

## Question 1:

### Twitter Sentimental Analysis:

Sentimental Analysis on the given data(static). Based on the data given, the result will be displayed as positive or negative.

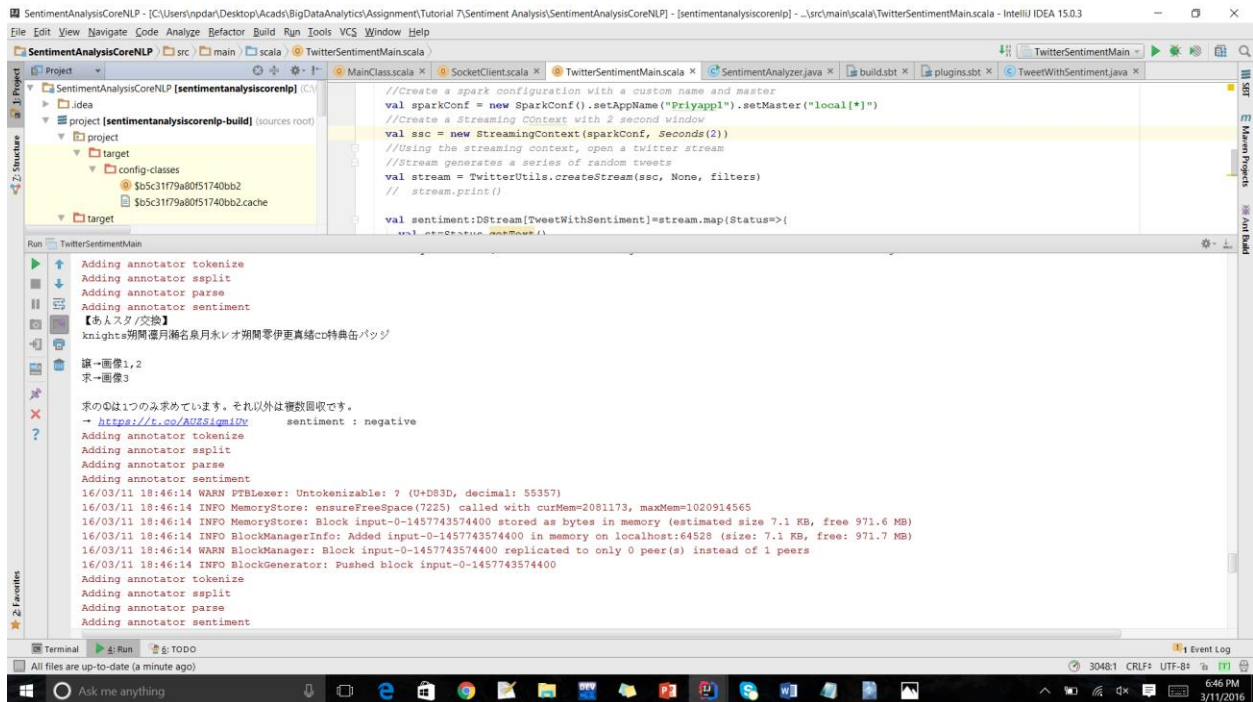
```
object MainClass {  
  def main(args: Array[String]) {  
    val sentimentAnalyzer: SentimentAnalyzer = new SentimentAnalyzer  
    val tweetWithSentiment: TweetWithSentiment = sentimentAnalyzer.findSentiment("I hate running out of time.")  
    System.out.println(tweetWithSentiment)  
  }  
}
```

Run MainClass  
"C:\Program Files\Java\jdk1.8.0\_65\bin\java" ...  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Loading parser from serialized file edu/stanford/nlp/models/lexparser/englishPCFG.ser.gz ... done [2.2 sec].  
Adding annotator sentiment  
TweetWithSentiment [line=I hate running out of time., cssClass=sentiment : negative]  
Process finished with exit code 0

```
object MainClass {  
  def main(args: Array[String]) {  
    val sentimentAnalyzer: SentimentAnalyzer = new SentimentAnalyzer  
    val tweetWithSentiment: TweetWithSentiment = sentimentAnalyzer.findSentiment("I love Scala.")  
    System.out.println(tweetWithSentiment)  
  }  
}
```

