

Company Reputation using Sentiment Analysis using Twitter Data

Soham Kolhe¹, Ankur Konwar², Anand Saxena³ Aditya P. Kamble⁴

MTech Student, Department of Computer Science Engineering, International Institute of Information Technology,
Bangalore, India¹

Chief Manager (Information Systems), Indian Oil Corporation Limited, Pipelines Division, Noida, India²

Young Professional-II (Information Technology), ICAR-Indian Institute of Soybean Research, Indore, India³

Senior Student, Portola High School, Irvine, CA, United States⁴

Abstract: The data from the Twitter sentimental analysis experiment has gained much renown as a topic of research. The potential to obtain information about public opinion by breaking down Twitter data and automatically classifying its sentimental polarity has consistently attracted researchers due to the concise language commonly used in tweets. The aim of this study was to use the Valence Aware Dictionary for sEntiment Reasoner (VADER), to classify the sentiments expressed in Twitter data. In this study, we developed a generic tool for analyzing tweets. We used VADER to categorize tweets on a particular keyword to predict recent trends about it. We successfully used the generated tool to classify multilingual tweets on any keyword to analyse public opinion.

Keywords: Natural Language Processing, NLP, Twitter, sentiment analysis, VADER.

I. INTRODUCTION

In recent years, countless people have been attracted to social-networking platforms. Social media technologies exist in several different forms, such as blogs, business networks, photo sharing, forums, microblogs, enterprise social networks, video sharing networks, and social networks. With the increase in the number of social media technologies, various online social networking platforms, such as Facebook, YouTube, and Twitter have become popular. People mainly use these social sites to express emotions, thoughts, and opinions about life events. Because of this, such platforms rapidly became eclectic sources of information. This is simply due to the varied nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive, negative or neutral sentiments in daily life. These networks enable users to have discussions with different people across the world and to post messages in the forms of texts, images, and videos [1], [2]. Moreover, social media is an enormous compendium of information for companies to monitor public opinion and receive polls about the products they manufacture. Microblogging services like Twitter have become the best known and the most commonly used platforms. They have evolved to become the significant hives of information that they are known for today [3]. Twitter is a popular microblogging service that allows users to share, deliver, and interpret real-time, short, and simple messages called tweets [4]. Therefore, Twitter provides a rich mine of data that is used in the field of opinion mining and sentiment analysis. Recently, sentiment analysis of Twitter data has been used to extract and classify information about the opinions expressed on Twitter concerning several topics, such as predictions, reviews, elections, and marketing. Companies have started to use the data provided by these general sentiment studies to examine user reactions and reply to users on microblogs. One major challenge in sentiment analysis has currently been observed in recent studies: building technology to detect and summarize an overall sentiment.

Sentiment Analysis, in a nutshell, is a text and data classification tool that leverages Natural language Processing (NLP) for data analysis and provides insight into the overall sentiment (positive, negative or neutral) about the data. This in turn, helps in determining the author's opinions and attitude towards a written piece of text. This technique uses a scaling system that tells the emotions and attitude of a text. This scoring makes it easier to bifurcate the text into three categories: positive, negative and neutral. Sentiment analysis, also known as Opinion Mining or Emotion AI, is widely used in the business sector to understand customer feedback and public opinion towards a brand or product. This knowledge can then be used for marketing strategies, advertising campaigns, market research, service or product improvement, and much more.

In this study, we took a look at one such popular social-networking platform called Twitter and developed a software tool to sort through "tweets" into positive, negative and neutral sentiment using keywords. The primary issue with Twitter sentiment analysis is the identification of the most suitable sentiment classifier that can correctly classify the tweets. In our work we used VADER (Valence Aware Dictionary and sEntiment Reasoner) [5]. It is a lexicon and rule-based

sentiment analysis tool that is specifically attuned to sentiments expressed in social media. It is fully open-sourced under the MIT License. We analysed the tweets on Indian Oil Corporation Limited (IOCL) to study the overall reputation of the company. The tweets in different Indian languages using different keywords were extracted using IOCL Twitter handle. We extracted the recent tweets and filtered those tweets having Indian Oil Corporation Limited in one form or the other. After that we cleaned, translated, and analysed the tweets. The hashtag is then extracted from each handle. The count of mostly used hashtags is displayed in different ways to judge the overall reputation of the company.

II. METHODS AND MATERIALS

A. Phases of the system

The current study consists of following Seven phases. The detail is shown in system flow chart (Fig. 1)

- Acquisition of Twitter Data- For this user can give any number of keywords to extract the tweets given by tweeters. The interface extracts only the most popular tweets on a particular keyword i.e tweets with more than ten counts.
- Storing the extracted twitter data - It stores the extracted tweets in a csv file database.
- Cleaning of twitter data-It focuses on initial pre-processing work carried out to clean and remove irrelevant information from the tweets. It includes removing handles, tweet id, URL and any special character's present in the tweets.
- Translation of multi-linguistic tweets into English- The tweets having the keywords may be in different languages. it needs translation of the multi-lingual tweets messages into English language.
- Sentiment Analysis using VADER- VADER analyser is used to classify and to give scores to the tweets. Based on sentiment whether neutral, positive or negative tweets are stored in a list [6].
- HashTag Analysis-The hashtags are extracted from each handle. The count of mostly used hashtags is displayed.
- Report generation-Based on the overall analysis a report is generated. Report contains strong hold and weak point or scope of development with a numerical count.

