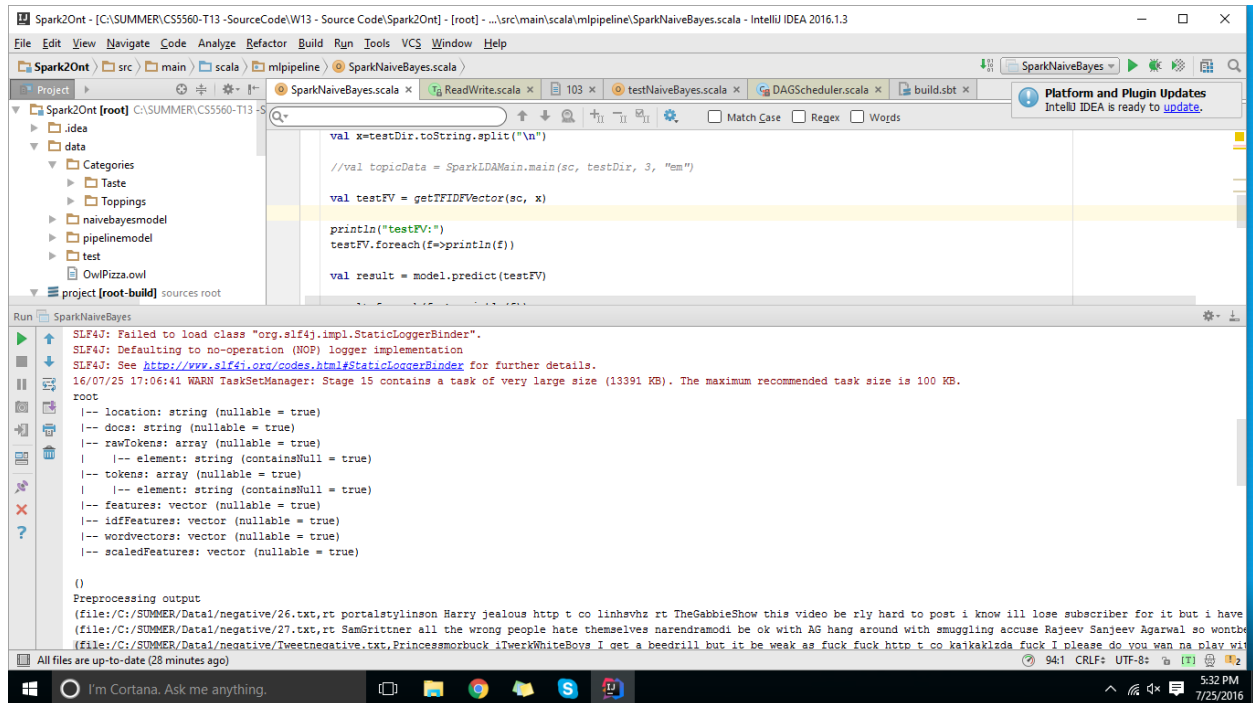


## Naive Bayes Model OUTPUT

The below images refers to the training data output. We are here first training the model by giving it both positive tweets and negative tweets. Once we execute the program we will get the below output.



```
val x=testDir.toString.split("\n")

//val topicData = SparkLDMain.main(sc, testDir, 3, "em")

val testFV = getTFIDFVector(sc, x)

println("testFV:")
testFV.foreach(f=>println(f))

val result = model.predict(testFV)
```

Run SparkNaiveBayes

SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".  
SLF4J: Defaulting to no-operation (NOP) logger implementation  
SLF4J: See <http://www.slf4j.org/codes.html#StaticLoggerBinder> for further details.  
16/07/25 17:06:41 WARN TaskSetManager: Stage 15 contains a task of very large size (13391 KB). The maximum recommended task size is 100 KB.

