Drug Usage Analysis by VADER Sentiment Analysis on Leading Countries

Sandipan Biswas* & Shivnath Ghosh†

Abstract

On Twitter, people from all parts of the planet will build their opinions and take feed from the opinions, which produces around five hundred million tweets daily that amount to 8TB of information. Data scrapped from Twitter may be helpful if analyzed as we can extract vital information via sentiment analysis. Opinions or comments regarding any news or launch of a product of quiet trend may be ascertained well in Twitter information. We aim to analyze tweets on the use of drugs for the treatment of COVID-19 diseases. In twitter sentiment analysis, we tend to categorize those tweets into positive and negative sentiment and cluster them along or a pack of clusters. We have conducted a study and found that the application will quickly and efficiently distinguish numerous tweets on the idea of their sentiment scores and proportion and may notice weak and powerful positive or negative tweets once clustered with results of various dictionaries and establish powerful support on our assumption. This paper surveys the polarity activity exploitation using VADER sentiment analysis on the utilization of drugs for COVID-19 treatment.

Keywords: Clustering, Opinion, Sentiments, Tweets, VADER

* Department of Computer Science & Engineering, Brainware University, West Bengal, India; sandipan_diet@rediffmail.com

† Department of Computer Science & Engineering, Brainware University, West Bengal, India; dsg.cs@brainwareuniversity.ac.in

1. Introduction

Social websites which are different forms like blogs, icon and forum sharing, video sharing social networks, microblogs, etc., have used various online social networking sites like Facebook, Instagram, YouTube, LinkedIn & Twitter. These sites and mobile apps share various people's views across the world.

On these social sites, different individuals worldwide can express their discussions, and comments in various styles like text, image, video, and emoji [1, 2]. Social media, with a huge source of knowledge, can gather user opinions and various polls regarding expression. Microblog has become the simplest familiar and, therefore, the source of various data [3]. Twitter is a microblogging service that enables users to share their emotions, and reply within a short time frame as tweets [4]. It provides a fashionable supply of knowledge that is utilized in various scientific studies using sentiment analysis to extract and analyze data that are expressed as tweets on various topics like market, election, and share-trade prediction.

Linguistic Inquiry & Word Count (LIWC) [5] is one of the tools of text extraction. Most of these tools require programming, here in our work, we have used Valence Aware wordbook and Sentiment Reasoning (VADER) [6], which work on sentiment analysis of tweets on COVID-19 pandemic situation on the usage of drugs for the treatment in various countries.

2. Related Work

Here, we have defined some references containing text mining and sentiment analysis approaches.

Dhirajgurkhe et.al [7] in their paper, have projected Twitter information that is collected and processed from numerous sources and removed the content which does not hold any polarity.

M. Bouazizi et.al [8] in their paper has used the tool SENTA for sentiment analysis of the tweets and calculating score according to sentiment.

Geetika Gautam et.al have used a review classification on tweets and leveled the data. They have used primary algorithms like Naïve Bayes, SVM, and Maximum Entropy on the platform NLTK, and Python.

Akshay Amolik et.al [9], in their paper, have used Twitter sentiment analysis on the Hollywood movie industry, and they compared Naïve Bayes and SVM algorithms on accuracy classification.

Subhabrata Mukherjee et.al [11] have used a hybrid sentiment analysis tool TwiSent in which spell check and linguistic handler have been defined. Davdov et.al [12], in their paper, has used a supervised sentiment analysis technique on twitter information. Neethu et.al [13] have used machine learning techniques on SVM, Naïve Bayes, and Maximum entropy on the MATLAB platform for classification data.

Pulkit et.al [14] have designed a structure for tweets on terrorism attack and their activities. Lots of tweets after the attack with the #tag have been used on the Naïve Bayes algorithm, which was used on huge data, and analysis has been performed for characteristics of the comments. Ali Hasan et.al [15] have projected a hybrid approach where the tweets are followed by the #tag on political trends. Several Urdu tweets are translated to English for analysis, where the Naïve Bayes and SVM approaches are used to build a structure.

