<Code>

MR. BHARATHKUMAR S

SENTIMENT ANAYSIS
DEMOTIZATION –PIG USECASE

```
#Load-data from local to pig-storage:
tweets= LOAD '/user/ubuntu/project1/d_tweets.csv' USING PigStorage(',')  AS (id:chararray,text:chararray,favorited:chararray,favoritecount:chararray,replaytosn:chararray,created:chararray,truncated:chararray,replaytosid:chararray,dupid:chararray,replaytouid:chararray,statussource:chararray,screenname:chararray,retweetcount:chararray,isretweet:chararray,retweeted:chararray);
#Required rows are id and tweets(text):
extract_details= FOREACH  tweets GENERATE  $0  as  id,$1 as text;
STORE extract_details into '/extract_details'
#Tokenize as a word:
tokens= FOREACH extract_details GENERATE id,text, FLATTEN(TOKENIZE(text)) AS word;
#Download and load affin dictionary to pigstorage:
dictionary = load  '/user/ubuntu/project1/AFINN.txt' using PigStorage ('\t')  AS (word:chararray,rating:int);
STORE dictionary into '/dictionary';
#Joining the text and affin-dictionary words:
word_rating = join  tokens  by  word  left outer, dictionary  by  word using 'replicated';
describe word_rating;
#Rating each word in a single tweet by using the dictionary rating:
rating= FOREACH  word_rating GENERATE  tokens::id AS id,tokens::text AS text ,dictionary::rating AS rate;
describe rating;
#Group the rating of the words based on id and tweets :
word_group =  group rating by (id,text);
#GROUP AVERAGE RATING(GROUP ALL THE TWEETS RATING eg: -ve to +ve ):
avg_rate=  FOREACH  word_group  GENERATE  group ,AVG(rating.rate) as tweet_rating;
STORE avg_rate  into '/avg_rate';
#Group the positive tweets by(rating>=0)positive ratings:
positive_tweets= FILTER  avg_rate  by  tweet_rating>=0;
STORE  positive_tweets '/positive_tweets';
#Group the negative tweets by(<0)negative ratings:
negative_tweets= FILTER avg_rate AS tweet_rating<0;
STORE  negative_tweets '/negative_tweets';
```