

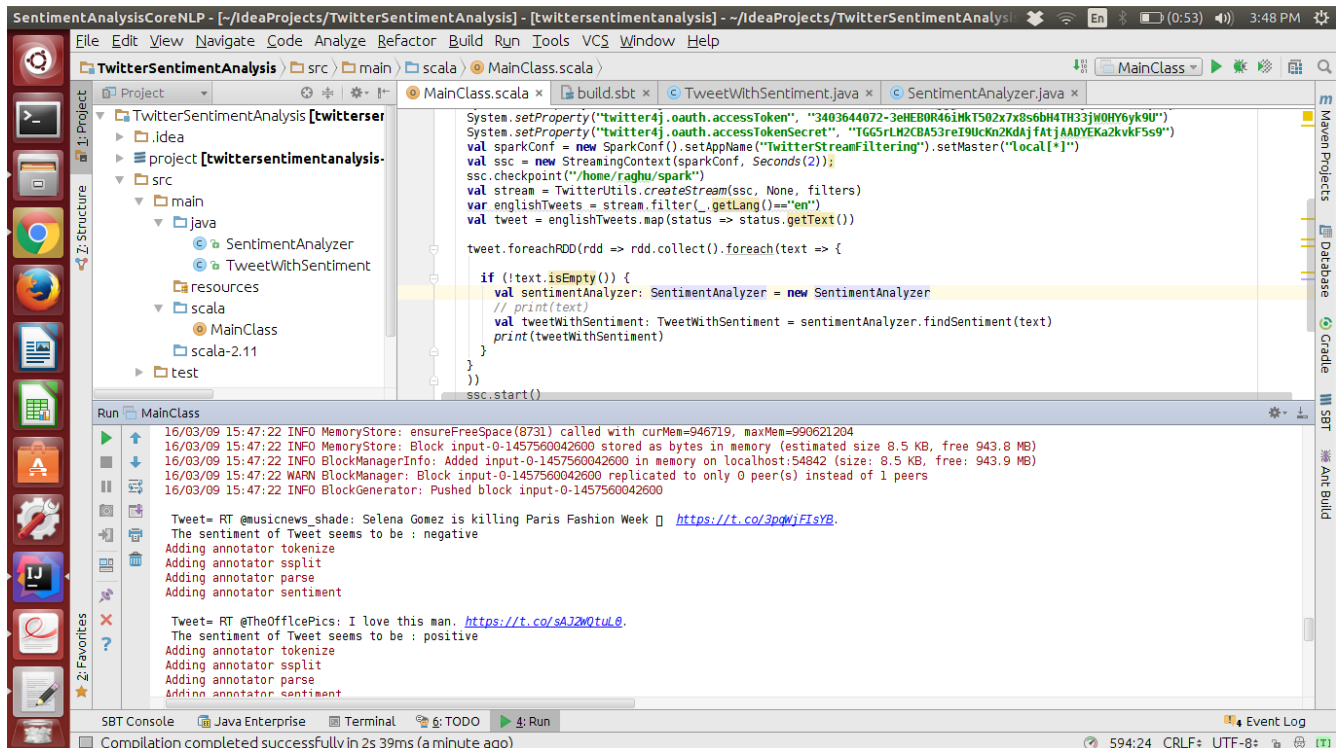
Lab 8 Assignment

Submitted by
Ravi Kiran Yadavalli -13
Ragunandan Rao Malangully -31

Task 1:

The task here was to perform sentiment analysis on Twitter streaming data. I performed the following steps to complete the task.

1. Create Twitter stream using Twitter4j package.
2. Filter the Tweets where the language is English.
3. Extract only the Tweet text from the Tweet.
4. For every Tweet text perform sentiment analysis using the Stanford NLP class.
5. Output the Tweet Text along with the sentiment of the Tweet text.



The screenshot displays an IDE window titled "SentimentAnalysisCoreNLP - [~/IdeaProjects/TwitterSentimentAnalysis] - [twittersentimentanalysis] - ~/IdeaProjects/TwitterSentimentAnalysis". The interface includes a project structure on the left, a central code editor, and a bottom console window.

