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