Hetu et. al [16] have designed and projected a sentiment analysis method on the python platform, and the data source was collected from Kaggle. Classification of users' emotions on positivity and negativity was done for accurate finding.

Feddah Alhumaidi Al Otaibi et al. [41] have structured a model both on supervised and unsupervised algorithms. They have used Twitter API for extracting 7000 tweets which are based on comments on McDonald's and KFC quality. The analysis was performed on the R programming language platform.

3. Sentiment Analysis on Twitter

This system primarily consists of the following stages. The stages square measure mentioned in Figure 1. shortly.

3.1. Input (Keyword)

This is the first stage of sentiment analysis on Twitter. Within the starting, we'll opt for the topic, and associated with that subject, we'll gather the tweet.

3.2. Tweets Retrieval

During this step, tweets are retrieved, it is often in any type like unstructured, structured, and semi-structured no changes are made to these data, and so these inputs are placed in table format in unaltered format.

3.3. Pre-processing

Within the tweet pre-processing, information is filtered by removing irrelevant information, inconsistent information, and yelling information. It is a way to consider each word as a feature of measure to capture whether a word exists or does not exist in a sentence.

3.4. Sentiment Detection

These measures numerous applications of sentiment analysis wherever it's necessary to seek out the sentiment, like tweet classification and tweet mining [18]. This step is important for determining sentiment weightage, which helps in the classification of sentiments.

3.5. Sentiment classification

Algorithmic rule Sentiment analysis is often classified into two approaches i.e., supervised learning and unattended learning. In supervised learning, Naïve Thomas Bayes, SVM, and most entropy square measures accustomed to perform the sentiment analysis [19-28].

3.6. Evaluation

The final output is analyzed to require a call on whether or not we must always prefer it or not [29-33].
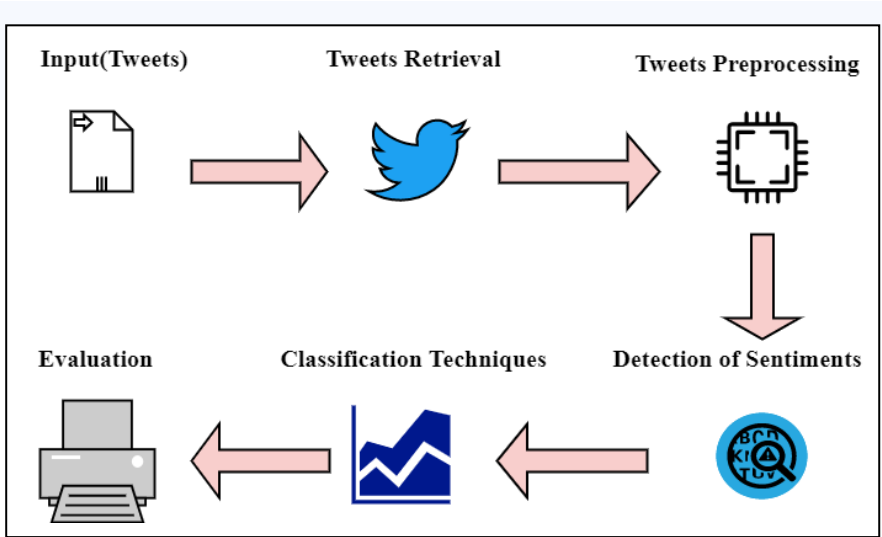

Fig.1: Stages of Sentiment Analysis

4.   Proposed Model Component Description

Twitter may be a social networking platform that enables its users to send and skim micro-blogs of up to 280-characters called "tweets". It allows registered users to browse and post their tweets through the net, short message service (SMS), and mobile applications.

