Institute of Computer Technology

B. Tech. Computer Science and Engineering

Sub: Big Data Analytics

Project Report

Group no. : 20

Project no. : 24

1. AIM of the project :

The primary objective of this project is to conduct a semantic analysis of Demonetization in India using Twitter data. The aim is to understand the diverse opinions and sentiments expressed by individuals on Twitter regarding demonetization. The project specifically focuses on classifying tweets as either positive or negative, providing insights into how the public perceived the demonetization move.

2. Name of students along with enrolment numbers who were included in the project

- **Student 1:** Shan Bhoharia - **Enrollment Number :** 21162121005

- **Student 2:** Devarsh juneja - **Enrollment Number :** 21162121010

- **Student 3:** Kshitij Nigam - **Enrollment Number :** 21162121012

3. Observation and Analysis

In the process of analyzing the demonetization Twitter data, diverse and contrasting viewpoints were observed. People expressed a range of sentiments, from support and approval to criticism and dissatisfaction. The mixed feelings captured in the tweets reflect the complexity and multifaceted nature of public opinion during the demonetization period.

4. Data Cleaning Steps

To ensure accurate analysis, the following steps were taken to clean the data

- Omitted special characters, including ",", """, "|", and ";"
- Substituted null values where applicable
- Removed unnecessary headers for better clarity

5. Tool & conceptualization

Tool Used: **Apache PIG**.

The chosen tool for this project is **PIG**.This tool was selected due to

its capabilities in handling large-scale data analysis, ease of use, and compatibility with the project requirements. The conceptualization involved leveraging the tool's features to perform sentiment analysis on the Twitter data and classify tweets as positive or negative.

6. Code/queries/commands used for execution of the project.

- ❖ Tweets = LOAD '/BDA-Project/demoetization.csv' USING PigStorage(' , ');
- ❖ Extracted_tweets = FOREACH tweets GENRATE $0 as id, $1 as text;
- ❖ Tokens =foreach extrated_tweets generate id,text FLATTEN(TOKENIZE(text)) AS word;
- ❖ Dump tokens

- ❖ AFFIN = LOAD '/BDA-Project/AFFIN.txt' USING PigStorage('\t') AS (word: chararray, rating: int);
- ❖ Dump AFFIN

- ❖ Word_rating = JOIN tokens BY word LEFT OUTER, AFFIN BY word USING 'replicated';
- ❖ rating = FOREACH word_rating GENRATE tokens::id AS id, tokens::text AS text, AFFIN:: rating as rate;
- ❖ Word_group = Group rating by (id, text);
- ❖ Avg_rate = FOREACH word_group GENRATE group,

AVG(rating.rate) AS tweet_rating;

❖ **Positive_tweets = FILTER avg_rate BY tweet_rating >= 0;**
❖ **negative_tweets = FILTER avg_rate BY tweet_rating < 0;**

❖ **positive_tweets_count = FOREACH (GROUP
   positive_tweets ALL) generate count(positive_tweets);**
❖ **dump positive_tweets_count;**

❖ **negative_tweets_count = FOREACH (GROUP
   negative_tweets ALL) GENERATE
   COUNT(negative_tweets);**
❖ **Dump negative_tweets_count;**