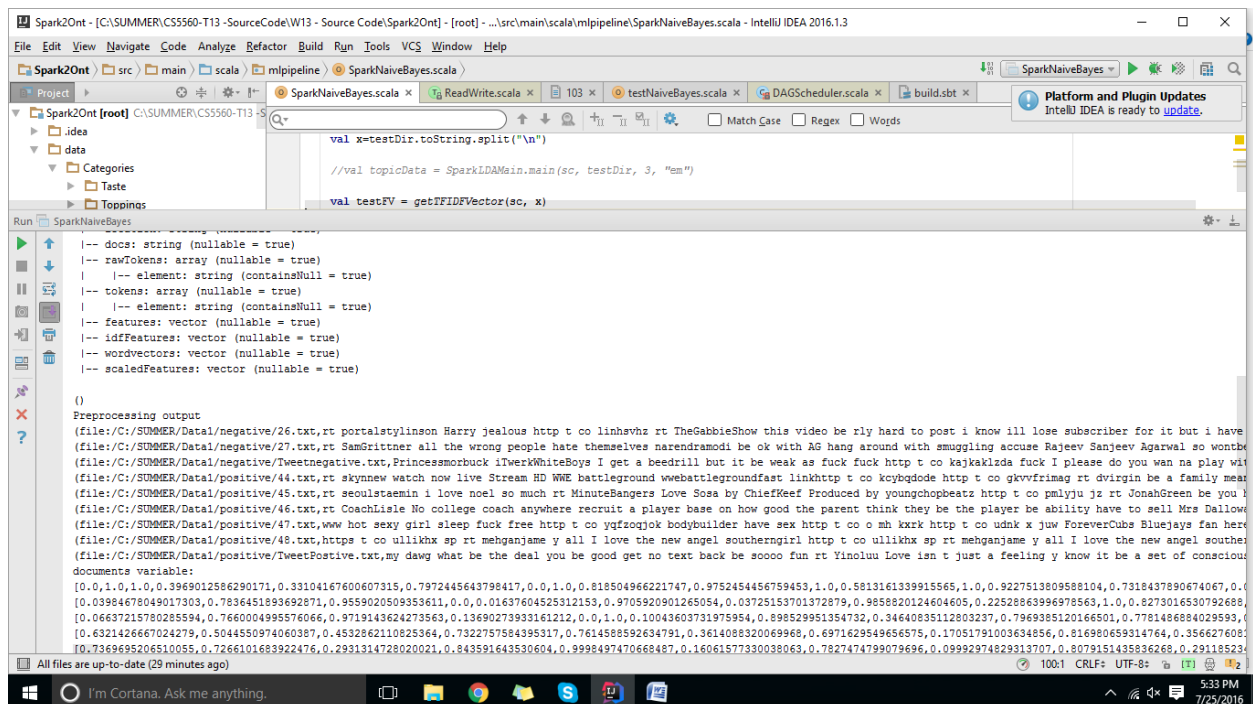
root

- location: string (nullable = true)
- docs: string (nullable = true)
- rawTokens: array (nullable = true)
  - element: string (containsNull = true)
- tokens: array (nullable = true)
  - element: string (containsNull = true)
- features: vector (nullable = true)
- idfFeatures: vector (nullable = true)
- wordVectors: vector (nullable = true)
- scaledFeatures: vector (nullable = true)

()