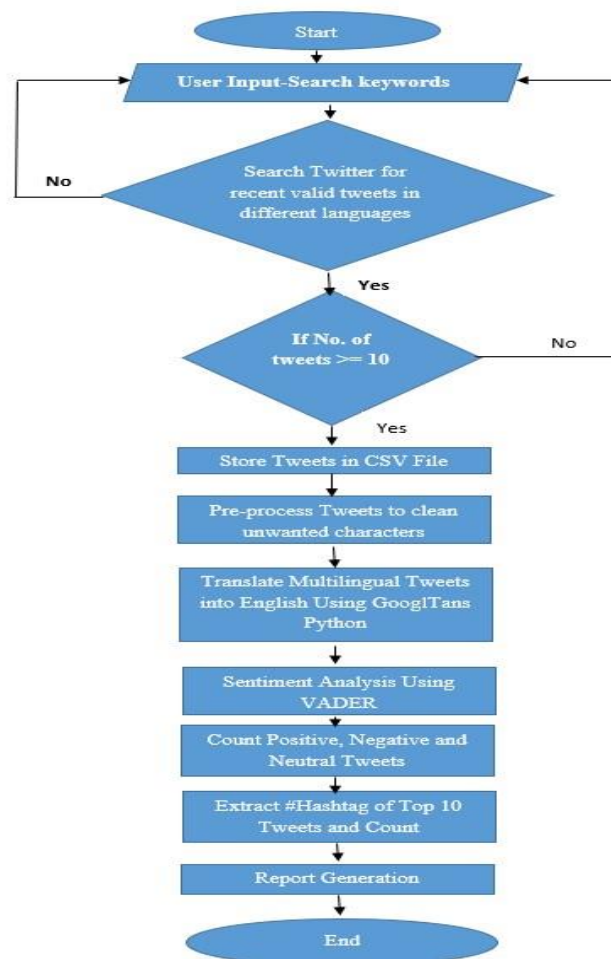
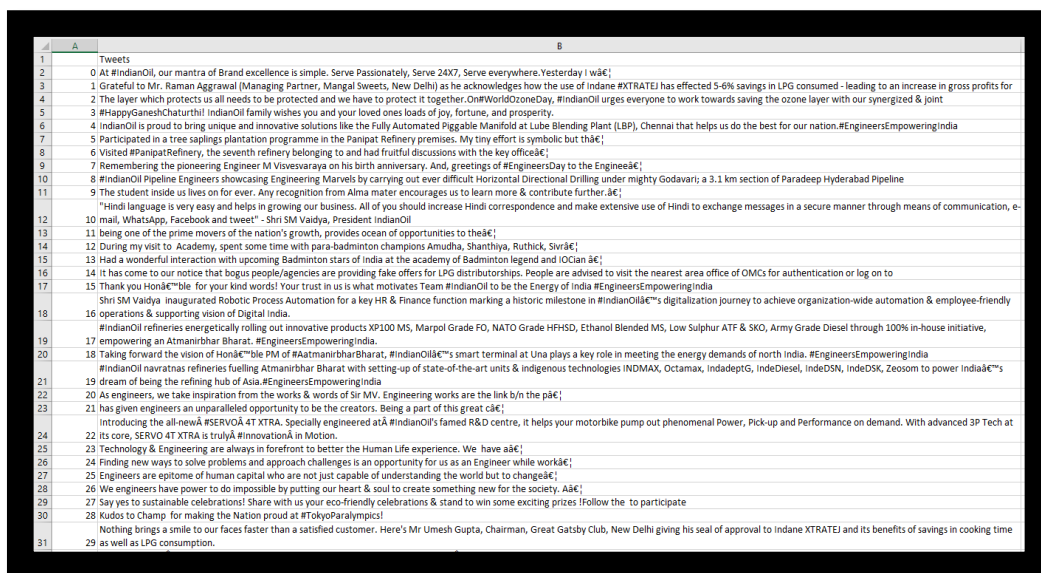