7. Screenshots and Output

```
ess
grunt> tweets = LOAD '/BDA-Project/demonetization.csv' USING PigStorage(',');
grunt> extracted_tweets = FOREACH tweets GENERATE $0 as id, $1 as text;
```

```
grunt> tokens = foreach extracted_tweets generate id,text,FLATTEN(TOKENIZE(text)) AS word;
```

```
tization related death. How 3
grunt> dump tokens
```

```
ve over DeMonetization &amp; he is also detained as he Tried 2 March�",he)
("7997","RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts mo
ve over DeMonetization &amp; he is also detained as he Tried 2 March�",Tried)
("7997","RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts mo
ve over DeMonetization &amp; he is also detained as he Tried 2 March�",2)
("7997","RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts mo
ve over DeMonetization &amp; he is also detained as he Tried 2 March�",March�)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,#demonetization)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,will)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,help)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,combat)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,terror)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,because)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,Pak)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,won't)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,be)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,able)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,to)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,print)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,new)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,notes!)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,And)
("7998","#demonetization will help combat terror because Pak won't be able to print
new notes! And now,now)

grunt> AFFIN = LOAD '/BDA-Project/AFFIN.txt' USING PigStorage('\t') AS (word: chararray, rating: int);
2023-11-24 23:56:47,577 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
j, use fs.defaultFS
2023-11-24 23:56:47,577 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Ins
tead, use mapreduce.jobtracker.address
2023-11-24 23:56:47,688 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-24 23:56:47,688 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt> S

grunt> dump AFFIN
```

```
(won,3)
(wonderful,4)
(woo,3)
(woohoo,3)
(wooo,4)
(woow,4)
(worn,-1)
(worried,-3)
(worry,-3)
(worrying,-3)
(worse,-3)
(worsen,-3)
(worsened,-3)
(worsening,-3)
(worsens,-3)
(worshiped,3)
(worst,-3)
(worth,2)
(worthless,-2)
(worthy,2)
(wow,4)
(wowow,4)
(wowww,4)
(wrathful,-3)
(wreck,-2)
(wrong,-2)
(wronged,-2)
(wtf,-4)
(yeah,1)
(yearning,1)
(yeees,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
```

```
grunt> word_rating = JOIN tokens BY word LEFT OUTER, AFFIN BY word USING 'replicated';
2023-11-25 00:04:14,271 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:04:14,271 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt> rating = FOREACH word_rating GENERATE tokens::id AS id, tokens::text AS text, AFFIN::rating AS rate;