Run MainClass  
"C:\Program Files\Java\jdk1.8.0\_65\bin\java" ...  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Loading parser from serialized file edu/stanford/nlp/models/lexparser/englishPCFG.ser.gz ... done [2.0 sec].  
Adding annotator sentiment  
TweetWithSentiment [line=I love Scala., cssClass=sentiment : positive]  
Process finished with exit code 0

In this, I am able to provide the twitter live streaming data as an input. Here the sentimental analysis is taking place on streaming data and is able to analyze the data to be positive or negative.



The screenshot shows the IntelliJ IDEA interface with the following components:

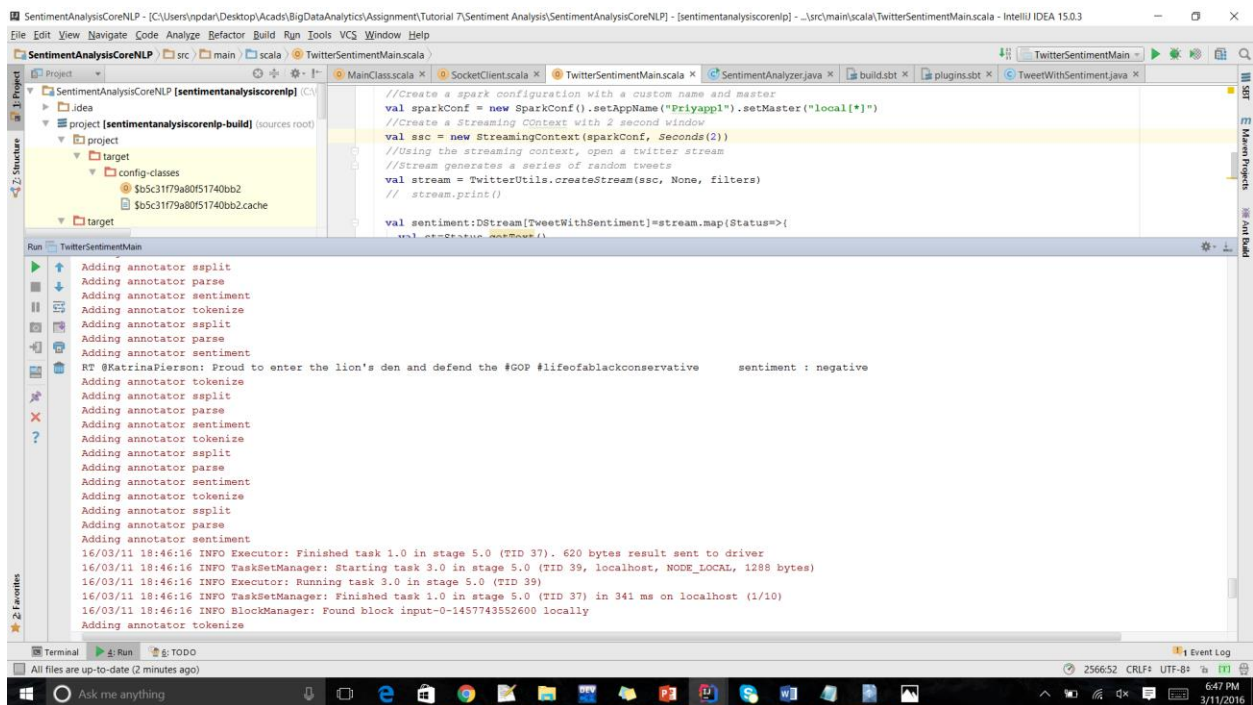
- Project Structure:** The project is named "SentimentAnalysisCoreNLP". It contains a "src" directory with "main" and "scala" subdirectories. The "main" directory contains "MainClass.scala" and "SocketClient.scala". The "scala" directory contains "TwitterSentimentMain.scala".
- Code Editor:** The "TwitterSentimentMain.scala" file is open. It contains the following code:

```
//Create a spark configuration with a custom name and master
val sparkConf = new SparkConf().setAppName("Priyap1").setMaster("local[*]")
//Create a Streaming Context with 2 second window
val ssc = new StreamingContext(sparkConf, Seconds(2))
//Using the streaming context, open a twitter stream
//Stream generates a series of random tweets
val stream = TwitterUtils.createStream(ssc, None, filters)
// stream.print()

val sentiment:DStream[TweetWithSentiment]=stream.map(Status=>{
  val status = new Tweet(status.getText(), status.getCreatedAt())
  sentiment = sentiment.union(stream.map(status => {
    val sentiment = sentimentAnalyzer.analyze(status.getText())
    new Tweet(status.getText(), status.getCreatedAt(), sentiment)
  }))
})
```
- Run Console:** The console shows the execution of the application. It displays the addition of various annotators and the processing of a tweet. The output is as follows:

```
Adding annotator tokenize
Adding annotator split
Adding annotator parse
Adding annotator sentiment
【あ人スタ/交換】
knights翔風温月蘭名泉月永レオ期間零伊更真緒cd特典缶パンジ
返→画像1,2
求→画像3

求の0は1つのみ求めています。それ以外は複数取収です。
→ https://t.co/AU8S4gm4Dc sentiment : negative
Adding annotator tokenize
Adding annotator split
Adding annotator parse
Adding annotator sentiment
16/03/11 18:46:14 WARN PTBlexer: Untokenizable: ? (U+DB3D, decimal: 55357)
16/03/11 18:46:14 INFO MemoryStore: ensureFreeSpace(7225) called with curMem=1020914565
16/03/11 18:46:14 INFO MemoryStore: Block input-0-1457743574400 stored as bytes in memory (estimated size 7.1 KB, free 971.6 MB)
16/03/11 18:46:14 INFO BlockManagerInfo: Added input-0-1457743574400 in memory on localhost:64528 (size: 7.1 KB, free: 971.7 MB)
16/03/11 18:46:14 WARN BlockManager: Block input-0-1457743574400 replicated to only 0 peer(s) instead of 1 peers
16/03/11 18:46:14 INFO BlockGenerator: Pushed block input-0-1457743574400
Adding annotator tokenize
Adding annotator split
Adding annotator parse
Adding annotator sentiment
```