As a worldwide period of a time communications platform, Twitter has quite four hundred million monthly guests and 255 million monthly active users around the world. Twitter's active cluster of registered members includes World leaders, major athletes, star performers, news organizations, and amusement retailers. It's presently accessible in additional than thirty-five languages.

Twitter was launched in 2006 by Jack Dorsey, Evan Williams, Biz Stone, and patriarch Glass. Twitter is headquartered in San Francisco, California, USA.

In Table 1, we've depicted a sample table of the scrapped information in table-wise wherever on a date and tweet texts square measure depicted for our result analysis. The whole analysis has been done with a full data set of entire date-wise scrapped information.

Table 1. Sample Tweet comments with post date and time

Date	Tweet Text
20-04-20 05:56	Thank you to our healthcare workers on the front-lines of battling COVID-19! The sacrifices you make, every day and especially during this pandemic. We applaud your dedication, commitment, courage and hearts!
20-04-20 05:56	Watch Poorna's story of covid-19 in @WIONews Indian news. A cautionary tale and a story of hope. Simply can't wait to see this amazing man again! RT @WIONews: #EXCLUSIVE| Dr. Poorna Gunasekeratels #WION how he survived #coronavirus The UK-based doctor hails his former students Listen in! #coronavirusupdates #UK
20-04-20 05:56	How students and parents are coping with home-school midst COVID-19 pandemic
20-04-2020 05:56	At noon today in Harrisburg, PA, thousands of protesters plan to gather They want to return to work "safely" now with over 32,000 cases of #COVID19 in PA, they fail to see the irony and the harm of their mass protest #wtpPA2020
20-04-20 05:56	Eastern Passage NSLC temporarily closes after employee tests positive for COVID-19 via @HalifaxToday
20-04-20 05:56	Good #MondayMorning to you! Here's a great restaurant take out list for any of my local SWFL connections: #coronavirus #Foodie #Naples #FtMyers
20-04-20 05:56	¡Standardized COVID-19 comparison among countries! Here it comes our research with @methaodos and @academia_joven about #COVID19 (it will be updated on Mondays and Thursdays):
20-04-20 05:56	In a tête-à-tête with the @FinancialXpress, our Founder Chairman @vineetnayar answers the glaring question: How do we respond to the Covid-19 pandemic which has resulted in schools shutting down, affecting the education of millions of children.
20-04-20 05:56	Live Q&A TODAY at 11:00A Central: @RWalensky of @MassGeneralNews discusses #publichealth strategies & policies for lifting shelter-in-place and quarantine and for returning the U.S. to a new post-first-wave normal. #JAMALive #covid19 #coronavirus

4.1. Tools

- Language Toolkit (NLTK)

NLTK may be a free ASCII text file Python package that has many tools for building various programs and then classifying them based on knowledge [34, 35].

- Valence Aware lexicon and Sentiment Ratiocinator (VADER)

VADER is a dictionary and rule-based sentiment analysis tool personalized specifically to the emotions expressed on social media[36, 37].

4.2. Pre-processing

Accents and punctuation have been removed from all instances of the information set , and various preprocessing techniques have been applied:

• Lemma extraction [38]: Lemma extraction is a method of output by lemma called "lemma", and it is not the root that is the output of the stem, but the word of the root. After extracting the lemma, get a valid word that means the same thing.

• Stemming: Stemming is a technique used to remove affixes and extract basic shapes. This is exactly the same as chopping a tree branch to the trunk.

• A Part of Speech (PoS) Tagging [39, 40]: Part-of-speech tagging (POS tagging or PoS tagging or POSTAlso known as grammatical tagging, words marking up a word in a text (corpus).

• Summarization: This is a core data mining concept that includes techniques for finding a compact description of a dataset.