2023-11-25 00:07:28,977 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:07:28,977 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt> word_group = Group rating by (id, text);
2023-11-25 00:08:02,148 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:08:02,148 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt> avg_rate = FOREACH word_group GENERATE group, AVG(rating.rate) AS tweet_rating;
2023-11-25 00:09:00,105 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:09:00,105 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
grunt>
```

```
"1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requir
s full disclosure &amp;�",fishy,,)
"1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requir
s full disclosure &amp;�",clearly,clearly,1)
"1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requir
s full disclosure &amp;�",&amp;�,,)
"1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requir
```

```
grunt> positive_tweets = FILTER avg_rate BY tweet_rating >= 0;
2023-11-25 00:34:06,066 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:34:06,066 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2023-11-25 00:34:06,066 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
grunt> negative_tweets = FILTER avg_rate BY tweet_rating < 0;
2023-11-25 00:34:44,101 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:34:44,101 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2023-11-25 00:34:44,101 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 2 time(s).
grunt>
```

```
grunt> positive_tweets_count = FOREACH (GROUP positive_tweets ALL) GENERATE COUNT(positive_tweets);
```

```
grunt> dump positive_tweets_count;
```

```
tead, use mapreduce.jobtracker.address
2023-11-25 00:42:25,068 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2023-11-25 00:42:25,078 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2023-11-25 00:42:25,078 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(2195)
```
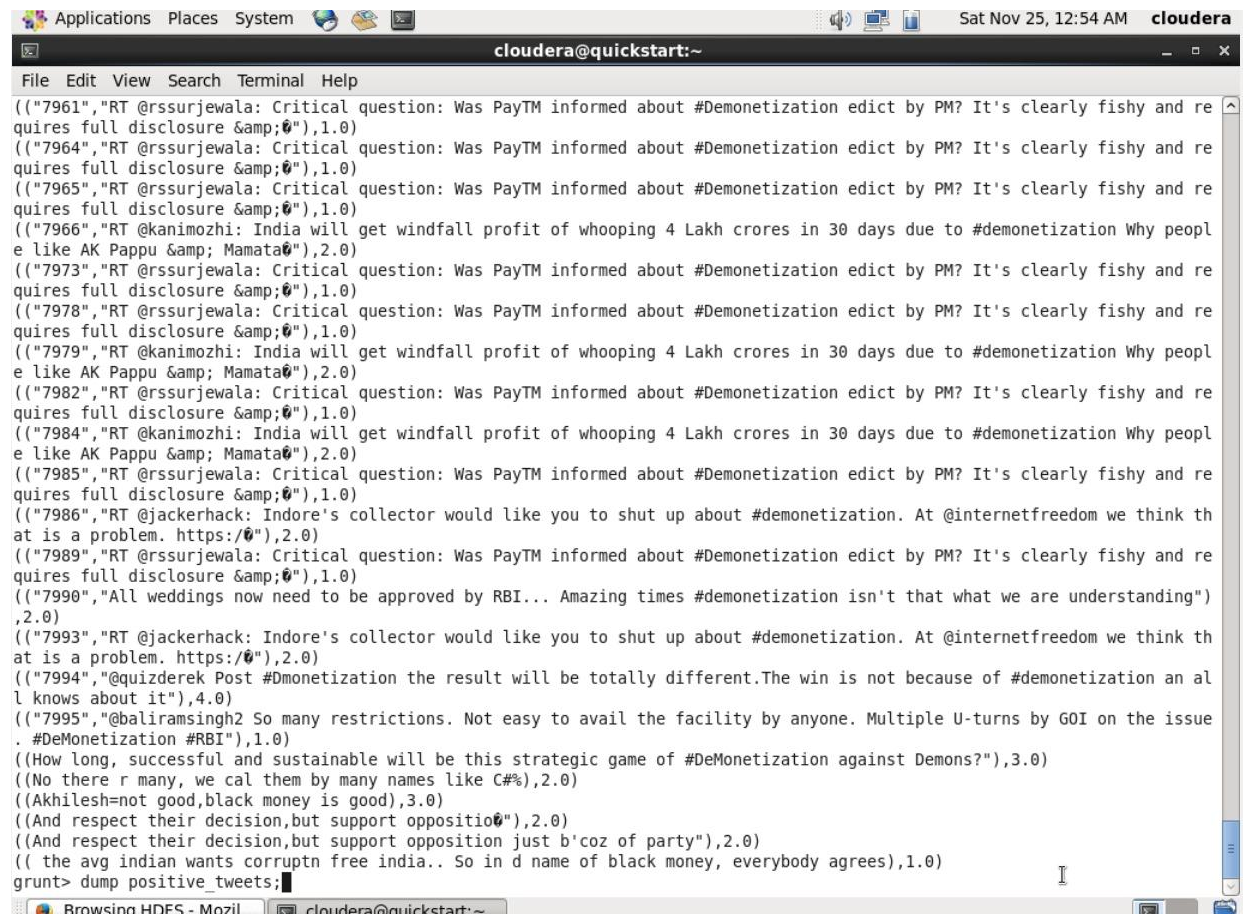
```
grunt> negative_tweets_count = FOREACH (GROUP negative_tweets ALL) GENERATE COUNT(negative_tweets);
2023-11-25 00:45:02,248 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-11-25 00:45:02,248 [main] WARN  org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2023-11-25 00:45:02,248 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 2 time(s).
grunt>
```
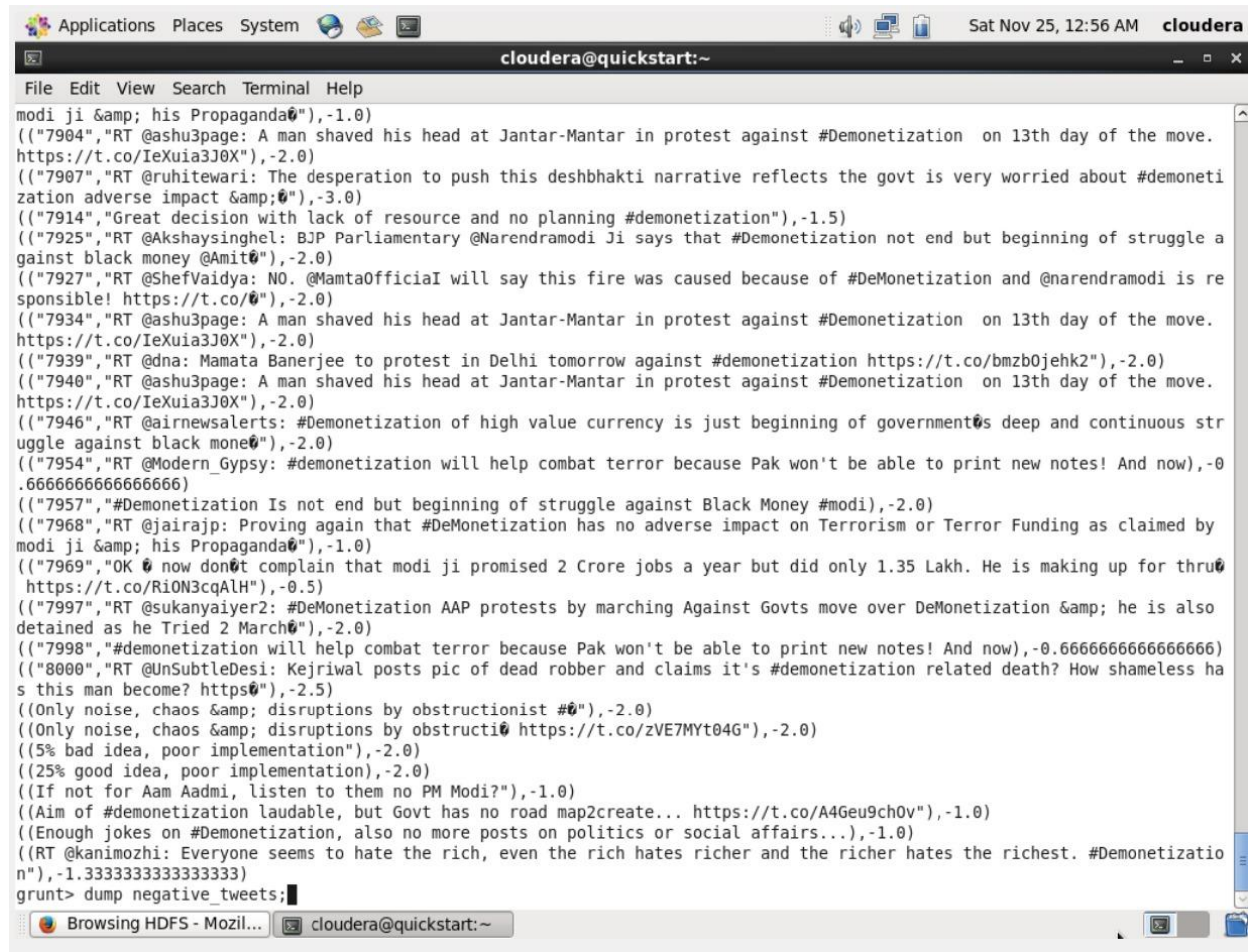
```
grunt> dump negative_tweets_count;
```

```
2023-11-25 00:48:44,713 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2023-11-25 00:48:44,713 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
cess : 1
(1258)
```

```
(("7961","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7964","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7965","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7966","RT @kanimozhi: India will get windfall profit of whooping 4 Lakh crores in 30 days due to #demonetization Why peopl