The screenshot shows the IntelliJ IDEA interface with the following components:

- Project Structure:** The project is named "SentimentAnalysisCoreNLP". It contains a "src" directory with "main" and "scala" subdirectories. The "main" directory contains "MainClass.scala" and "SocketClient.scala". The "scala" directory contains "TwitterSentimentMain.scala".
- Code Editor:** The "TwitterSentimentMain.scala" file is open. It contains the following code:

```
//Create a spark configuration with a custom name and master
val sparkConf = new SparkConf().setAppName("Priyap1").setMaster("local[*]")
//Create a Streaming Context with 2 second window
val ssc = new StreamingContext(sparkConf, Seconds(2))
//Using the streaming context, open a twitter stream
//Stream generates a series of random tweets
val stream = TwitterUtils.createStream(ssc, None, filters)
// stream.print()

val sentiment:DStream[TweetWithSentiment]=stream.map(Status=>{
  val status = new Tweet(status.getText(), status.getCreatedAt())
  sentiment = sentiment.union(stream.map(status => {
    val sentiment = sentimentAnalyzer.analyze(status.getText())
    new Tweet(status.getText(), status.getCreatedAt(), sentiment)
  }))
})
```
- Run Console:** The console shows the execution of the application. It displays the addition of various annotators and the processing of a tweet. The output is as follows:

```
Adding annotator split
Adding annotator parse
Adding annotator sentiment
Adding annotator tokenize
Adding annotator split
Adding annotator parse
Adding annotator sentiment
RT @fatimaPierson: Proud to enter the lion's den and defend the #GOP #lifeofblackconservative sentiment : negative
Adding annotator tokenize
Adding annotator split
Adding annotator parse
Adding annotator sentiment
Adding annotator tokenize
Adding annotator split
Adding annotator parse
Adding annotator sentiment
16/03/11 18:46:16 INFO Executor: Finished task 1.0 in stage 5.0 (TID 37). 620 bytes result sent to driver
16/03/11 18:46:16 INFO TaskSetManager: Starting task 3.0 in stage 5.0 (TID 39, localhost, NODE_LOCAL, 1288 bytes)
16/03/11 18:46:16 INFO Executor: Running task 3.0 in stage 5.0 (TID 39)
16/03/11 18:46:16 INFO TaskSetManager: Finished task 1.0 in stage 5.0 (TID 37) in 341 ms on localhost (1/10)
16/03/11 18:46:16 INFO BlockManager: Found block input-0-1457743552600 locally
Adding annotator tokenize
```