Preprocessing output

(file:/C:/SUMMER/Data/negative/26.txt,rt portalstylinson Harry jealous http t co linshvz rt TheGabbieShow this video be rly hard to post i know ill lose subscriber for it but i have  
(file:/C:/SUMMER/Data/negative/27.txt,rt SamGrittner all the wrong people hate themselves narendramodi be ok with AG hang around with smuggling accuse Rajeev Sanjeev Agarwal so wontbe  
(file:/C:/SUMMER/Data/negative/Tweetnegative.txt,Princesasmoribuck iTwerkWhiteBoys I get a beedrill but it be weak as fuck fuck http t co kajaklida fuck I please do you wan na play wit



```
val x=testDir.toString.split("\n")

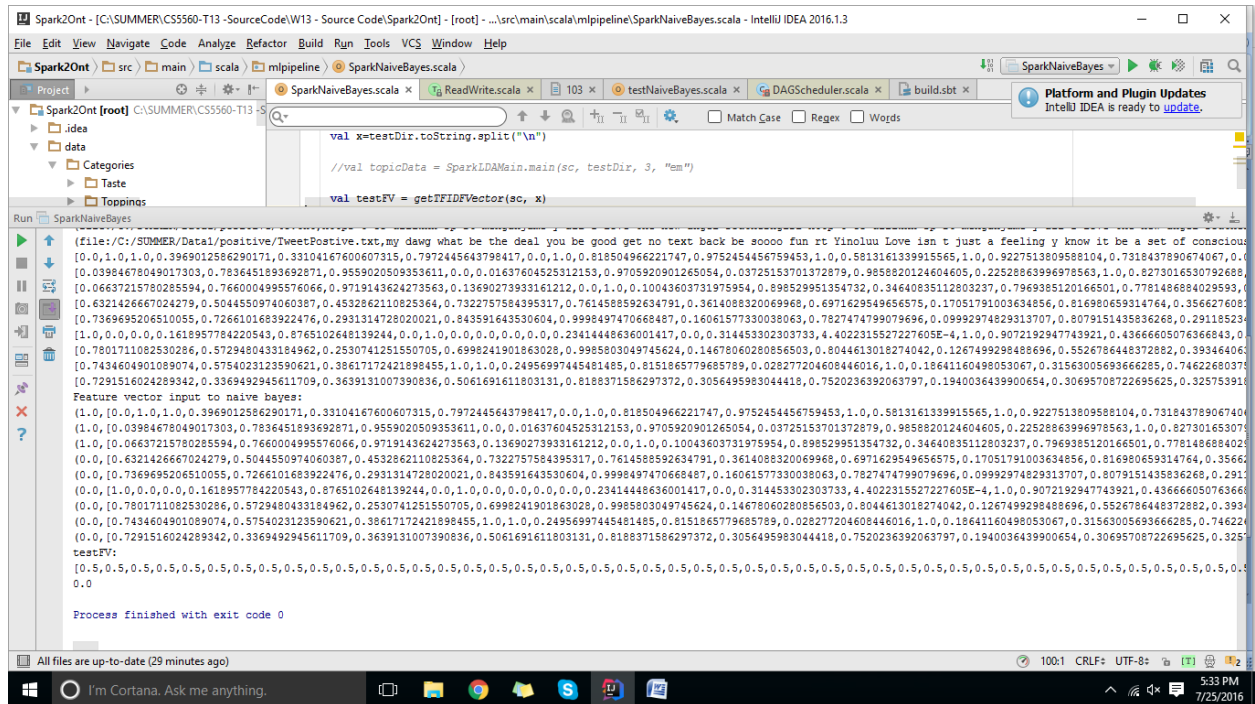
//val topicData = SparkLDMain.main(sc, testDir, 3, "em")

val testFV = getTFIDFVector(sc, x)
```

Run SparkNaiveBayes

Preprocessing output

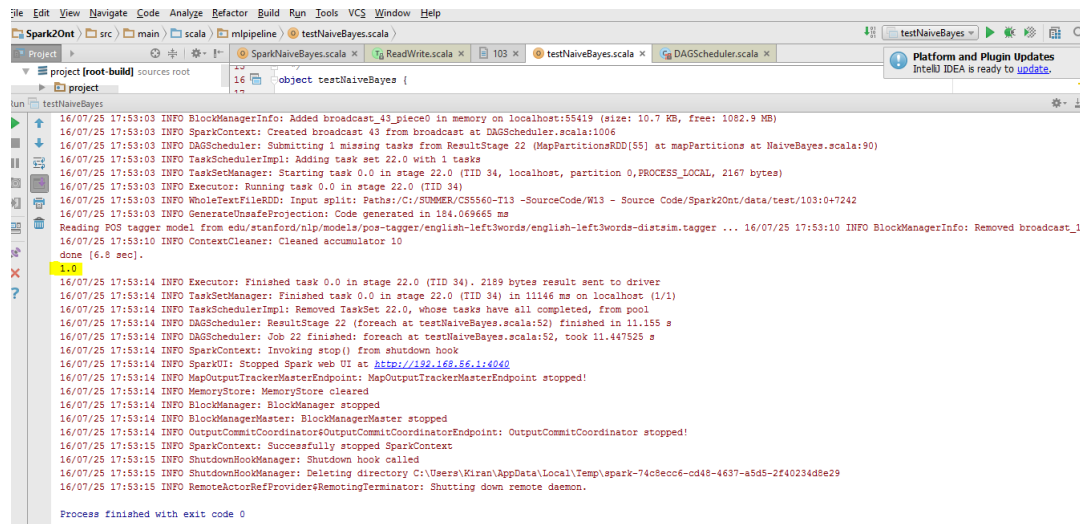
(file:/C:/SUMMER/Data/negative/26.txt,rt portalstylinson Harry jealous http t co linshvz rt TheGabbieShow this video be rly hard to post i know ill lose subscriber for it but i have  
(file:/C:/SUMMER/Data/negative/27.txt,rt SamGrittner all the wrong people hate themselves narendramodi be ok with AG hang around with smuggling accuse Rajeev Sanjeev Agarwal so wontbe  