Fig. 1 System Flow chart

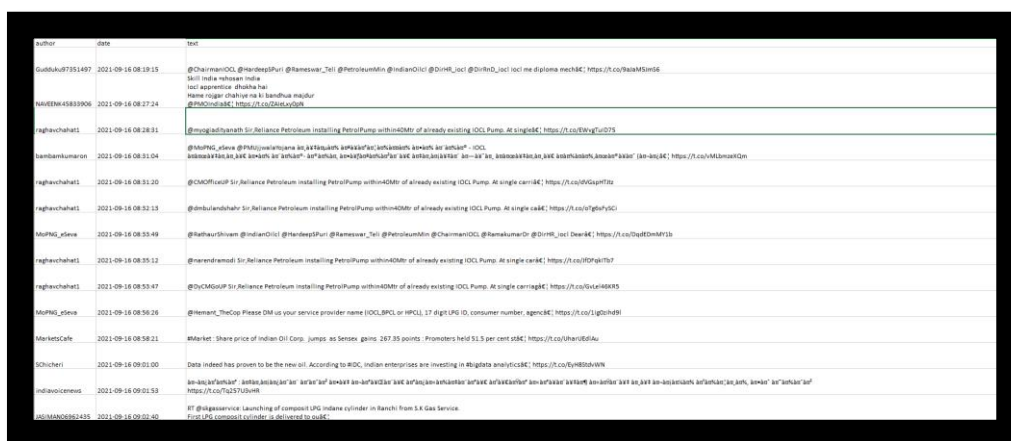
1) Data Extraction: The methods used to extract Twitter data allowed real-time access to publicly available raw tweets. To gather the data, we used Tweepy [7]. Tweepy is a python library for accessing Twitter API. Using Twitter API, we extracted recent tweets made form the user. Also tweet stream is possible which enable us to filter out the recent tweets. Tweet stream is a real time process so performed for a particular time interval. Using tweet stream from twitter API we extracted those tweets which has company name mentioned in one form or the other. Thus, recent tweets concerning any keyword e.g Oil or Indian Oil which were published on Twitter’s public message board and posted recently were extracted. In order to collect the most relevant tweets, we used hashtags containing the company’s name and product like Indian Oil, Oil, IOCL, Indian Oil Corporation Ltd. These were used as keywords to retrieve tweets, such as #Indian Oil, #Oil, #IOCL, #Indian Oil Corporation Limited. Once we get the enough tweet, we store them in a csv file. The latest 100 tweets are extracted from IOCL Twitter handle. Given below are the latest 30 tweets (Fig. 2). The tweets are in raw format and need to be cleaned for analysis.



	A	B
1		Tweets
2	0	At #IndianOil, our mantra of Brand excellence is simple. Serve Passionately, Serve 24x7, Serve everywhere. Yesterday I wāc;
3	1	Grateful to Mr. Raman Aggrawal (Managing Partner, Mangal Sweets, New Delhi) as he acknowledges how the use of Indane #XTRATEJ has effected 5-6% savings in LPG consumed - leading to an increase in gross profits for
4	2	The layer which protects us all needs to be protected and we have to protect it together. On#WorldOzoneDay, #IndianOil urges everyone to work towards saving the ozone layer with our synergized & joint
5	3	#HappyGaneshChaturthi! IndianOil family wishes you and your loved ones loads of joy, fortune, and prosperity.
6	4	IndianOil is proud to bring unique and innovative solutions like the Fully Automated Piggable Manifold at Lube Blending Plant (LBP), Chennai that helps us do the best for our nation. #EngineersEmpoweringIndia
7	5	Participated in a tree saplings plantation programme in the Panipat Refinery premises. My tiny effort is symbolic but thāc;
8	6	Visited #PanipatRefinery, the seventh refinery belonging to and had fruitful discussions with the key officeāc;
9	7	Remembering the pioneering Engineer M Visvesvaraya on his birth anniversary. And, greetings of #EngineersDay to the Engineāc;
10	8	#IndianOil Pipeline Engineers showcasing Engineering Marvels by carrying out ever difficult Horizontal Directional Drilling under mighty Godavari; a 3.1 km section of Paradeep Hyderabad Pipeline
11	9	The student inside us lives on for ever. Any recognition from Alma mater encourages us to learn more & contribute further. āc;
12	10	"Hindi language is very easy and helps in growing our business. All of you should increase Hindi correspondence and make extensive use of Hindi to exchange messages in a secure manner through means of communication, e-
13	11	mail, WhatsApp, Facebook and tweet". Shri SM Vaidya, President IndianOil
14	12	11 being one of the prime movers of the nation's growth, provides ocean of opportunities to theāc;
15	13	During my visit to Academy, spent some time with para-badminton champions Amudha, Shanthy, Ruthick, Sivāc;
16	14	Had a wonderful interaction with upcoming Badminton stars of India at the academy of Badminton legend and IOCian āc;
17	15	It has come to our notice that bogus people/agencies are providing fake offers for LPG distributorships. People are advised to visit the nearest area office of OMCs for authentication or log on to
18	16	15 Thank You Honāc"ble for your kind words! Your trust in us is what motivates Team #IndianOil to be the Energy of India #EngineersEmpoweringIndia
19	17	Shri SM Vaidya inaugurated Robotic Process Automation for a key HR & Finance function marking a historic milestone in #IndianOilāc"s digitalization journey to achieve organization-wide automation & employee-friendly
20	18	operations & supporting vision of Digital India.
21	19	#IndianOil refineries energetically rolling out innovative products XP100 MS, Marpol Grade FO, NATO Grade HFHSD, Ethanol Blended MS, Low Sulphur ATF & SKO, Army Grade Diesel through 100% in-house initiative,
22	20	17 empowering an Atmanirbhar Bharat. #EngineersEmpoweringIndia.
23	21	18 Taking forward the vision of Honāc"ble PM of #AtmanirbharBharat, #IndianOilāc"s smart terminal at Una plays a key role in meeting the energy demands of north india. #EngineersEmpoweringIndia
24	22	#IndianOil narvatas refineries fuelling Atmanirbhar Bharat with setting-up of state-of-the-art units & indigenous technologies INDMAx, Octamax, IndadeptG, IndeDiesel, IndeDSN, IndeDSK, Zeosom to power indiaāc"s
25	23	19 dream of being the refining hub of Asia. #EngineersEmpoweringIndia
26	24	20 As engineers, we take inspiration from the works & words of Sir MV. Engineering works are the link b/h the pāc;
27	25	21 has given engineers an unparalleled opportunity to be the creators. Being a part of this great cāc;
28	26	Introducing the all-newāc"SERVOāc" 4T XTRA. Specially engineered atāc" IndianOil's famed R&D centre, it helps your motorbike pump out phenomenal Power, Pick-up and Performance on demand. With advanced 3P Tech at
29	27	22 its core, SERVO 4T XTRA is trulyāc" #Innovationāc" in Motion.
30	28	23 Technology & Engineering are always in forefront to better the Human Life experience. We have āāc;
31	29	24 Finding new ways to solve problems and approach challenges is an opportunity for us as an Engineer while workāc;
32	30	25 Engineers are epitome of human capital who are not just capable of understanding the world but to changeāc;
33	31	26 We engineers have power to do impossible by putting our heart & soul to create something new for the society. āāc;
34	32	27 Say yes to sustainable celebrations! Share with us your eco-friendly celebrations & stand to win some exciting prizes! Follow the to participate
35	33	28 Kudos to Champ for making the Nation proud at #TokyoParalympics!
36	34	29 Nothing brings a smile to our faces faster than a satisfied customer. Here's Mr Umesh Gupta, Chairman, Great Gatsby Club, New Delhi giving his seal of approval to Indane XTRATEJ and its benefits of savings in cooking time
37	35	30 as well as LPG consumption.