SentimentAnalysisCoreNLP - [C:\Users\npdar\Desktop\Acad\BigDataAnalytics\Assignment\Tutorial 7\Sentiment Analysis\SentimentAnalysisCoreNLP] - [sentimentanalysiscorenlp] - ...src\main\scala\TwitterSentimentMain.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SentimentAnalysisCoreNLP [src] [main] [scala] [TwitterSentimentMain.scala]

Project: SentimentAnalysisCoreNLP [sentimentanalysiscorenlp] (C:\...)  
Idea  
MainClass.scala X SocketClient.scala X TwitterSentimentMain.scala X SentimentAnalyzer.java X build.sbt X plugins.sbt X TweetWithSentiment.java X

Run: TwitterSentimentMain

```
//Create a spark configuration with a custom name and master
val sparkConf = new SparkConf().setAppName("Priyapri").setMaster("local[*]")
//Create a Streaming Context with 2 second window
```

16/03/11 18:46:16 INFO BlockGenerator: Pushed block input-0-1457743576400  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
I really enjoyed trying out this blush stick by @testique, I like the fact that you dont need a brush to apply :) <https://t.co/2Faob6nyKt> sentiment : positive  
16/03/11 18:46:16 INFO MemoryStore: ensureFreeSpace(6356) called with curMem=2082453, maxMem=1020914565  
16/03/11 18:46:16 INFO MemoryStore: Block input-0-1457743576600 stored as bytes in memory (estimated size 6.2 KB, free 971.6 MB)  
16/03/11 18:46:16 INFO BlockManagerInfo: Added input-0-1457743576600 in memory on localhost:64528 (size: 6.2 KB, free: 971.7 MB)  
16/03/11 18:46:16 WARN BlockManager: Block input-0-1457743576600 replicated to only 0 peer(s) instead of 1 peers  
16/03/11 18:46:16 INFO BlockGenerator: Pushed block input-0-1457743576600  
16/03/11 18:46:16 INFO BlockManagerInfo: Removed broadcast\_4\_piece0 on localhost:64528 in memory (size: 1051.0 B, free: 971.7 MB)  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
RT @officialMonstax: [7월] two you의 베레 커피가 공개됐습니다~~ 보아드릴 수 있어 좋아요~ 우리 승우할아버지 항상 행복되길테니까 건강하세요~ <https://t.co/18Qqr5huqc> sentiment : negative  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment

Terminal Run TODO  
All files are up-to-date (2 minutes ago)

256652 CRLF UTF-8 6:48 PM 3/11/2016

SentimentAnalysisCoreNLP - [C:\Users\npdar\Desktop\Acad\BigDataAnalytics\Assignment\Tutorial 7\Sentiment Analysis\SentimentAnalysisCoreNLP] - [sentimentanalysiscorenlp] - ...src\main\scala\TwitterSentimentMain.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SentimentAnalysisCoreNLP [src] [main] [scala] [TwitterSentimentMain.scala]