(file:/C:/SUMMER/Data/negative/Tweetnegative.txt,Princesasmoribuck iTwerkWhiteBoys I get a beedrill but it be weak as fuck fuck http t co kajaklida fuck I please do you wan na play wit  
(file:/C:/SUMMER/Data/positive/44.txt,rt skynnew watch now live Stream HD WME battleground weebattlegroundfast linkhttp t co koybdode http t co gkvvfrimag rt dvirgin be a family near  
(file:/C:/SUMMER/Data/positive/46.txt,rt CoachLisla No college coach anywhere recruit a player base on how good the parent think they be the player be ability have to sell Mrs Dalloway  
(file:/C:/SUMMER/Data/positive/47.txt,www hot sexy girl sleep fuck free http t co yqfzqjok bodybuilder have sex http t co o mh kxrk http t co udnk x juw ForeverCubs Bluejays fan here  
(file:/C:/SUMMER/Data/positive/48.txt,htps t co ullikxh sp rt mehganjame y all i love the new angel southerngirl http t co ullikxh sp rt mehganjame y all i love the new angel southern  
(file:/C:/SUMMER/Data/positive/TweetPositive.txt,my dawg what be the deal you be good get no text back be soooo fun rt Yinoluu Love isn t just a feeling y know it be a set of conscious documents variable:



After training this model, we have checked the working of this model by giving the positive tweets and negative tweets as input. It worked perfectly. Below image is the example of Negative tweets. Here we are getting 1.0 as output which tells us the file contains negative tweets.

## Output :

When given Negative tweets as input

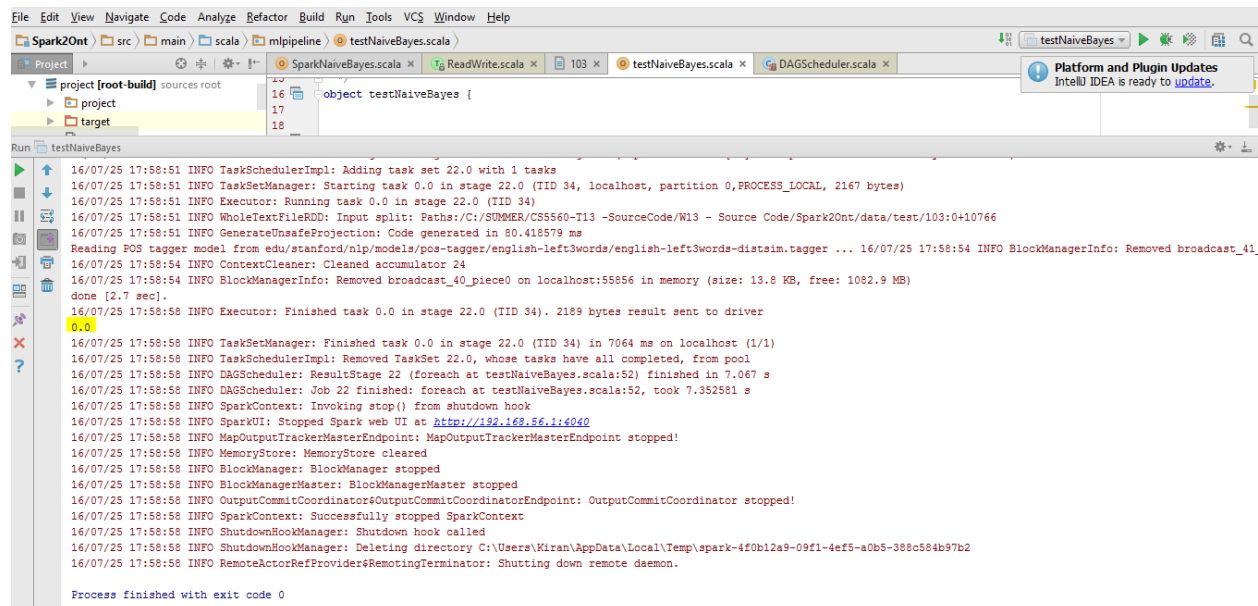


```
16/07/25 17:53:03 INFO BlockManagerInfo: Added broadcast_43_piece0 in memory on localhost:55419 (size: 10.7 KB, free: 1082.9 MB)
16/07/25 17:53:03 INFO SparkContext: Created broadcast 43 from broadcast at DAGScheduler.scala:1006
16/07/25 17:53:03 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 22 (MapPartitionsRDD[55] at mapPartitions at NaiveBayes.scala:90)
16/07/25 17:53:03 INFO TaskSchedulerImpl: Adding task set 22.0 with 1 tasks
16/07/25 17:53:03 INFO TaskSetManager: Starting task 0.0 in stage 22.0 (TID 34, localhost, partition 0, PROCESS_LOCAL, 2167 bytes)
16/07/25 17:53:03 INFO Executor: Running task 0.0 in stage 22.0 (TID 34)
16/07/25 17:53:03 INFO WholeTextFileRDD: Input split: Paths: C:/SUMMER/CS5560-T13 -SourceCode/W13 - Source Code/Spark2Ont/data/test/103:0+7242
16/07/25 17:53:03 INFO GenerateUnsafeProjection: Code generated in 184.069665 ms
16/07/25 17:53:03 INFO POS tagger model from edu/stanford/nlp/models/pos-tagger/english-left3words/english-left3words-distsim.tagger ... 16/07/25 17:53:10 INFO BlockManagerInfo: Removed broadcast_19
16/07/25 17:53:10 INFO ContextCleaner: Cleaned accumulator 10
done [6.8 sec].
1.0
16/07/25 17:53:14 INFO Executor: Finished task 0.0 in stage 22.0 (TID 34). 2189 bytes result sent to driver
16/07/25 17:53:14 INFO TaskSetManager: Finished task 0.0 in stage 22.0 (TID 34) in 11146 ms on localhost (1/1)
16/07/25 17:53:14 INFO TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool
16/07/25 17:53:14 INFO DAGScheduler: ResultStage 22 (foreach at testNaiveBayes.scala:52) finished in 11.155 s
16/07/25 17:53:14 INFO DAGScheduler: Job 22 finished: foreach at testNaiveBayes.scala:52, took 11.447525 s
16/07/25 17:53:14 INFO SparkContext: Invoking stop() from shutdown hook
16/07/25 17:53:14 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040
16/07/25 17:53:14 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/07/25 17:53:14 INFO MemoryStore: MemoryStore cleared
16/07/25 17:53:14 INFO BlockManager: BlockManager stopped
16/07/25 17:53:14 INFO BlockManagerMaster: BlockManagerMaster stopped
16/07/25 17:53:14 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/07/25 17:53:15 INFO SparkContext: Successfully stopped SparkContext
16/07/25 17:53:15 INFO ShutdownHookManager: Shutdown hook called
16/07/25 17:53:15 INFO ShutdownHookManager: Deleting directory C:\Users\Kiran\AppData\Local\Temp\spark-74c8ec6c-0d48-4637-a5d5-2f40234d8e29
16/07/25 17:53:15 INFO RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.

Process finished with exit code 0
```