Fig. 2 Latest 30 tweets having the Indian Oil Corporation Ltd in different forms

We extracted the recent tweets and filtered those tweets having Indian oil corporation limited in one or the other forms. After that we cleaned, translated, and analysed the tweets. (Fig. 3)



author	data	text
Gurukul87351497	2021-09-16 08:19:15	@CharmenOil @Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
roosekha5839906	2021-09-16 08:27:24	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
regheeshah3	2021-09-16 08:28:31	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
sanjaykumar1	2021-09-16 08:31:04	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
regheeshah3	2021-09-16 08:31:20	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
regheeshah3	2021-09-16 08:31:13	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
roosekha5839906	2021-09-16 08:33:49	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
regheeshah3	2021-09-16 08:35:12	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
regheeshah3	2021-09-16 08:53:47	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
roosekha5839906	2021-09-16 08:56:26	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
MarketsSafe	2021-09-16 08:58:21	Market: Share price of Indian Oil Corp. jumps as Sensex gains 287.35 points - Traders held \$1.5 per cent āc" https://t.co/8eakM5d56
Schicheri	2021-09-16 09:01:00	Data indeed has proven to be the new oil. According to IDC, Indian enterprises are investing in #bigdata analyticsāc" https://t.co/8eakM5d56
roosekha5839906	2021-09-16 09:01:53	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56
roosekha5839906	2021-09-16 09:01:49	@Harddeep5Puri @Rameswar_Tel @PetroneumIndia @IndianOil @Dinh_Loc @DinhLoc (sc) me d phone meāc" https://t.co/8eakM5d56

Fig. 3 Filtered tweets having Indian oil corporation limited in one or the other forms

The steps are discussed in detail as under.

2) Data cleaning: The tweets extracted are in raw format. They contain text and embed URLs, pictures, usernames emoji, user handle, tweet id etc. They also contain misspellings. Hence, a series of pre-processing steps were carried out to remove irrelevant information from the tweets [8]. The reason is that the cleaner the data, the more suitable they are

for accuracy of the sentiment analysis. The tweets were also pre-processed to eliminate duplicate tweets and retweets from the dataset. Each tweet was processed to extract its main message. We used Python's Natural Language Toolkit (NLTK) to pre-process these data, NLTK is a free open-source Python package that provides several tools for building programs and classifying data [9]. We used tweet-pre-processor and re python library. This module provides regular expression matching. A regular expression (Regex) in Python was run to detect and discard tweets with special characters, such as URLs ("http://url"), retweet (RT), user mention (@), and unwanted punctuation. Because hashtags (#) often explain the subject of the tweet and contain useful information related to the topic of the tweet, they are added as a part of the tweet, but the "#" symbol was removed. These cleaned tweets are passed for further process.

3) **Data translation:** After cleaning the data, the tweets in different languages were translated to English. To translate tweets in English for VADER sentiment analysis, we used Googletrans which is a free and unlimited python library that implemented Google Translate API [10]. This uses the Google Translate Ajax API to make calls to such methods as detect and translate. When all the pre-processing steps were complete, the dataset was ready for sentiment classification.