Project: SentimentAnalysisCoreNLP [sentimentanalysiscorenlp] (C:\...)  
MainClass.scala X SocketClient.scala X TwitterSentimentMain.scala X SentimentAnalyzer.java X build.sbt X plugins.sbt X TweetWithSentiment.java X

Run: TwitterSentimentMain

```
//Create a spark configuration with a custom name and master
val sparkConf = new SparkConf().setAppName("Priyapri").setMaster("local[*]")
//Create a Streaming Context with 2 second window
```

Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
16/03/11 18:46:17 WARN PTBTokenizer: Untokenizable: ? (U+D83D, decimal: 55357)  
16/03/11 18:46:17 INFO MemoryStore: ensureFreeSpace(18518) called with curMem=2152339, maxMem=1020914565  
16/03/11 18:46:17 INFO MemoryStore: Block input-0-1457743577600 stored as bytes in memory (estimated size 2.5 KB, free 971.6 MB)  
16/03/11 18:46:17 INFO BlockManagerInfo: Added input-0-1457743577600 in memory on localhost:64528 (size: 2.5 KB, free: 971.6 MB)  
16/03/11 18:46:17 WARN BlockManager: Block input-0-1457743577600 replicated to only 0 peer(s) instead of 1 peers  
16/03/11 18:46:17 INFO BlockGenerator: Pushed block input-0-1457743577600  
RT @rt00004: The #MGM loves to trot out partisan party hacks to masquerade as objective political analysts #MichiganPrimary <https://t.co/y...> sentiment : negative  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
16/03/11 18:46:17 WARN PTBTokenizer: Untokenizable: ? (U+D83D, decimal: 55357)  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
16/03/11 18:46:17 WARN PTBTokenizer: Untokenizable: ? (U+D83C, decimal: 55356)  
Adding annotator tokenize  
Adding annotator split  
Adding annotator parse  
Adding annotator sentiment  
16/03/11 18:46:17 WARN PTBTokenizer: Untokenizable: ? (U+D83D, decimal: 55357)  
16/03/11 18:46:18 INFO MemoryStore: ensureFreeSpace(18940) called with curMem=2154857, maxMem=1020914565  
RT @ufc: UFC employees get a sneak preview of #EASPORTS UFC2! Check out all the action on Snapchat! <https://t.co/bKasPFnuPp> sentiment : negative  
16/03/11 18:46:18 INFO MemoryStore: Block input-0-1457743577800 stored as bytes in memory (estimated size 18.5 KB, free 971.5 MB)  
16/03/11 18:46:18 INFO BlockManagerInfo: Added input-0-1457743577800 in memory on localhost:64528 (size: 18.5 KB, free: 971.6 MB)  
16/03/11 18:46:18 WARN BlockManager: Block input-0-1457743577800 replicated to only 0 peer(s) instead of 1 peers  
16/03/11 18:46:18 INFO BlockGenerator: Pushed block input-0-1457743577800  
Adding annotator tokenize  
Adding annotator split

Terminal Run TODO  
All files are up-to-date (10 minutes ago)