Project Structure:

- TwitterSentimentAnalysis [twittersentimentanalysis]
 - .idea
 - project [twittersentimentanalysis]
 - src
 - main
 - java
 - SentimentAnalyzer
 - TweetWithSentiment
 - resources
 - scala
 - MainClass
 - test

Code Editor: The file "MainClass.scala" is open, showing the following code:

```
System.setProperty("twitter4j.oauth.accessToken", "3403644072-3eHEB0R46iMkT502x7x8s6bH4TH33jW0HY6yk9U")
System.setProperty("twitter4j.oauth.accessTokenSecret", "IG65rLH2CBAS3reI9Uckn2KdJfATjAADYEKa2kvf5s9")
val sparkConf = new SparkConf().setAppName("TwitterStreamFiltering").setMaster("local[*]")
val ssc = new StreamingContext(sparkConf, Seconds(2))
ssc.checkpoint("/home/raghu/spark")
val stream = TwitterUtils.createStream(ssc, None, filters)
val englishTweets = stream.filter(_.getLang() == "en")
val tweet = englishTweets.map(status => status.getText())

tweet.foreachRDD(rdd => rdd.collect().foreach(text => {
    if (!text.isEmpty()) {
        val sentimentAnalyzer: SentimentAnalyzer = new SentimentAnalyzer
        // print(text)
        val tweetWithSentiment: TweetWithSentiment = sentimentAnalyzer.findSentiment(text)
        print(tweetWithSentiment)
    }
}))
ssc.start()
```

Console Output:

```
16/03/09 15:47:22 INFO MemoryStore: ensureFreeSpace(8731) called with curMem=946719, maxMem=990621204
16/03/09 15:47:22 INFO MemoryStore: Block input-0-1457560042600 stored as bytes in memory (estimated size 8.5 KB, free 943.8 MB)
16/03/09 15:47:22 INFO BlockManagerInfo: Added input-0-1457560042600 in memory on localhost:54842 (size: 8.5 KB, free: 943.9 MB)
16/03/09 15:47:22 WARN BlockManager: Block input-0-1457560042600 replicated to only 0 peer(s) instead of 1 peers
16/03/09 15:47:22 INFO BlockGenerator: Pushed block input-0-1457560042600

Tweet= RT @musicnews_shade: Selena Gomez is killing Paris Fashion Week https://t.co/3pQWjFt5yB.
The sentiment of Tweet seems to be : negative
Adding annotator tokenize
Adding annotator ssplit
Adding annotator parse
Adding annotator sentiment

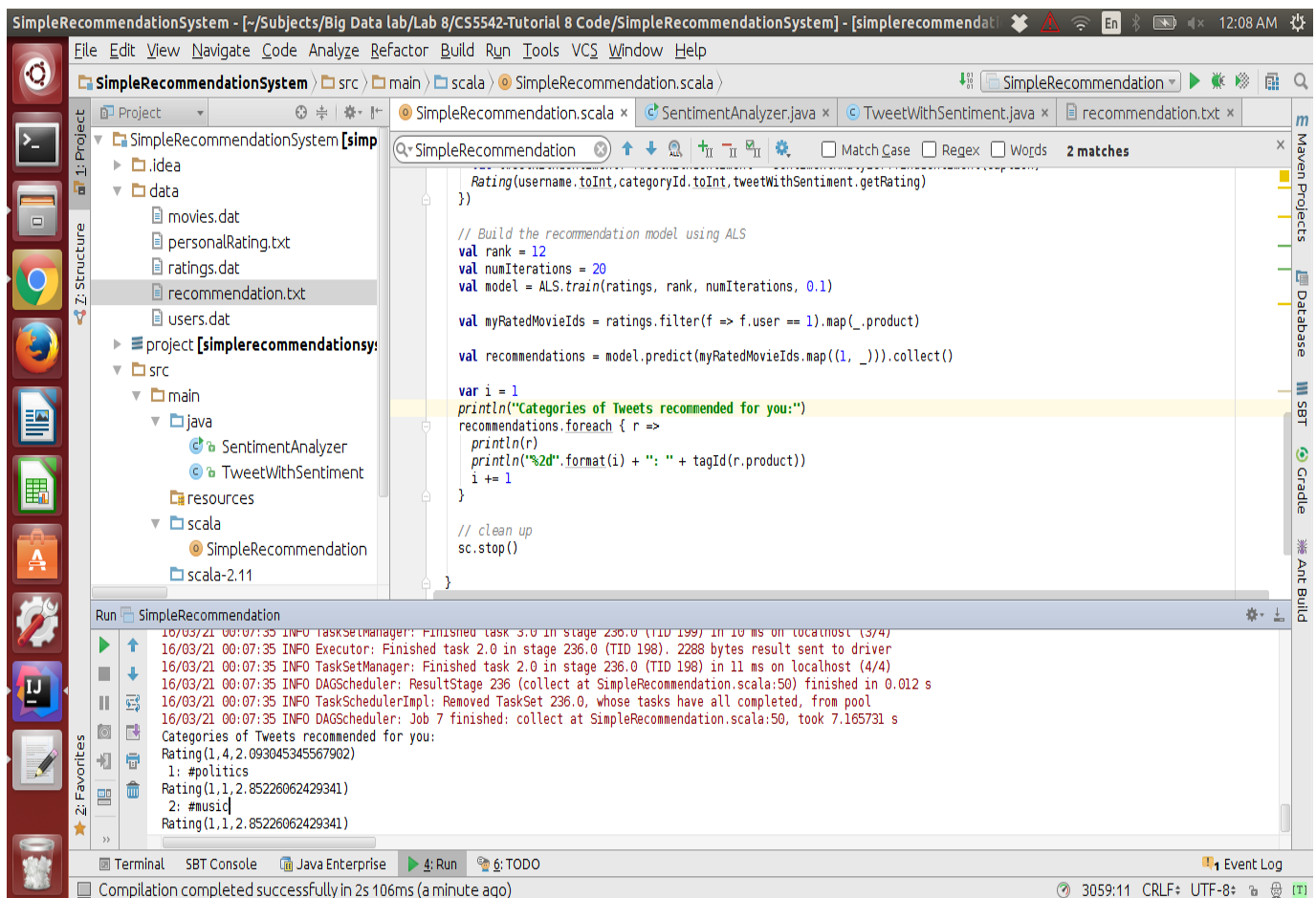
Tweet= RT @TheOfficePics: I love this man. https://t.co/SA72M0tUu0.
The sentiment of Tweet seems to be : positive
Adding annotator tokenize
Adding annotator ssplit
Adding annotator parse
Adding annotator sentiment
```

