IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676, p-ISSN: 2320-3331, Volume 17, Issue 2 Ser. II (Mar. – Apr. 2022), PP 01-05
www.iosrjournals.org

Sentiment Analysis of Twitter Dataset-AText Mining Approach on Oil Prices

Saad Nabeel, *Master Student, North China University of Water Resources and Electric Power, Zhengzhou*
Hongtao Guo, *North China University of Water Resources and Electric Power | NCWU · Department of Information Engineering Master of Engineering*

Abstract *The world population, more than 7 billion people has faced COVD 9 which placed people in quarantine and impacted the global supply chain. Moreover, it hindered the transportation as well as stopped every type of mobile activities. The freeze in these activities also impacted the global oil market. Theoil prices on 20 April 2020 dropped below zero for the first time in the history due to the access supply of crude oil. This rapid decrease in oil price resulted in anxiety among the people. Now a day's social media platform plays an impactful role and websites like Twitter can be used to share thoughts and feelings. Sentiment analysis is the process of classifying text into sentiments. The purpose of this study is to analyse the tweets gathered during 9 March 2020 to 15 May 2020. POS tagging and sentiment score of 0.1 million tweets has done using Lexicon-based approach. High level programming language, python based Natural language toolkit, is used classify the tweets. It is observed that maximum number of tweets reflective negative sentiments.*

Date of Submission: 17-03-2022 Date of Acceptance: 02-04-2022

I. Introduction

Now a day's millions of people use social medialike Twitter, Facebook, WhatsApp and WeChat in daily life. Social media are getting verymuch popular among people. In social media we cancommunicate with each other by using internet. Internet is a trend in the world. Mostly internet user spends more time with social media by the total time spent on mobile device. The comments and reviews in social networks and web are considered as one of the most trustworthy and powerful sources of information. It is a fact that analysing social media network posts and web content is very important [1]. Twitter is one of the most popular social media which, according to the statistic, currently has over 330 million user accounts. With the rapid growth of the World Wide Web, people are using social media such as Twitter which generates big volumes ofopinion texts in the form of tweets which is available for the sentiment analysis [2]. In Twitter people can communicate and share their views by tweets on avariety of topics. In Twitter user can share information with tweets. The word limit of tweetsare 140 characters [3]. People discuss current social issues, complain or express positive or negativesentiment for products they use every day. Sometimes people react in neutral way. A collectionof tweets is used as the primary corpus for sentimentanalysis, which refers to the use of opinion mining or natural language processing [4]. Twitter with more than 500 million users and million messages per day, has quickly become a valuable asset for organizations to invigilate their reputation and brands by extracting and analyzing the sentiment ofthe tweets by the public about their products, services market and even about competitors [5]

Sentiment analysis has many applications for different domains. In businesses to get feedbacks for products by which companies can learn users' feedback and reviews on social media. Organizations use sentiment analysis as a tool to know about what people think about their product.

The fluctuation in global oil price has significant impact on economy as the crude oil is themost traded commodity. High oil price may slow down the economic growth due to higher expenses on transportation. The crude oil price may crash dueto fears of a financial downturn and sharp decrease in oil price are not adequate for the growth of financial development. If oil price fall adequately, it can cause some oil firms to leave business resulted in substantially increase in debts. The recent oilprice crash of 2020 is result of the financial downturn and costs have fallen so far that many oil firms has been constrained bankrupt and causedsubstantially decrease in investment in oil sector andjob losses.

Highly influential people and public users shared their point of views on twitter regarding oil prices. In this study the sentiments analysis of oil price data has been conducted by using machine learning algorithm based on python.

Sentiment Analysis

Sentiment Analysis on Twitter data is increasing popularity of social networking and Sentiment Analysis. Sentiment analysis has been handled as a Natural Language Processing (NLP). NLP techniques are based on Machine Learning which uses a general learning algorithm combined with a large sample of data to learn the rules. NLTKis a Python library to build a program to process human language data.