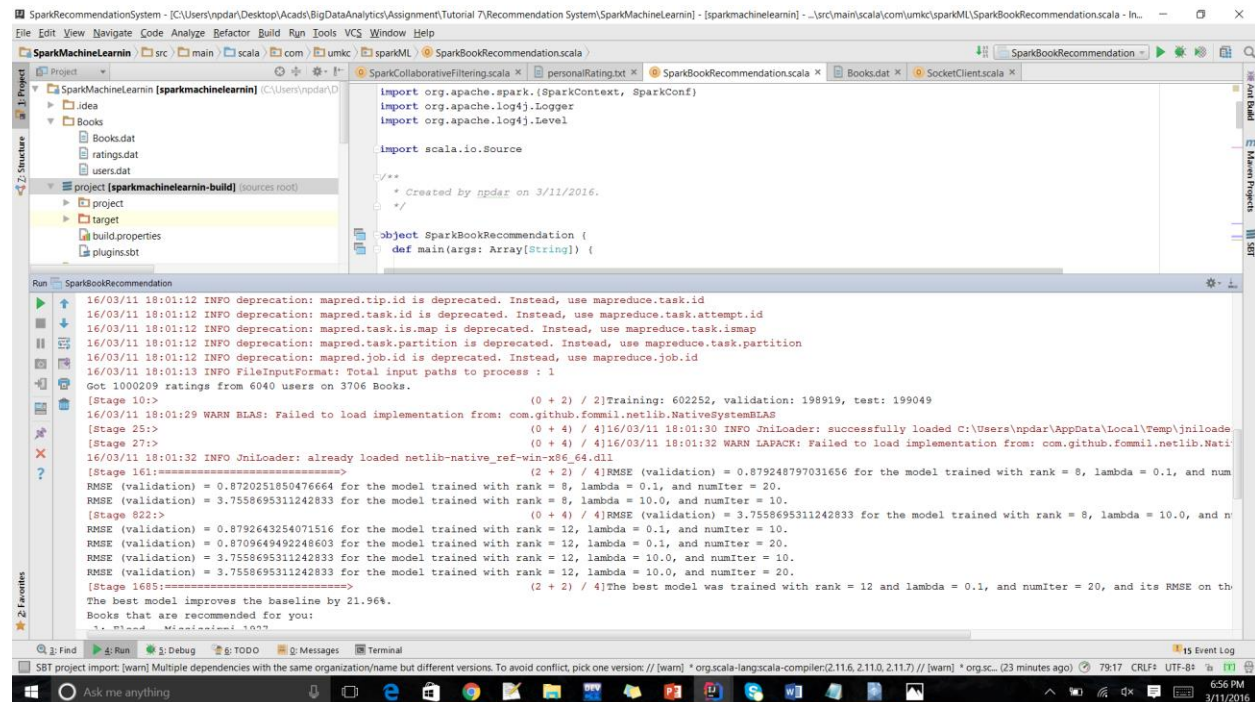
30481 CRLF UTF-8 6:55 PM 3/11/2016



## Question 2:

### Recommendation System

In this I am able to provide the details of Books with the user id, ratings etc. My recommendation system will analyze the results based on the training data and will recommend the books.



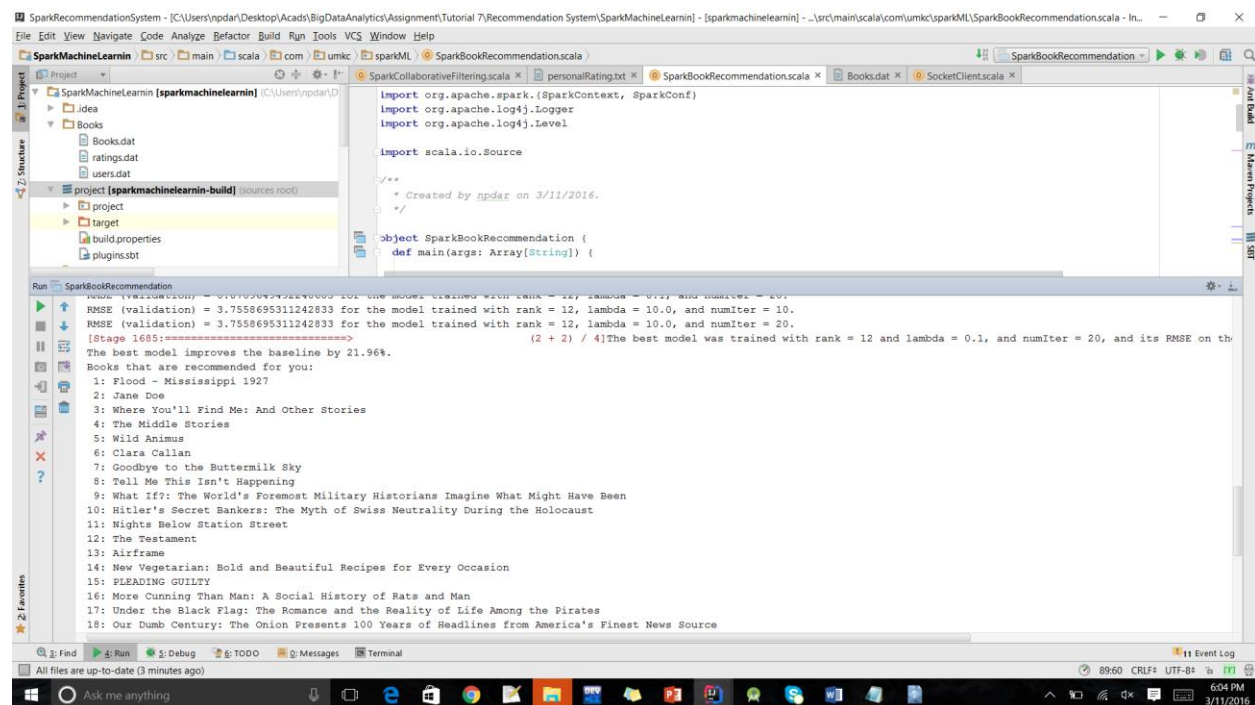
```
import org.apache.spark.{SparkContext, SparkConf}
import org.apache.log4j.Logger
import org.apache.log4j.Level

import scala.io.Source

/**
 * Created by npdar on 3/11/2016.
 */

object SparkBookRecommendation {
  def main(args: Array[String]) {
```

