Sentiment Analysis on Demonetization

Let us find out the views of different people on the demonetization by analysing the tweets from twitter. Here is the dataset where twitter tweets are gathered in CSV format.

Metadata
- id
- Text (Tweets)
- favorited
- favoriteCount
- replyToSN
- created
- truncated
- replyToSID
- id
- replyToUID
- statusSource
- screenName
- retweetCount
- isRetweet
- retweeted

Problem statements:
1. Load data into Pig
2. Extract columns id and Text from the loaded data
3. Divide the data into words
4. Load dictionary into Pig
5. Perform map side join between dictionary and bag of individual words
6. View the schema of the joined bag
7. Extract id, text, word rating (from dictionary) to a new bag
8. Group the rating of all the words in a tweet
9. Perform average operation on the rating of the words per each tweet
10. Filter positive and negative tweets

1) Load data into Pig

Answer :

```
load_tweets = LOAD '/demonetization-tweets.csv' USING PigStorage(',');
```

2) Extract columns id and Text from the loaded data

Answer :

```
extract_details = FOREACH load_tweets GENERATE $0 as id,$1 as text;
```

3) Divide the data into words

Answer :

```
tokens = foreach extract_details generate id,text, FLATTEN(TOKENIZE(text)) As word;
("1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requires full disclosure &amp;�",RT)
describe tokens;
tokens: {id: bytearray,text: bytearray,word: chararray}
```

4) Load dictionary into Pig

Answer :

```
dictionary = load '/AFINN.txt' using PigStorage('\t') AS(word:chararray,rating:int)
```

5) Perform map side join between dictionary and bag of individual words

Answer :

```
word_rating = join tokens by word left outer, dictionary by word using 'replicated';
```

6) View the schema of the joined bag

Answer :

```
describe word_rating;
word_rating: {tokens::id: bytearray,tokens::text: bytearray,tokens::word:chararray,dictionary::word: chararray,dictionary::rating: int}
```

7) Extract id, text, word rating (from dictionary) to a new bag

Answer :

```
rating = foreach word_rating generate tokens::id as id,tokens::text as text,dictionary::rating as rate;
rating: {id: bytearray,text: bytearray,rate: int}
```

8) Group the rating of all the words in a tweet

Answer :

```
word_group = group rating by (id,text);
```

9) Perform average operation on the rating of the words per each tweet

Answer :

```
avg_rate = foreach word_group generate group, AVG(rating.rate) as tweet_rating;
```

10) Filter positive and negative tweets

Answer :

```
positive_tweets = filter avg_rate by tweet_rating>=0;
negative_tweets = filter avg_rate by tweet_rating<0;
```