4) **Sentiment Analysis:** For sentiment analysis VADER (Valence Aware Dictionary and sEntiment Reasoner) was used [11]. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. It is fully open sourced under the [MIT License]. VADER takes only English as an input, so all the multi lingual tweets were translated using Googletrans. The translated tweets were passed as an input to VADER which returned the sentiment score. The negative means inclined to negative sentiment. While positive means inclined to positive sentiment. Based on the score we classified the tweets on negative sentiment, positive sentiment, and neutral sentiment. The sentiment score is further classified as -1(negative), 0(neutral) and +1(positive). Each classification is further stored along with the tweets in the csv file.

5) **Report Generation:** Once this overall sentiment analysis is done and tweets are classified. Hashtags are extracted for each classification. Word cloud is generated for company tweets. Based on hashtag count plot is generated which represented count of public on that opinion.

B. Tools Used

Different tools used in the study are described below.

1) **GoogleTranslate:** Googletrans is a free and unlimited python library that implements Google Translate API. It uses the Google Translate Ajax API to detect and translate text. This specific python library was used in this study due to its immense linguistic efficiency because Googletrans is fast and reliable. Additionally, its automatic language detection and ability to complete translations in bulk, allow for multiple tweets to be translated and simplified quite easily. This and countless other elements (such as the presence of customizable service URLs, connection pooling, and HTTP/2 support Python) made it one of the best translation tools available.

2) **VADER (Valence Aware Dictionary and sEntiment Reasoner):** It is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative. VADER not only provides a Positivity and Negativity score, but also tells us about how positive or negative a sentiment is. By informing how positive and how negative a sentiment is, VADER allows users to determine the exact polarity of a sentiment. And in turn, it provides more detailed, complex information to users that a simple binary algorithm could never match.

3) **Word Cloud:** Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analysing data from social network websites [12]. For generating word cloud in Python, modules needed are – matplotlib, pandas, and wordcloud. By revealing the most commonly used words in tweets, Word Cloud further enhances the sentimental analysis done by VADER's algorithm. This is because users now have insights regarding not only sentimental polarity, but also regarding the emotion highlighted by the keywords in a tweet.

4) **Tweet-preprocessor:** Preprocessor is a pre-processing library for tweet data written in Python. When building Machine Learning systems based on tweet data, a pre-processing is required. This library makes it easy to clean, parse or tokenize the tweets. It is a tool which help in cleaning the twitter tweets. Currently it supports cleaning, tokenizing, and parsing of twitter tweets:URLs, Hashtags, Mentions, Reserved words (RT, FAV), Emojis, Smileys, JSON and .txt file support.

C. Code Development

The code development is described in below sub-sections.

1) System requirement: There are few libraries that are needed to be installed first (Fig. 4). The below code installs googletrans, vaderSentiment, Tweepy, wordcloud alsnd tweet-preprocessor.

```
!pip install googletrans==3.1.0a0
!pip install vaderSentiment
!pip install tweepy
!pip install wordcloud
!pip install tweet-preprocessor
!pip install googletrans==3.1.0a0
```

Fig. 4 Code for setup of system environment

2) Importing Tools: After the libraries are installed, there is a need to import the tools that are being used from these libraries. The code given below (Fig. 5) imports the tools needed.

```
import time
import re
import nltk
import requests
import json
import sys
import tweepy
import csv
import preprocessor
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from nltk.stem.porter import *
from wordcloud import WordCloud, STOPWORDS
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from googletrans import Translator

%matplotlib inline
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

Fig. 5 Code for importing different tools required for the analysis

3) Twitter API: We used twitter API to get the tweets. For ease of access, Tweepy was used (Fig. 6). All the credentials were got from twitter developer portal. After all setup and authorization of the credentials, we can extract tweets from twitter.

```
consumer_key = 
consumer_secret = 
access_token = 
access_token_secret = 

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth)
```

Fig. 6 code for setup of twitter API and authorization of the credentials

4) Setup Google Translator: We used google translator for translating the tweets. For setup of google translator and the tweeter preprocessor to clean the tweets by removing emoji, URL, handle etc. except hashtags. we used following code (Fig. 7)

```
sns.set()
analyser = SentimentIntensityAnalyzer()
translator = Translator()
preprocessor.set_options(preprocessor.OPT.URL, preprocessor.OPT.EMOJI, preprocessor.OPT.MENTION, preprocessor.OPT.SMILEY)
```

Fig. 7 Code for setup of google translator and tweeter preprocessor

5) Sentiment Analysis: We get the polarity score using Vader to judge how much negative or positive is the content. Score ranges from +1 to -1. We created a function that will give me the classification of the sentiment (Fig. 8). This function takes any tweet and will return us the sentiment of the tweet i.e., positive negative or neutral.

```
def sentiment_analyzer_scores(text):
    trans = translator.translate(text).text
    return sentiment_scores(trans)

def sentiment_scores(text):
    score = analyser.polarity_scores(text)
    lb = score['compound']

    if lb > -0.25:
        return 1
    elif (lb < -0.25 and lb > -0.5) :
        return 0
    else:
        return -1
```

Fig. 8 Code for classification of sentiment of the tweets

6) Translation: Before passing to Vader, we need to translate text to English as Vader can predict sentiment for English language only. We need to translate the multi-lingual tweets to generate word cloud of the tweets. The function in Fig. 9 takes the tweets list and translate them which were then used to generate the word cloud. This gave us the major words used.

