.В. Степанов

(И.О. Фамилия)

Москва, 2024

Цель: ознакомиться с работой Spark в языке программирования Java.

Задание: сделать 10 выборки данных по выбранной предметной области

Был выбран Dataset с пользователями различных социальных сетей и их эмоциями.

Код класса Main:

```
package atomic.engineering
```

```
import org.apache.spark.sql.SparkSession
```

```
//https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotional-well-being
```

```
object Spark {
```

```
    def main(args: Array[String]): Unit = {
```

```
        val spark = SparkSession.builder().master("local[1]").appName("AtomicEngineering").getOrCreate()
```

```
        println("name:" + spark.sparkContext.appName)
```

```
        val path = "src/train.csv"
```

```
        val df = spark.read.option("header", "true").csv(path)
```

```
//        df.show()
```

```
        df.createOrReplaceTempView("data")
```

```
        spark.sql("SELECT * FROM data limit 10").show
```

```

        spark.sql("SELECT  AVG(Messages_Sent_Per_Day)  FROM  data
WHERE Platform = 'Twitter'").show

        spark.sql("SELECT  *  FROM  data  WHERE  Gender = 'Female'
AND Age > 18 AND Age<24").show

        spark.sql("SELECT  Dominant_Emotion  FROM  data  WHERE
Gender = 'Male' AND Posts_Per_Day>5").show

        spark.sql("SELECT  AVG(Likes_Received_Per_Day)  FROM  data
WHERE Platform = 'Instagram' AND Gender = 'Female'").show

        spark.sql("SELECT  AVG(Likes_Received_Per_Day)  FROM  data
WHERE Platform = 'Instagram' AND Posts_Per_Day>1").show

        spark.sql("SELECT  Dominant_Emotion  FROM  data  WHERE
Age>30 AND Posts_Per_Day>1").show

        spark.sql("SELECT  Age  FROM  data  WHERE  Platform  =
'Facebook' AND Dominant_Emotion = 'Boredom'").show

        spark.sql("SELECT  *  FROM  data  WHERE  Platform  =
'Facebook' AND Gender = 'Female' AND Age < 25").show

        spark.sql("SELECT  Age  FROM  data  WHERE  Platform  =
'Instagram' AND Likes_Received_Per_Day > 100").show


        spark.stop()

    }

}

```

Работа программы показана на рисунке 1.

User_ID	Age	Gender	Platform	Daily_Usage_Time (minutes)	Posts_Per_Day	Likes_Received_Per_Day	Comments_Received_Per_Day	Messages_Sent_Per_Day	Dominant_Emotion	
1	97	22	Female	Facebook	70	1	12	5	10	Neutral
1	197	22	Female	Facebook	70	1	14	6	10	Neutral
1	297	22	Female	Facebook	70	1	14	6	10	Neutral
1	397	22	Female	Facebook	70	1	14	6	10	Neutral
1	497	22	Female	Facebook	70	1	14	6	10	Neutral
1	597	22	Female	Facebook	70	1	14	6	10	Neutral
1	697	22	Female	Facebook	70	1	14	6	10	Neutral
1	797	22	Female	Facebook	70	1	14	6	10	Neutral
1	897	22	Female	Facebook	70	1	14	6	10	Neutral
1	997	22	Female	Facebook	70	1	14	6	10	Neutral

Рисунок 1 – Работа программы

Вывод: был изучен Spark в языке программирования Java.