e like AK Pappu &amp; Mamata�"),2.0)
(("7973","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7978","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7979","RT @kanimozhi: India will get windfall profit of whooping 4 Lakh crores in 30 days due to #demonetization Why peopl
e like AK Pappu &amp; Mamata�"),2.0)
(("7982","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7984","RT @kanimozhi: India will get windfall profit of whooping 4 Lakh crores in 30 days due to #demonetization Why peopl
e like AK Pappu &amp; Mamata�"),2.0)
(("7985","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7986","RT @jackerhack: Indore's collector would like you to shut up about #demonetization. At @internetfreedom we think th
at is a problem. https:/�"),2.0)
(("7989","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and re
quires full disclosure &amp;�"),1.0)
(("7990","All weddings now need to be approved by RBI... Amazing times #demonetization isn't that what we are understanding")
,2.0)
(("7993","RT @jackerhack: Indore's collector would like you to shut up about #demonetization. At @internetfreedom we think th
at is a problem. https:/�"),2.0)
(("7994","@quizderek Post #Dmonetization the result will be totally different.The win is not because of #demonetization an al
l knows about it"),4.0)
(("7995","@baliramsingh2 So many restrictions. Not easy to avail the facility by anyone. Multiple U-turns by GOI on the issue
. #DeMonetization #RBI"),1.0)
((How long, successful and sustainable will be this strategic game of #DeMonetization against Demons?"),3.0)
((No there r many, we cal them by many names like C#%),2.0)
((Akhilesh=not good,black money is good),3.0)
((And respect their decision,but support oppositio�"),2.0)
((And respect their decision,but support opposition just b'coz of party"),2.0)
(( the avg indian wants corruptn free india.. So in d name of black money, everybody agrees),1.0)
grunt> dump positive_tweets;
```

```
modi ji &amp; his Propaganda0"),-1.0)
(("7904","RT @ashu3page: A man shaved his head at Jantar-Mantar in protest against #Demonetization  on 13th day of the move.
https://t.co/IeXuia3J0X"),-2.0)
(("7907","RT @ruhitewari: The desperation to push this deshbhakti narrative reflects the govt is very worried about #demoneti
zation adverse impact &amp;0"),-3.0)
(("7914","Great decision with lack of resource and no planning #demonetization"),-1.5)
(("7925","RT @Akshaysinghel: BJP Parliamentary @Narendramodi Ji says that #Demonetization not end but beginning of struggle a
gainst black money @Amit0"),-2.0)
(("7927","RT @ShefVaidya: NO. @MamtaOfficiaI will say this fire was caused because of #DeMonetization and @narendramodi is re
sponsible! https://t.co/0"),-2.0)
(("7934","RT @ashu3page: A man shaved his head at Jantar-Mantar in protest against #Demonetization  on 13th day of the move.
https://t.co/IeXuia3J0X"),-2.0)
(("7939","RT @dna: Mamata Banerjee to protest in Delhi tomorrow against #demonetization https://t.co/bmzbOjehk2"),-2.0)