4.3. Algorithm and Flowchart of our Sentiment Analysis Model

In this section, we have described our sentiment analysis model in Figure 2. In the flowchart, at first, the comments are collected from online data sources (we have collected data from Twitter). Then using special keywords, data is accessed from the data source. Next, the sentiment analysis and mining process is done in three

stages, the first stage is feature extraction (transforming raw data into numerical features saving the information in the original data source), the second sentiment analysis, and last polarity classification (the orientation of the expressed sentiment whether the text expresses the positive, negative or neutral sentiment).
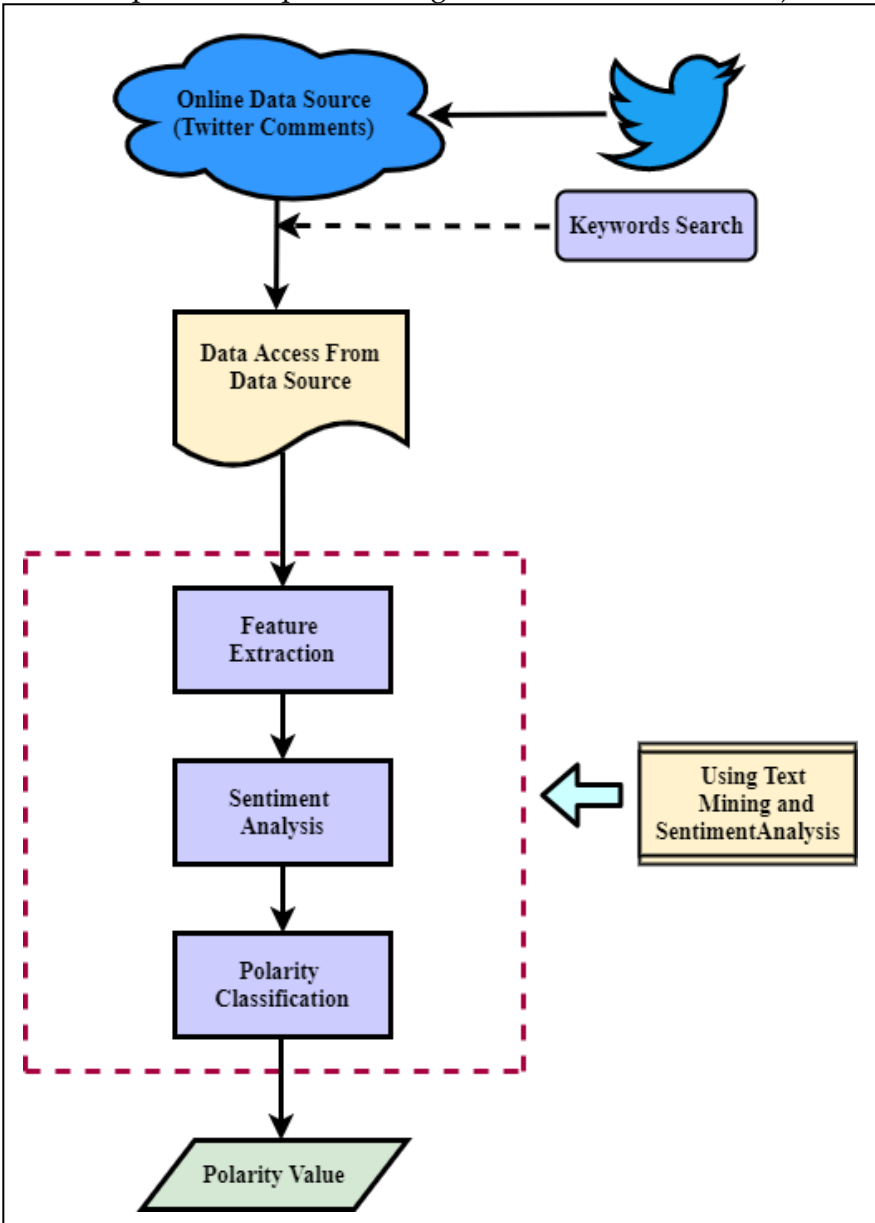

Fig 2: Algorithm Flowchart

Here in Algorithm 1, we have also discussed the steps to carry on the analysis process on the news articles. In the algorithm twitter comments, classifications are done step by step. Generally, text comments include nouns, adjectives, and adverbs. These comments are given to outputs into three categories, sentiment polarity score/value greater than or equal to 0.5 which gives, sentiment polarity score/value less than or equal to -0.05, and sentiment polarity score/value equal to zero. The algorithm output is divided into three parts one positive value, one negative value, another is neutral.