16/03/11 18:01:12 INFO deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id  
16/03/11 18:01:12 INFO deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id  
16/03/11 18:01:12 INFO deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap  
16/03/11 18:01:12 INFO deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition  
16/03/11 18:01:12 INFO deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id  
16/03/11 18:01:13 INFO FileInputFormat: Total input paths to process : 1  
Got 1000209 ratings from 6040 users on 3706 Books.  
[Stage 10:]>  
16/03/11 18:01:29 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS  
[Stage 25:]>  
16/03/11 18:01:32 INFO JniLoader: already loaded netlib-native\_ref-win-x86\_64.dll  
[Stage 161:]>  
RMSE (validation) = 0.8720251850476664 for the model trained with rank = 8, lambda = 0.1, and numIter = 20.  
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 8, lambda = 10.0, and numIter = 10.  
RMSE (validation) = 0.8792643254071516 for the model trained with rank = 12, lambda = 0.1, and numIter = 10.  
RMSE (validation) = 0.870964942248603 for the model trained with rank = 12, lambda = 0.1, and numIter = 20.  
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 10.  
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 20.  
[Stage 165:]>  
The best model improves the baseline by 21.96%.  
Books that are recommended for you:  
1: Flood - Mississippi 1927



```
import org.apache.spark.{SparkContext, SparkConf}
import org.apache.log4j.Logger
import org.apache.log4j.Level

import scala.io.Source

/**
 * Created by npdar on 3/11/2016.
 */

object SparkBookRecommendation {
  def main(args: Array[String]) {
```

RMSE (validation) = 0.870964942248603 for the model trained with rank = 12, lambda = 0.1, and numIter = 20.  
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 10.  
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 20.  
[Stage 165:]>  
The best model improves the baseline by 21.96%.  
Books that are recommended for you:  
1: Flood - Mississippi 1927  
2: Jane Doe  
3: Where You'll Find Me: And Other Stories  
4: The Middle Stories  
5: Wild Animus  
6: Clara Callan  
7: Goodbye to the Buttermilk Sky  
8: Tell Me This Isn't Happening  
9: What If?: The World's Foremost Military Historians Imagine What Might Have Been  
10: Hitler's Secret Bankers: The Myth of Swiss Neutrality During the Holocaust  
11: Nights Below Station Street  
12: The Testament  
13: Airframe  
14: New Vegetarian: Bold and Beautiful Recipes for Every Occasion  
15: PLEADING GUILTY  
16: More Cunning Than Man: A Social History of Rats and Man  
17: Under the Black Flag: The Romance and the Reality of Life Among the Pirates  
18: Our Dumb Century: The Onion Presents 100 Years of Headlines from America's Finest News Source

### Question 3:

This question is related to the question-1. In this the data collected from the twitter will be analyzed using sentimental analysis and will be sent to any android device such as watch/ mobile in the form of notification.