we can see 1.0 as output.

when positive tweets are given as input

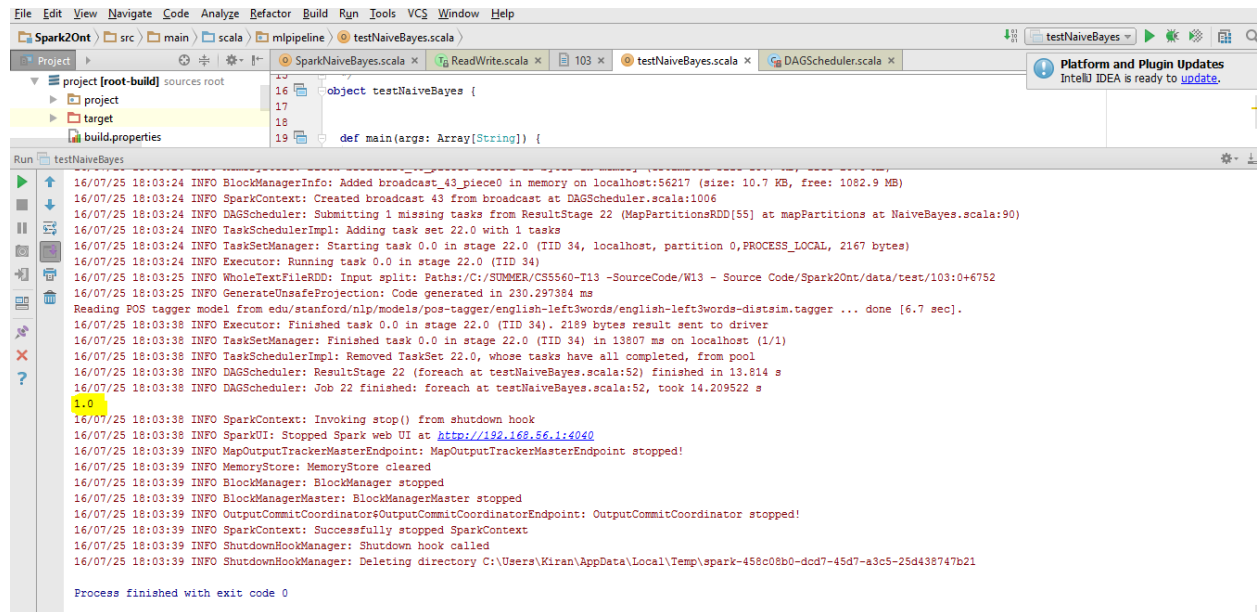


```
16/07/25 17:58:51 INFO TaskSchedulerImpl: Adding task set 22.0 with 1 tasks
16/07/25 17:58:51 INFO TaskSetManager: Starting task 0.0 in stage 22.0 (TID 34, localhost, partition 0, PROCESS_LOCAL, 2167 bytes)
16/07/25 17:58:51 INFO Executor: Running task 0.0 in stage 22.0 (TID 34)
16/07/25 17:58:51 INFO WholeTextFileRDD: Input split: Paths: C:/SUMMER/CS5560-T13 -SourceCode/W13 - Source Code/Spark2Ont/data/test/103:0+10766
16/07/25 17:58:51 INFO GenerateUnsafeProjection: Code generated in 80.416579 ms
16/07/25 17:58:51 INFO POS tagger model from edu/stanford/nlp/models/pos-tagger/english-left3words/english-left3words-distsim.tagger ... 16/07/25 17:58:54 INFO BlockManagerInfo: Removed broadcast_41
16/07/25 17:58:54 INFO ContextCleaner: Cleaned accumulator 24
16/07/25 17:58:54 INFO BlockManagerInfo: Removed broadcast_40_piece0 on localhost:55856 in memory (size: 13.8 KB, free: 1082.9 MB)
done [2.7 sec].
0.0
16/07/25 17:58:58 INFO Executor: Finished task 0.0 in stage 22.0 (TID 34). 2189 bytes result sent to driver
16/07/25 17:58:58 INFO TaskSetManager: Finished task 0.0 in stage 22.0 (TID 34) in 7064 ms on localhost (1/1)
16/07/25 17:58:58 INFO TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool
16/07/25 17:58:58 INFO DAGScheduler: ResultStage 22 (foreach at testNaiveBayes.scala:52) finished in 7.067 s
16/07/25 17:58:58 INFO DAGScheduler: Job 22 finished: foreach at testNaiveBayes.scala:52, took 7.352581 s
16/07/25 17:58:58 INFO SparkContext: Invoking stop() from shutdown hook
16/07/25 17:58:58 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040
16/07/25 17:58:58 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/07/25 17:58:58 INFO MemoryStore: MemoryStore cleared
16/07/25 17:58:58 INFO BlockManager: BlockManager stopped
16/07/25 17:58:58 INFO BlockManagerMaster: BlockManagerMaster stopped
16/07/25 17:58:58 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/07/25 17:58:58 INFO SparkContext: Successfully stopped SparkContext
16/07/25 17:58:58 INFO ShutdownHookManager: Shutdown hook called
16/07/25 17:58:58 INFO ShutdownHookManager: Deleting directory C:\Users\Kiran\AppData\Local\Temp\spark-4f0b12a9-09f1-4ef5-a0b5-388c584b97b2
16/07/25 17:58:58 INFO RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.

Process finished with exit code 0
```

we can see we got 0.0 as output

Now we have given 30 positive and 30 negative tweets as input to know how this model works. It has given output as 1.0 which tells us that if the file contains negative data with equal number of positive tweets, it considers as negative.