There are many studies conducted related to sentiment analysis. In previous studies most active research was on the area of user-generated material which was discussed on social media, blogs and forums. Since most sentiment analysis studies use or depend on machine learning techniques which has been applied in most of sentiment analysis, therefore there is available unlimited data. In 2015, Aldahawi collected the twitter data of business companies and then investigated the difference in sentiment classification between human classification and machine learning classification in python [6]. Aliza Sarlan, Shuib and Chayanit conducted experiments on twitter data in which they simply extracted the tweets in Jason format and used python lexicon dictionary to assign polarity to the tweets [7].

In these studies, regarding Sentiment analysis two things has been given emphasized. Subjectivity or objectivity of the text and to identify the polarity oftext.

There have been done sentiment analysis on movie responses by Pang et al. (2002) [8] on product views by Dave et al., (2003) [9], Na et al., (2004) have also conducted a study on views about the products [10], and on news sentiment analysis has been done by Godbole et al., (2007) and Bautinet al., (2008) [11] [12].

Objectives of the Study

- Study the sentiment analysis of web 2.0websites.
- Learning the sentiments of users regardingoil price from their tweets.
- Building a Classification Model to predictthe sentiment of the tweets.
- Evaluation of the model using evaluationmetrics.

Method of the Study

Data Collection

Tweets data set was collecting using twitter API. The raw data was scraped in the form of (Java script Object notation) JSON file. Total 0.1 milliontweets were scraped from 2020-03-09 22:24:00 UTC till 2020-05-15 23:49:00 UTC.

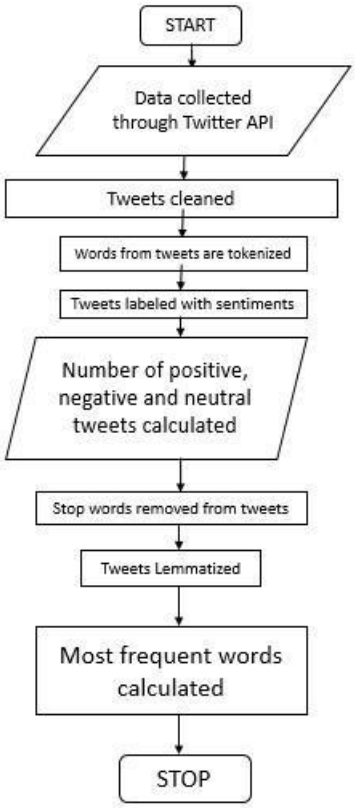

Figure 1: Overall Structure of the state of art classifierbased proposed model

Total 12 columns/features associated with each tweets containing hashtag of #oilprice were returned by the API but in this study, our main area of interest is tweets text. We can see that initially tweets count was stable around 2,000 tweets per daybut there was sharp increase in tweet count on 20 April 2020. Around more than 14,000 tweets were posted on this day followed by 4,000 in next two consecutive days. Then they were again stabilized at the rate of 2,000 tweets. The reason behind this sharp increase of tweet was the negative oil prices that compelled user to tweet more about oil price.

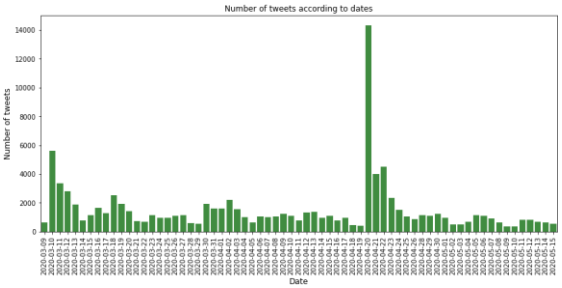

Figure 2: Tweet count over the period

Data Pre-processing

Data pre-processing plays an important role in any process of analysis. First of all, data was cleaned from repeated values. Sometimes, text contain unnecessary characters and symbols thatneeded to be eliminated to get ready our data for analysis. For this purpose, tweets were cleaned, and username, hashtags, URLs and punctuations were removed for further analysis.