1:09 PM

Friday, Mar 11, 2016



Clear

Lycamobile



Twitter Trend notification

1:08 PM

Performing sentiment analysis



USB debugging connected

Touch to disable USB debugging.



Connected as a media device

Touch for other USB options.



1:08 PM

Hello World!

NOTIFY WEARABLE

I'm waiting here: 1234

SiteLocalAddress: 10.205.1.253



74%



1:11 PM

Hello World!

NOTIFY WEARABLE

I'm waiting here: 1234

SiteLocalAddress: 10.205.1.253

#1 from /10.151.1.25:52868

TweetWithSentiment [line=RT @CassieLeuffen: TRUMP: "Clinton, boy is she being nice to Obama, she agrees w/ him on everything. We can't have another 4 years.. She's ..., cssClass=sentiment : negative]replayed: Hello from Android, you are #1

#2 from /10.151.1.25:52869

TweetWithSentiment [line=RT @CassieLeuffen: TRUMP: "Clinton, boy is she being nice to Obama, she agrees w/ him on everything. We can't have another 4 years.. She's ..., cssClass=sentiment : negative]replayed: Hello from Android, you are #2

#3 from /10.151.1.25:52870

TweetWithSentiment [line=Bernie Sanders wins big with Michigan Muslims <https://t.co/P8SAz7wOBn>, cssClass=sentiment : positive]replayed: Hello from Android, you are #3

#4 from /10.151.1.25:52871

TweetWithSentiment [line=RT @BonnieMadden: I'm amazed at the effort the GOP is making against the people's chosen candidate. Don't let the establishment win. VOTE T..., cssClass=sentiment : negative]replayed: Hello from Android, you are #4

#5 from /10.151.1.25:52872

TweetWithSentiment [line=RT @politiCOHEN\_: a roundup of violence toward protestors at trump events. hint: it's a long roundup. <https://t.co/NS3ilX7Ppf>, cssClass=sentiment : negative]replayed:



I'm waiting here: /::1234

SiteLocalAddress: 10.205.1.253

TweetWithSentiment [line=RT @larrymadill: Ben Carson endorsing Donald Trump is like that moment in Star Wars when Jar Jar Binks nominates Senator Palpatine., cssClass=sentiment :

negative]TweetWithSentiment [line=RT @horoscoponegro: #LEO: a veces, crees demasiado en la gente que te rodea, y debes tener cuidado. No todo el mundo que te "admira y halag...,

cssClass=sentiment : negative]TweetWithSentiment [line=Oscar frågar sig om man verkligen behöver bygga sin egen backend? Han visar att det finns andra möjligheter. <https://t.co/3Vhe6aWmF2>,

cssClass=sentiment : positive]TweetWithSentiment [line=RT @vdare: The endless triangulation by the Beltway Right is tiresome. Calling Trump a "scam" is pure projection. <https://t.co/jFklX6do1F>,

cssClass=sentiment : negative]TweetWithSentiment [line=RT @WriterUSA: #Ohio leads in jobs lost to Clinton's NAFTA <https://t.co/dsvMNTqPu5> #p2 #feelthebern #ohprimary #ohleg #1u <https://t.co/6Wz...>, cssClass=sentiment :

negative]TweetWithSentiment [line=RT @DiarioTalCual: TIENEN LA TAREA ATRASADA. Oscar Figuera: GPP no tiene plan para enfrenar el referéndum <https://t.co/6ayB7wwimz> <https://...>,

cssClass=sentiment : positive]TweetWithSentiment [line=RT @AreZeroZeroEm: I can definitely respect @RealBenCarson for endorsing @realDonaldTrump for the 2016 presidential election.

#MakeAmericaG..., cssClass=sentiment : positive]TweetWithSentiment [line=This hot dog shop is mocking Donald Trump in the best way