The status bar at the bottom indicates "Compilation completed successfully in 2s 39ms (a minute ago)" and "594/24 CRLF: UTF-8".

Task 2:

The task here was to create a recommendation system based on Twitter stream. For accomplishing this task the following steps were followed.

1. Create a Twitter Stream to stream Tweets and perform real time classification based on what categories they belonged to.
2. Then part of this data was used to train the same classification algorithm.
3. Live Tweets from Twitter stream was categorized using the classification algorithm.
4. Then the data was stored as the format of `userId:Tweet:Category:CategoryId`.
5. This was used as a training data for the recommendation system.
6. Then recommendations were performed on the data.



```
SimpleRecommendationSystem - [~/Subjects/Big Data lab/Lab 8/CS5542-Tutorial 8 Code/SimpleRecommendationSystem] - [simplerecommendat... 12:08 AM
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SimpleRecommendationSystem > src > main > scala > SimpleRecommendation.scala
SimpleRecommendation.scala x SentimentAnalyzer.java x TweetWithSentiment.java x recommendation.txt x
SimpleRecommendation 2 matches
Rating(username.toInt, categoryID.toInt, tweetWithSentiment.getRating)
})
// Build the recommendation model using ALS
val rank = 12
val numIterations = 20
val model = ALS.train(ratings, rank, numIterations, 0.1)
val myRatedMovieIds = ratings.filter(f => f.user == 1).map(_.product)
val recommendations = model.predict(myRatedMovieIds.map((1, _)).collect())
var i = 1
println("Categories of Tweets recommended for you:")
recommendations.foreach { r =>
  println(r)
  println("%2d".format(i) + ": " + tagId(r.product))
  i += 1
}
// clean up
sc.stop()
}

Run SimpleRecommendation
16/03/21 00:07:35 INFO TaskSetManager: Finished task 3.0 in stage 236.0 (TID 199) in 10 ms on localhost (3/4)
16/03/21 00:07:35 INFO Executor: Finished task 2.0 in stage 236.0 (TID 198). 2288 bytes result sent to driver
16/03/21 00:07:35 INFO TaskSetManager: Finished task 2.0 in stage 236.0 (TID 198) in 11 ms on localhost (4/4)
16/03/21 00:07:35 INFO DAGScheduler: ResultStage 236 (collect at SimpleRecommendation.scala:50) finished in 0.012 s
16/03/21 00:07:35 INFO TaskSchedulerImpl: Removed TaskSet 236.0, whose tasks have all completed, from pool
16/03/21 00:07:35 INFO DAGScheduler: Job 7 finished: collect at SimpleRecommendation.scala:50, took 7.165731 s
Categories of Tweets recommended for you:
Rating(1,4,2.093045345567902)
1: #politics
Rating(1,1,2.85226062429341)
2: #music
Rating(1,1,2.85226062429341)

Terminal SBT Console Java Enterprise Run TODO
Compilation completed successfully in 2s 106ms (a minute ago) 3059:11 CRLF+ UTF-8+
```