The screenshot shows the IntelliJ IDEA interface with the `testNaiveBayes.scala` file open. The code defines an object `testNaiveBayes` with a `main` method. The Run console displays the execution logs, which include Spark framework messages and the final output `1.0`.

```
16/07/25 18:03:24 INFO BlockManagerInfo: Added broadcast_43_piece0 in memory on localhost:56217 (size: 10.7 KB, free: 1082.9 MB)
16/07/25 18:03:24 INFO SparkContext: Created broadcast 43 from broadcast at DAGScheduler.scala:1006
16/07/25 18:03:24 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 22 (MapPartitionsRDD[55] at mapPartitions at NaiveBayes.scala:90)
16/07/25 18:03:24 INFO TaskSchedulerImpl: Adding task set 22.0 with 1 tasks
16/07/25 18:03:24 INFO TaskSetManager: Starting task 0.0 in stage 22.0 (TID 34, localhost, partition 0, PROCESS_LOCAL, 2167 bytes)
16/07/25 18:03:24 INFO Executor: Running task 0.0 in stage 22.0 (TID 34)
16/07/25 18:03:25 INFO WholeTextFileRDD: Input split: Paths: C:/SUMMER/CS5560-T13 -SourceCode/W13 - Source Code/Spark2Ont/data/test/103:0+6752
16/07/25 18:03:25 INFO GenerateUnsafeProjection: Code generated in 230.297384 ms
16/07/25 18:03:25 INFO Reading POS tagger model from edu/stanford/nlp/models/pos-tagger/english-left3words/english-left3words-distisim.tagger ... done [6.7 sec].
16/07/25 18:03:38 INFO Executor: Finished task 0.0 in stage 22.0 (TID 34). 2189 bytes result sent to driver
16/07/25 18:03:38 INFO TaskSetManager: Finished task 0.0 in stage 22.0 (TID 34) in 13807 ms on localhost (1/1)
16/07/25 18:03:38 INFO TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool
16/07/25 18:03:38 INFO DAGScheduler: ResultStage 22 (foreach at testNaiveBayes.scala:52) finished in 13.814 s
16/07/25 18:03:38 INFO DAGScheduler: Job 22 finished: foreach at testNaiveBayes.scala:52, took 14.209522 s
1.0
16/07/25 18:03:38 INFO SparkContext: Invoking stop() from shutdown hook
16/07/25 18:03:38 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040
16/07/25 18:03:38 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/07/25 18:03:38 INFO MemoryStore: MemoryStore cleared
16/07/25 18:03:38 INFO BlockManager: BlockManager stopped
16/07/25 18:03:38 INFO BlockManagerMaster: BlockManagerMaster stopped
16/07/25 18:03:38 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/07/25 18:03:38 INFO SparkContext: Successfully stopped SparkContext
16/07/25 18:03:38 INFO ShutdownHookManager: Shutdown hook called
16/07/25 18:03:38 INFO ShutdownHookManager: Deleting directory C:\Users\Kiran\AppData\Local\Temp\spark-458c08b0-dcd7-45d7-a3c5-25d438747b21

Process finished with exit code 0
```

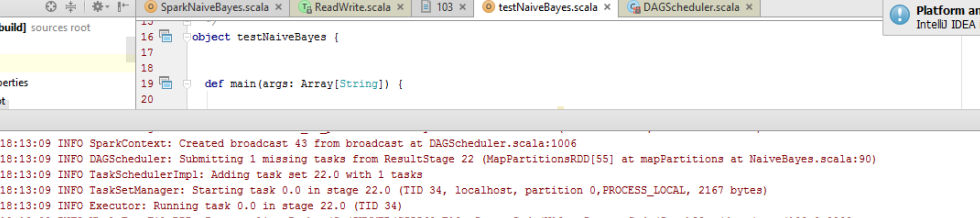
We are getting 1.0 as output which says us that data is having /having meaning like negative tweets.

The screenshot displays the IntelliJ IDEA IDE interface. The top menu bar includes File, Edit, View, Navigate, Code, Analyze, Refactor, Build, Run, Tools, VCS, Window, and Help. The main editor window shows the file `testNaiveBayes.scala` with the following code:

```
object testNaiveBayes {  
  16  
  17  
  18
```

The Run console at the bottom shows the execution output for the `testNaiveBayes` test. The output includes various INFO messages from the Spark framework, such as `WholeTextFileRDD: Input split`, `GenerateUnasfProjection: Code generated in 167.563577 ms`, `BlockManagerInfo: Removed broadcast_41`, `ContextCleaner: Cleaned accumulator 24`, `BlockManagerInfo: Removed broadcast_40_piece0 on localhost:56735 in memory (size: 13.8 KB, free: 1082.9 MB)`, `BlockManagerInfo: Removed broadcast_39_piece0 on localhost:56735 in memory (size: 15.1 KB, free: 1082.9 MB)`, `ContextCleaner: Cleaned accumulator 23`, `done [4.8 sec]`, `BlockManagerInfo: Removed broadcast_28_piece0 on localhost:56735 in memory (size: 3.3 KB, free: 1082.9 MB)`, `ContextCleaner: Cleaned accumulator 15`, `BlockManagerInfo: Removed broadcast_27_piece0 on localhost:56735 in memory (size: 13.8 KB, free: 1082.9 MB)`, `0.0`, `Executor: Finished task 0.0 in stage 22.0 (TID 34). 2189 bytes result sent to driver`, `TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool`, `DAGScheduler: Result Stage 22 (foreach at testNaiveBayes.scala:52) finished in 13.521 s`, `DAGScheduler: Job 22 finished: foreach at testNaiveBayes.scala:52, took 13.814247 s`, `SparkContext: Invoking stop() from shutdown hook`, `SparkUI: Stopped Spark web UI at http://192.168.56.1:4040`, `MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!`, `MemoryStore: MemoryStore cleared`, `BlockManager: BlockManager stopped`, `BlockManagerMaster: BlockManagerMaster stopped`, `OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!`, `SparkContext: Successfully stopped SparkContext`, `ShutdownHookManager: Shutdown hook called`, and `ShutdownHookManager: Deleting directory C:\Users\Kiran\AppData\Local\Temp\spark-b2bf673-9c2f-4d48-9ea5-3d6b03df2524`. The console concludes with `Process finished with exit code 0`.

