Sentiment Analysis on Demonetization

Let us find out the views of different people on the demonetization by analysing the tweets from twitter.

Now we will load the data into pig using PigStorage as follows:

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

Now after loading successfully, you can see the tweets loaded successfully into pig by using the **dump** command.

Here is the sample tweet

Metadata of the tweets are as follows:

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

Now from this columns, we will extract the **id** and the **tweet_text** as follows
extract_details = FOREACH load_tweets GENERATE $0 as id,$1 as text;

Now if you dump the extracted columns, you will get the id and the tweet_text
Now we will divide the tweet_text into words to calculate the sentiment of the whole tweet.

tokens = foreach extract_details generate id,text, FLATTEN(TOKENIZE(text)) As word;

For every word in the tweet_text, each word will be taken and created as a new row
You can use the dump command to check the same. Here is the sample.

("1","RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by PM? It's clearly fishy and requires full disclosure &

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

In the above statement **describe word_rating**, we can see that the word_rating has joined the tokens (consists of id, tweet text, word) statement and the dictionary(consists of word, rating).
Now we will extract the **id,tweet text** and **word rating(**from the dictionary**)** by using the below relation.

rating = foreach word_rating generate tokens::id as id,tokens::text as text, dictionary::rating as rate;

We can now see the schema of the relation rating by using the command **describe rating**.
rating: {id: bytearray,text: bytearray,rate: int}
In the above statement **describe rating** we can see that our relation now consists of **id,tweet text** and **rate(**for each word**).**
Now, we will group the **rating of all the words in a tweet** by using the below relation:

word_group = group rating by (id,text);

Here we have grouped by two constraints, **id** and **tweet text**.
Now, let's perform the **Average** operation on the **rating of the words per each tweet**.

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

Now we have calculated the Average rating of the tweet using the rating of each word.
From the above relation, we will get all the tweets i.e., both positive and negative. Here, we can classify the positive tweets by taking the rating of the tweet which can be from **0-5.** We can classify the negative tweets by taking the rating of the tweet from **-5 to -1.**

We have now successfully performed the Sentiment Analysis on Twitter data using Pig. We now have the tweets and its rating, so let's perform an operation to filter out the positive tweets.
Now we will filter the positive tweets using the below statement:

positive_tweets = filter avg_rate by tweet_rating>=0;

Here are the sample tweets with positive ratings.
Like this we will also filter the negative tweets as follows:

negative_tweets = filter avg_rate by tweet_rating<0;

Like this, you can perform sentiment analysis using Pig.