POS-Tagging and Sentiment Score

Pang and Lee, (2008) described that POS isused to disambiguate sense which in turn is used to guide feature selection [13]. For example, with POS tags, we can identify adjectives and adverbs which are usually used as sentiment indicators. Accordingto Turney, (2002) adjectives performed worse than the same number of uni-grams selected on the basisof frequency [14].

Sentiment Analysis can be done using twodifferent approaches. The first one is lexicon-based approach and other is Machine learning model. In lexicon-based model, dictionary of words associated with their semantic scores is used to check the polarity of tweet.

Semantic score associated with each word was added up to calculate the score of complete sentences. The positive score indicates the positive tweet, and negative score denotes negative tweet.

Porter Stemmer module from NLTK library was used to reduce a word to its word stem by removing prefix and suffix. While WordNet Lemmatizer was used to group together different forms of word. Built in stop words from English library was used to remove the most common words.

Each corpus of tweet was tokenized, after applying porter stemmer word net lemmatizer. Then POS tagging and positive and negative sentiment score associated with each token were added. If the positive score is greater than negative, then the tweet tagged as positive. In case of lower positive score, tweet is tagged as negative. While the corpus havingzero positive and negative score is tagged as neutral.

Few examples of tweet text along with their positive, negative and sentiment score have given below. As the positive score was greater than negative the tweet was tagged as positive. It is shown in Table 1.

Table 1 *Few examples of postive, neutral and negative* tweets

Pos score	Negscore	Sent score	Tweet text
0	0.125	-1	Oil Price Charts https://oilprice.com #oilpricespic.twitter.com/yWxaolcnJ6
0	0	0	China Refinery Runs Jump As Nation Emerges From Lockdown | http://OilPrice.com https://oilprice.com/Latest-Energy- News/World-News/China-Refinery-Runs-Jump-As-Nation-Emerges-From-Lockdown.html â€¦ #oilprice
0	0.25	-1	Analyzing Bitcoinâ€™s Correlation to the #Oil Price - #Analysis #BitcoinAnalysis #BTCUSD #Premium - https://paulcrypto.com/2020/05/15/an alyzing-bitcoins-correlation-to-the- oil-price/ â€¦pic.twitter.com/P7SRGBNEuq

0.5	0.75	-1	why oil price declined so sharply? what Saudis did and what USA facedin post corona situation? read the science, geography and politics of crude oil: https://valuablemineral.blogspot.com /
1.75	0.125	1	I read one commentary that actuallysaid Russia, and next Iran (both having already made public the switch to EU $), were the targets ofthat orchestrated oil price drop. ButRussia & Iran's economies were strong enough to sustain the blow while Venezuela...
3	1.875	1	Yet in every credit downgrade Alberta has received (NDP&UCP) credit agencies have cited over dependence on resource revenue &revenue projections not likely to berealized do to overly optimistic oil price projections. How Alberta budgets its revenue does affect it credit rating pic.twitter.com/7x38GFZe7C
0.375	0.25	1	The Nigerian Economic Summit Group | Blog Post: Macro-Economic Outlook: COVIDâ€“19, Global Oil Price and The Nigerian Economy https://www.nesgroup.org/blog/Macro-Economic-Outlook-- COVID%E2%80%9319,-Global-Oil-Price-and-The-Nigerian-Economy#.Xr8kNpguh6k.twitter â€¦
0	0.625	-1	When will all this stop after stealing from us now oil price has fallen andis stamp duty charge nija govt hmmmmmmm
0.25	1.125	-1	Heading home tomorrow from Angola into Furlough. Major oil company with one days notice endedthe one year support contract. We immediately sailed to port. Reality ofthe oil price caused by over supply & difficulties in travelling to vesselsworldwide due to COVID-19.
0	0.125	-1	#Oil price update: $32.74 (#Brent $Crude) #oilprice source http://oilpriceapi.com

Word cloud Visualization forsentiments

In Word cloud visualization method text is represented in a chart, where the words having more importance are shown in written bold fonts and largetext, on the other hand less important words are written in simple fonts and small text. After the twitter analysis, the following word cloud isgenerated. It shows the words that are frequently used in the twitter account for oil prices.