```
def translate_tweets(lst):
    res=[]
    for twt in lst:
        twt = translator.translate(twt).text
        res.append(twt)
    return res
```

Fig. 9 Code for translation of tweets and generating word cloud

7) Analysis: This is a function that analyses the tweet input for this tweet list and it will generate the graph showing count of positive, negative and neutral tweets (Fig. 10). This function is also used to analyse the user tweets.

```
def anl_tweets(lst, title='Tweets Sentiment'):
    sents = []
    for tw in lst:
        st = sentiment_scores(tw)
        sents.append(st)
    ax = sns.distplot(sents, kde=False, bins=3)
    ax.set(xlabel = 'Negative Neutral Positive', ylabel='Number of Tweets',title="Tweets from "+title)
    return sents
```

Fig. 10 Python code of function for analysing the user tweets.

8) Extracting Tweets: This is the function that extracts the tweets from the keyword given by the user. This function takes user handle and the count. And return the list of tweets with given count from the handle provided. (Fig. 11)

```
def list_tweets(user_id, count):
    tweets = api.user_timeline("@" + user_id, count=count, tweet_mode='extended')
    tw = []
    for t in tweets:
        tw.append(t.full_text)
    return tw
```

Fig. 11 Python code of the function to extract list of tweets

9) Cleaning Tweets: The tweets extracted is cleaned so that we can analyse it to get good quality of results. Below is the function (Fig. 12) that cleans the tweet, it takes tweet list and returns the cleaned tweet list.

```
def clean_tweets(lst):
    for l in lst:
        l = preprocessor.clean(l)

    # remove twitter Return handles (RT @xxx:)
    lst = np.vectorize(remove_pattern)(lst, "RT @[\\w]*:")

    # remove twitter handles (@xxx)
    lst = np.vectorize(remove_pattern)(lst, "@[\\w]*")

    # remove URL Links (httpxxx)
    lst = np.vectorize(remove_pattern)(lst, "https?:/[A-Za-z0-9./]*")

    # remove new line
    lst = np.vectorize(remove_pattern)(lst, "[\\n]*")

    # remove special characters, numbers, punctuations (except for #)
    lst = np.core.defchararray.replace(lst, "[^a-zA-Z#]", " ")

    return lst
```

Fig. 12 Python code for cleaning the tweets

10) Hash tag Analysis: Hash tags are very important as they are the keywords. We need to analyse the hashtag. For this created the following function (Fig. 13) that gives the visual output of the hashtag. This returns the count graph with count of major hashtags.

```
def hashtag_analysis(HT_list):
    if len(HT_list):
        a = nltk.FreqDist(HT_list)
        d = pd.DataFrame({'Hashtag': list(a.keys()), 'Count': list(a.values())})

        # selecting top 10 most frequent hashtags
        d = d.nlargest(columns="Count", n = 10)
        plt.figure(figsize=(16,5))
        ax = sns.barplot(data=d, x= "Hashtag", y = "Count")
        ax.set(ylabel = 'Count')
        plt.show()
    else:
        print ('NO HASHTAG PRESENT')
```

Fig. 13 Python code for Hashtag analysis

III. RESULTS AND DISCUSSION

The exact functioning and application of the overall system is described in detail in this section.

A. Translation

The Image given below (Fig. 14) shows the how the translation of Hindi tweet to English. In this way tweet in any language can be translated to English.