ALGORITHM 1:	Twitter Comments Sentiment Classification
Input:	Text File (Twitter Comments which include Nouns, Adjectives, Adverbs)
Output:	Values = Positive, Values = Negative, Values = Neutral
Begin:	1. Sentiment Analysis () ← File 2. For each row in rows 3. if Sentiment Polarity Score(line) > = 0.05 then 4. Sentiment ← Positive 5. else 6. if Sentiment Polarity Score (line) < = - 0.05 then 7. Sentiment ← Negative 8. else Sentiment ← Neutral 9. end 10. end 11. end 12. end
End:	

5. Result Analysis

For the result analysis, we have proceeded by analyzing the information which was collected from twitter information. Collected information is then processed to spot positive or negative sentiments. For our result analysis, we have applied VEDAR sentiment analysis to four economically developed countries where COVID 19 cases were found and also people of those countries are consuming medicines for the treatment of COVID 19.

Here below the graphs measure the polarity with their sentiment score and collect data on the information by using VEDAR sentiment analysis.

Here in Figure 3, the graphs below represent the comparison of the various political heads of different countries with the result and conjointly represent the trend of medicine HCQS use in respect of the treatment of COVID-19. We have concluded that each polarity

analysis has a positivism perspective for the utilization of HCQS for the treatment of COVID-19 which has been taken from the Political Head of different countries' Tweets. We will work adds a lot of work on the opinions/sentiment's alternative optimization techniques.
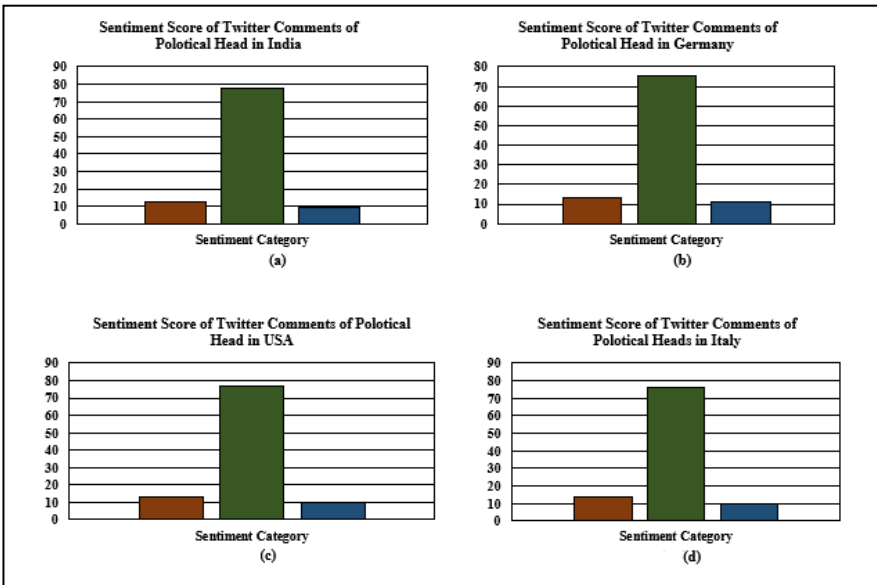

Fig 3: Sentiment analysis result Twitter comments of different countries

7. Conclusion

In our paper on the concurrent result of analyzing totally various twitter comments of various countries' political heads on drug usage trends with reference to COVID-19 of the twitter views of the consultants which have an effect on the traders. The World Health Organization wishes to speculate in the medicine market. This comparative study has proved the results where VADER sentiment analysis ways result in effective performance. The results have shown a robust co-relation between Twitter comments in sentiment polarity. Further, analysis work is applied in sequence to urge a lot of improved leads to this domain. We will jointly investigate the results of the pandemic situation on the results caused by coronavirus infections around the world. We'll analyze many sorts of news with numerical values to alter perceived trends in COVID-

19 scenario. support historical patterns. In our result, we'll contemplate views and opinions denoted by reviewers on twitter.