Word cloud analysis shows that the word “Oil Price” was most frequently used in positive, negative and neutral way because the study is relatedto oil prices. The majority of the words that have been portrayed in each of the sentiments has been visualized using the Word Cloud modules. The othermost frequently used words by the users are “drop”,“collapse”, “market”, “covid ”, “crude oil” and so on. So, we can clearly get a visual understand of what actually is going on in the user’s mind by looking at those frequently used words. Fig. 3a shows that most used words in positive sentiments that have been represented by this Word Clouds. Fig. 3b describes the neutral responses where Fig. 3c shows the most used negative words regarding the oil price. The word “war” is frequently used inpositive responses as well as in negative responses by the users.

II. Conclusion

We have discussed the mechanism to analysis sentiment of user in tweeter regarding crude oilpricing based on the reviews it is found most of the user have negative sentiment regarding crude oil prices. On 20 April 2020 sudden increase in tweet count and negative sentiment regarding oil market crash is a major concern. The purposed methodologyof extracting data from twitter and analysing and visualising the sentiment regarding oil price may help government and other organization in making strategies related to this business as well as boostingthe economy.

Figure 3: Word Clouds Visualization for Positive (a), Neutral (b), Negative(c)

References

[1]. Ray, S. K., & Shaalan, K. Using Freeware Resources to Analyse Sentiments in Social Media. In *Developments of E-Systems Engineering (DeSE),2015 International Conference on* (pp. 204-209). IEEE. 2015

[2]. P. Lai, "Extracting Strong Sentiment Trend fromTwitter". Stanford University, 2012.

[3]. A. K. Jose, N. Bhatia, and S. Krishna,"TwitterSentimentAnalysis".NationalInstituteofTechnologyCalicut,2010.

[4]. M.Rambocas, and J. Gama, "Marketing Research: The Role of Sentiment Analysis". The 5thSNA-KDD Workshop'11. University of Porto,2013.

[5]. Saif, H, Y.He, and H. Alani, "SemanticSentiment Analysisof Twitter," Proceeding of the Workshop on Information Extraction and Entity Analytics on Social Media Data. United Kingdom: Knowledge Media Institute, 2011.

[6]. Aldahawi, H. (2015). Mining and analysing social network in the oil business: Twitter sentimentanalysis and prediction approaches (Doctoral dissertation, Cardiff University).

[7]. Aliza Sarlan, Chayanit Nadam and Shuib Basri"Twitter Sentiment Analysis". Universiti TeknologiPETRONAS Perak, Malaysia, 2014.

[8]. Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in NLP, pages 79–86, Philadelphia,PA.2002.

[9]. Dave, K., Lawrence, S., and Pennock, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews.2003.

[10]. Na, J.-C., Sui, H., Khoo, C., Chan, S., and Zhou, Y. Effectiveness of simple lin-guistic processing in automatic sentiment classification of product reviews. Conferenceof the International Society of Knowledge Organization (ISKO), p. 49–54. 2004.

[11]. Godbole, N.; Srinivasaiah, M.; and Skiena, S. Large-Scale Sentiment Analysis for News and Blogs. In ICWSM'07.2007.

[12]. Bautin, M., L. Vijayarenu, and S. Skiena. International sentiment analysis for news and blogs.In Proceedings of the International Conference on Weblogs and Social Media (ICWSM).2008.

[13]. Pang, Bo and Lillian Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135.Proceedings ofthe World Wide Web Conference.2008.

[14]. Turney, P. D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for ComputationalLinguistics.2002.