when negative 50 positive 30 is given as input



The screenshot displays the IntelliJ IDEA interface. The top menu bar includes File, Edit, View, Navigate, Code, Analyze, Refactor, Build, Run, Tools, VCS, and Window. The main editor window shows the file `testNaiveBayes.scala` with the following code:

```
16 object testNaiveBayes {  
17  
18  
19 def main(args: Array[String]) {  
20
```

The left sidebar shows the project structure for `sources root`, including `project`, `target`, `build.properties`, and `plugins.sbt`. The bottom panel shows the `Run` tab with the output of the `testNaiveBayes` execution. The output log shows the following messages:

```
16/07/25 18:13:09 INFO SparkContext: Created broadcast 43 from broadcast at DAGScheduler.scala:1006  
16/07/25 18:13:09 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 22 (MapPartitionsRDD[55] at mapPartitions at NaiveBayes.scala:90)  
16/07/25 18:13:09 INFO TaskSchedulerImpl: Adding task set 22.0 with 1 tasks  
16/07/25 18:13:09 INFO TaskSetManager: Starting task 0.0 in stage 22.0 (TID 34, localhost, partition 0, PROCESS_LOCAL, 2167 bytes)  
16/07/25 18:13:09 INFO Executor: Running task 0.0 in stage 22.0 (TID 34)  
16/07/25 18:13:09 INFO WholeTextFileRDD: Input split: Paths[C:/SUMMER/CSS560-T13 - SourceCode/W13 - Source Code/Spark2ont/data/test/103:0+8839  
16/07/25 18:13:09 INFO GenerateUnsafeProjection: Code generated in 165.796989 ms  
Reading POS tagger model from edu.stanford.nlp/models/pos-tagger/english-left3words/english-left3words-distsim.tagger ... done [3.8 sec].  
16/07/25 18:13:18 INFO Executor: Finished task 0.0 in stage 22.0 (TID 34). 2189 bytes result sent to driver  
  
16/07/25 18:13:18 INFO TaskSetManager: Finished task 0.0 in stage 22.0 (TID 34) in 8601 ms on localhost (1/1)  
16/07/25 18:13:18 INFO TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool  
16/07/25 18:13:18 INFO DAGScheduler: ResultStage 22 (foreach at testNaiveBayes.scala:52) finished in 8.605 s  
16/07/25 18:13:18 INFO DAGScheduler: Job 22 finished: foreach at testNaiveBayes.scala:52, took 9.371377 s  
16/07/25 18:13:18 INFO SparkContext: Invoking stop() from shutdown hook  
16/07/25 18:13:18 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040  
16/07/25 18:13:18 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!  
16/07/25 18:13:18 INFO MemoryStore: MemoryStore cleared  
16/07/25 18:13:18 INFO BlockManager: BlockManager stopped  
16/07/25 18:13:18 INFO BlockManagerMaster: BlockManagerMaster stopped  
16/07/25 18:13:22 INFO SparkContext: Successfully stopped SparkContext  
16/07/25 18:13:22 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!  
16/07/25 18:13:22 INFO ShutdownHookManager: Shutdown hook called  
16/07/25 18:13:22 INFO ShutdownHookManager: Deleting directory C:\Users\Kiran\AppData\Local\Temp\spark-e383b0c6-36c5-48fd-96cd-a30b8b438660  
  
Process finished with exit code 0
```

we can see that output is 1.0 which tells us that here positive tweets are more in number and it is domination the negative tweets.