References

[1] B. J. Jansen, M. Z hang, K. Sobeland & A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American Society for Information Science and Technology,Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2169–2188, 2009.

[2] V. Kharde & P. Sonawane., "Sentiment analysis of twitter data: a survey of techniques," *International Journal of Computer Applications*, ArXiv1601.06971, 2016.

[3] P. Selvaperumal & A. Suruliandi, "A Short Message Classification Algorithm for Tweet Classification," *Int. Conf. Recent Trends Inf. Technol.*, pp. 1–3, 2014.

[4] T. Singh & M. Kumari , "Role of Text Pre-processing in Twitter Sentiment Analysis," *Procedia Comput. Sci.*, vol. 89, pp. 549–554, 2016.

[5] Y. R. Tausczik & J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, 2010.

[6] C. 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)*. 2016.

[7] D. Gurkhe, N. Pal & R. Bhatia., "Effective Sentiment Analysis of Social Media Datasets using Naïve Bayesian Classification," *International Journal of Computer Applications*, 2014.

[8] M. Bouazizi & T. Ohtsuki, "Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer," *IEEE Access*. 6, 64486-64502, 2018.

[9] G. Gautam & D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," *Seventh International Conference on Contemporary Computing (IC3)*, 2014.

[10] A. Amolik, N. Jivane, J. Mahavir Bhandari & M. Venkatesan, "Twitter sentiment analysis of movie reviews using machine learning techniques," *International Journal of Engineering and Technology* 7.6: 1-7, 2016.

[11] S. Mukherjee, A. Malu, A.R. Balamurali, & P. Bhattacharyya, "TwiSent: A Multistage System for Analyzing Sentiment in Twitter,"*Proceedings of the 21st ACM international conference on Information and knowledge management*, 2013.

[12] D. Davidov, O. Tsur & A. Rappoport , "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys,"*Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, August 2010.

[13] M. Neethu & R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," *Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2013.

[14] P. Garg, H. Garg & V. Ranga, "Sentiment Analysis of the Uri Terror Attack Using Twitter" *International Conference on Computing, Communication and Automation (ICCCA2017)*, 2017.

[15] A. Hasan, S. Moin, A. Karim & S. Shamshirb, "Machine Learning Based Sentiment Analysis for Twitter Accounts," *Licensee MDPI*, Basel, Switzerland, 2018.

[16] H. Bhavsar & R. Manglani, "Sentiment Analysis of Twitter Data using Python," *International Research Journal of Engineering and Technology (IRJET)*, 2019.

[17] S. Sirsat, S. Rao & B. Wukkadada, "Sentiment Analysis on Twitter Data for product evaluation," *IOSR Journal of Engineering (IOSRJEN)*, pp. 22-25, 2019.

[18] S. Behdenna, F. Barigou & G. Belalem, "Document Level Sentiment Analysis: A survey," *EAI Endorsed Transactions on Context-aware Systems and Applications*, 2018.

[19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll & M. Stede , "Lexicon-based methods for sentiment analysis," *Comput Linguist J*, pp. 267–307, 2011.

[20] R. M. Tong , "An operational system for detecting and tracking opinions in on-line discussions," *In: Working Notes of the SIGIR Workshop on Operational Text Classification*, pp. 1–6 , 2001.

[21] P. Turney & M. Littman, "Measuring praise and criticism: inference of semantic orientation from association," *ACM Transact Inform Syst J* 21(4):315–346, 2003.