(("7940","RT @ashu3page: A man shaved his head at Jantar-Mantar in protest against #Demonetization  on 13th day of the move.
https://t.co/IeXuia3J0X"),-2.0)
(("7946","RT @airnewsalerts: #Demonetization of high value currency is just beginning of government0s deep and continuous str
uggle against black mone0"),-2.0)
(("7954","RT @Modern_Gypsy: #demonetization will help combat terror because Pak won't be able to print new notes! And now),-0
.6666666666666666)
(("7957","#Demonetization Is not end but beginning of struggle against Black Money #modi),-2.0)
(("7968","RT @jairajp: Proving again that #DeMonetization has no adverse impact on Terrorism or Terror Funding as claimed by
modi ji &amp; his Propaganda0"),-1.0)
(("7969","OK 0 now don0t complain that modi ji promised 2 Crore jobs a year but did only 1.35 Lakh. He is making up for thru0
 https://t.co/RiON3cqAlH"),-0.5)
(("7997","RT @sukanyaiyer2: #DeMonetization AAP protests by marching Against Govts move over DeMonetization &amp; he is also
detained as he Tried 2 March0"),-2.0)
(("7998","#demonetization will help combat terror because Pak won't be able to print new notes! And now),-0.6666666666666666)
(("8000","RT @UnSubtleDesi: Kejriwal posts pic of dead robber and claims it's #demonetization related death? How shameless ha
s this man become? https0"),-2.5)
((Only noise, chaos &amp; disruptions by obstructionist #0"),-2.0)
((Only noise, chaos &amp; disruptions by obstructi0 https://t.co/zVE7MYt04G"),-2.0)
((5% bad idea, poor implementation"),-2.0)
((25% good idea, poor implementation),-2.0)
((If not for Aam Aadmi, listen to them no PM Modi?"),-1.0)
((Aim of #demonetization laudable, but Govt has no road map2create... https://t.co/A4Geu9chOv"),-1.0)
((Enough jokes on #Demonetization, also no more posts on politics or social affairs...),-1.0)
((RT @kanimozhi: Everyone seems to hate the rich, even the rich hates richer and the richer hates the richest. #Demonetizatio
n"),-1.3333333333333333)
grunt> dump negative_tweets;
```

8. Conclusion :

Based on the sentiment analysis of demonetization tweets, it can be concluded that the public perception was diverse. The sentiment polarity varied across tweets, indicating a mix of positive and negative sentiments. Graphical representations further enhance the understanding of sentiment distribution.

9. What else can you use the data for?? (Like for historical analysis/ future prediction/ behavior of something/ pattern)

Apart from the current analysis, the demonetization Twitter data can be utilized for:

- Historical analysis of public sentiment trends
- Predictive modeling for future events or policy changes
- Understanding patterns in public reactions to socio-economic developments

This project sheds light on the significance of leveraging big data analytics for extracting meaningful insights from social media data, providing valuable perspectives on public sentiment during critical events like demonetization.