```
txt = 'जीवनपर्यंत जन कल्याण के लिए समर्पित रहे कल्याण सिंह जी के अंतिम दर्शन किए। उनके परिजनों से मिला। प्रभु श्रीराम उनके परिजनों को इस अपार दुख को सहने की  
txt = translator.translate(txt).text  
print(txt)
```

He had the last darshan of Kalyan Singh ji, who was devoted to public welfare throughout his life. Met his family members. May Lord Shri Ram give strength to his family members to bear this immense loss.

Fig. 14 Screenshot showing translation of Hindi tweet to English

B. Analysis Of tweets

The latest 100 IOCL tweets were translated and analysed. After analysis they were classified into -Negative, Neutral and Positive, based on the sentiments. After classification the count plot was drawn for the Same (Fig. 15).

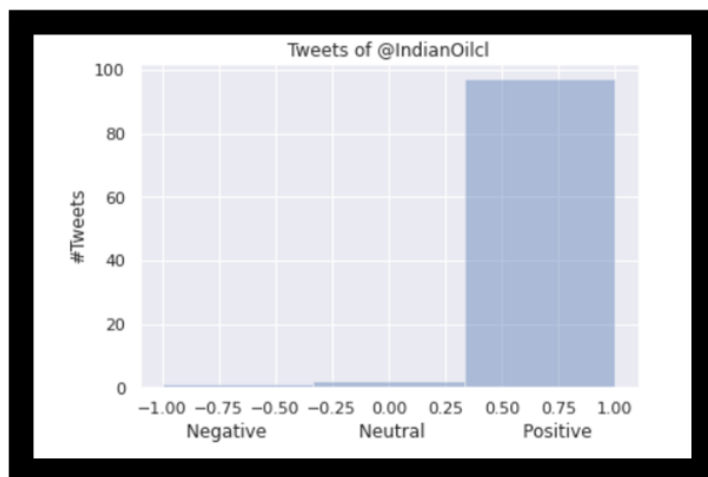


Fig. 15 Classification of the tweets based on sentiments

C. Public Opinion Analysis

The real time public opinion was analysed to judge the reputation of IOCL by drawing the count plot based on the sentiment (Fig. 16).

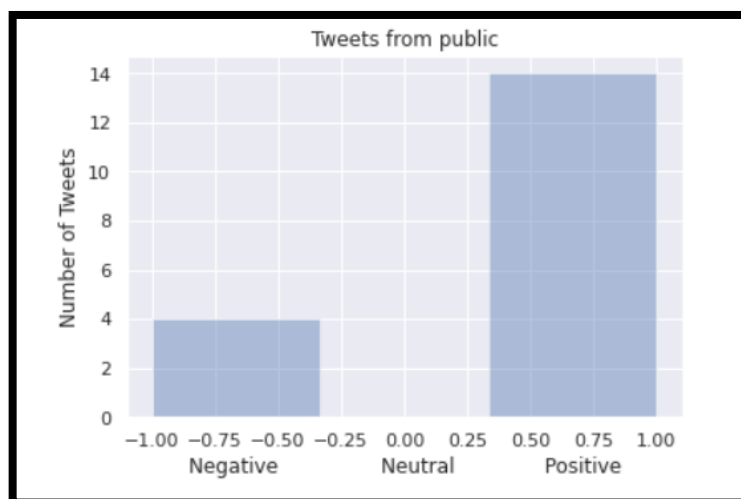


Fig. 16 Real time public opinion analysis

D. Word Cloud

The image below (Fig. 17) shows the word cloud for latest 100 tweets of Indian Oil Corporation Limited (IOCL).

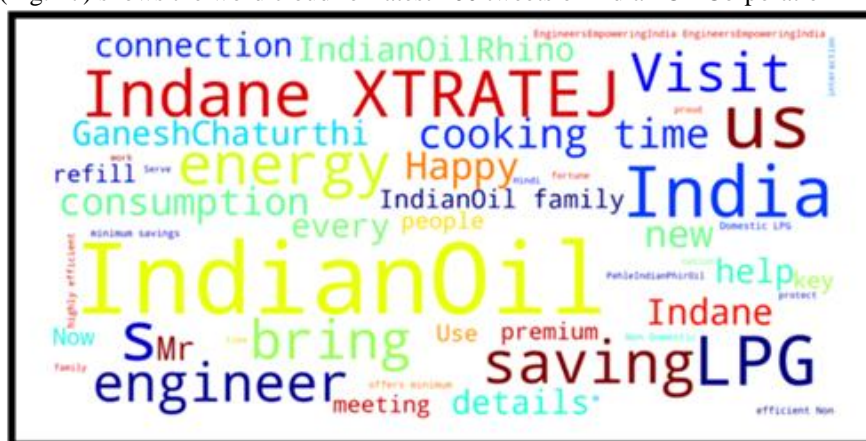


Fig. 17 Word cloud for latest tweets of IOCL

E. Word Cloud Public

The word cloud was drawn for the most common words used by public in their tweets regarding IOCL (Fig. 18). This contains both positive and negative tweets.

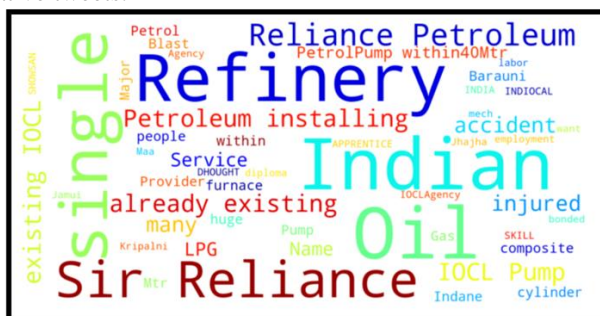


Fig. 18 Word cloud for the most common words used by public

We draw the word cloud from all the positive tweets based on sentiment (Fig. 19).

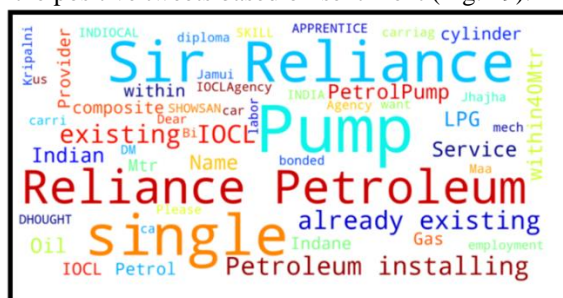


Fig. 19 Word cloud for positive tweets by public

The word cloud from all the negative tweets was then drawn based on sentiment (Fig. 20).



Fig. 20 Word cloud for all negative tweets by public

F. Hashtags

The count plot of the hashtags was drawn. Hashtags represents the main key points and their count plot represents number of people with the same opinion. From positive count plot (Fig. 21) we can interpret what are the things public like where the company can promote further to contribute better for the public.

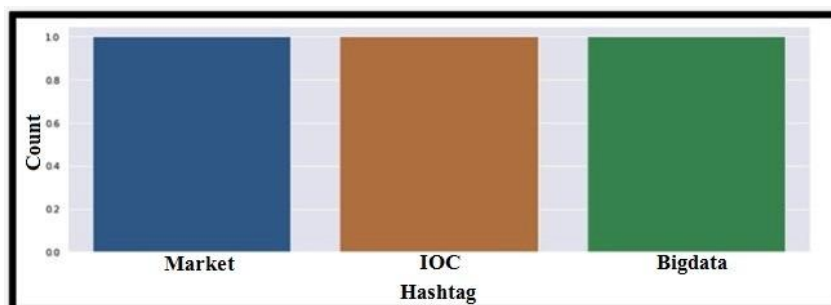


Fig. 21 The positive count plot of hashtags

The negative count plot (Fig. 22) helps the company to interpret what public dislikes. This helps to know the points of action that need to be improved further to strengthen the overall reputation of the firm.

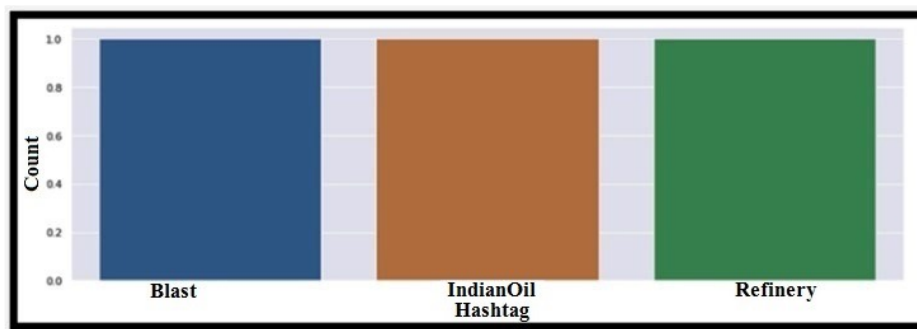


Fig. 22 The negative count plot of hashtags

IV. CONCLUSION