[22] L. Kaur, "Review Paper on Twitter Sentiment Analysis Techniques," *In International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Volume 4, Issue X, October 2016. ISSN: 2321-9653,2016.

[23] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross & K. J. Miller, "Introduction to Word Net: An on-line lexical database," *International journal of lexicography*, 3(4), pp. 235-244, 1990.

[24] S. Mohammad, C. Dunne, B. Dorr, "Generating high-coverage semantic orientation lexicons from overly marked words and a thesaurus," *In: Proceedings of the conference on Empirical Methods in Natural Language Processing*, 2009.

[25] A. Harb, M. Plantie, G. Dray, M. Roche, F. Trousset & F. Poncelet, "Web opinion mining: How to extract opinions from blogs?" *In Proceedings of the 5th International Conference on Soft Computing as Transdisciplinary Science and Technology*, CSTST 08, pp. 211–217, 2008.

[26] P. D. Turney, "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, pp. 417–424, 2002.

[27] G. Wang, & K. Araki, "Modifying SOPMI for Japanese weblog opinion mining by using a balancing factor and detecting neutral expressions," *In Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*; Companion Volume, Short Papers, pp. 189–192, 2007.

[28] D. R. Rice & C. Zorn, "Corpus-based dictionaries for sentiment analysis of specialized vocabularies," *Proceedings of NDATAD*, pp. 98-115, 2013.

[29] K. Y. Su, T. H. Chiang & J. S. Chang, "An Overview of Corpus-Based Statistics Oriented (CBSO) Techniques for Natural Language Processing," *Computational Linguistics and Chinese Language Processing* 1(1), pp. 101–157, 1996.

[30] K. Y. Su & T. H. Chiang, "Some Key Issues in Designing MT Systems. *Machine Translation*," Vol. 5, No. 4, pp. 265-300, 1990.

[31] K. Y. Su & T. H. Chiang, "Why Corpus-Based Statistics Oriented Machine Translation," *Proceedings of 4th Int. Conf. on Theoretical and Methodological Issue in Machine Translation*, pp. 249-262, Montreal, Canada, June 25-27, 1992.

[32] K. Y. Su, J. S. Chang & Y. L. Una Hsu, "A Corpus-based Two-Way Design for Parameterized MT Systems: Rationale, Architecture and Training Issues," *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, TMI-95, vol. 2, pp. 334-353, 1995.

[33] H. E. Moss, R. K. Ostrin, L. K. Tyler & W. D. Marslen, "Accessing Different Types of Lexical Se-mantic Information: Evidence from Priming," *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21:863-883(1995).

[34] Natural Language Toolkit http://www.nltk.org/ (Date Last Accessed, November 20, 2018.

[35] S. Bird, E. Loper & E. Klein, "Natural Language Processing with Python," *O'Reilly Media Inc*, 2009.

[36] C. 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)*, 2014.

[37] S. Elbagir & J. Yang, "Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment," *In Lecture Notes in Engineering and Computer Science: Proceedings of The International Multiconference of Engineers and Computer Scientists*, 2019.

[38] Natural Language Processing with neural networks http://nilc.icmc.usp.br/nlpnet/.

[39] E. R. Fonseca & J. L. G. Rosa, "A Two-Step Convolutional Neural Network Approach for Semantic Role Labeling," *Proceedings of the 2013 International Joint Conference on Neural Networks*, 2013. pp. 2955-2961, 2013.

[40] E. R. Fonseca & J. L. G. Rosa, "Mac-Morpho Revisited: Towards Robust Part-of-Speech Tagging," *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*. pp. 98-107, 2013.

[41] S. A. El Rahman, F. A. Al Otaibi & W. A. Al Shehri, "Sentiment Analysis of Twitter Data," *2019 International Conference on Computer and Information Sciences (ICCIS),* 2019, pp. 1-4, doi: 10.1109/ICCISci.2019.8716464.