In this study, the NLTK and the VADER analyzer were applied to conduct a sentiment analysis of Twitter data. The tweets in different Indian languages using different keywords were collected for Indian Oil Corporation Ltd using twitter API. It is then converted into English language using “Google Translator” to generate word cloud of the tweets. The hashtag is then extracted from each handle. The count of mostly used hashtags was used to judge the overall reputation of the company. The case study was to judge the public opinion about Indian Oil Corporation Limited based on the 100 most recent tweets in different languages. The results indicated that the VADER Sentiment Analyzer was an effective choice for sentiment analysis classification using Twitter data. VADER easily and quickly classified huge amounts of data. The positive opinion helped to interpret what are the things public like about the organization so that it can promote further to contribute better for the public. Whereas from negative opinion it was easy to interpret what public dislike about the organization that needs to be improved further to strengthen the overall reputation of the firm. The tool developed in the study can be used in future to study public opinion on any topic of interest by giving a suitable keyword.

REFERENCES

- [1]. B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, Twitter power: Tweets as electronic word of mouth, J. Am. Soc. Inf. Sci. Technol., vol. 60, no. 11, pp. 2169–2188, 2009.
- [2]. V. Kharde and P. Sonawane, Sentiment analysis of twitter data: a survey of techniques, arXiv Prepr. arXiv1601.06971, 2016.
- [3]. P. Selvaperumal and D. A. Suruliandi, A Short Message Classification Algorithm for Tweet Classification, Int. Conf. Recent Trends Inf. Technol., pp. 1–3, 2014.
- [4]. T. Singh and M. Kumari, Role of Text Pre-processing in Twitter Sentiment Analysis, Procedia Comput. Sci., vol. 89, pp. 549–554, 2016.
- [5]. <https://pypi.org/project/vaderSentiment/>
- [6]. “Sentiment Analysis” <https://brand24.com/blog/sentiment-analysis/>
- [7]. <https://www.tweepy.org/>
- [8]. <https://pypi.org/project/tweet-preprocessor/>
- [9]. Shihab Elbagir and Jing Yang, “Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment Proceedings of the International MultiConference of Engineers and Computer Scientists 2019 IMECS 2019, March 13-15, 2019, Hong Kong, ISBN: 978-988-14048-5-5 ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)
- [10]. <https://pypi.org/project/googletrans/>
- [11]. J. H. E. Gilbert, “Vader: A parsimonious rule-based model for sentiment analysis of social media text,” in Eighth International Conference on Weblogs and social media (ICWSM-14). Available at (20/04/16) <http://comp.social.gatech.edu/papers/icws14.vader.hutto.pdf>, 2014.
- [12]. <https://pypi.org/project/wordcloud/>