ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse56932020.pdf
https://doi.org/10.30534/ijatcse/2020/56932020

Analysing COVID-19 News Impact on Social Media Aggregation

Dheeraj.K
Department of Computer Science and Engineering, National Institute of Technology, Warangal, India
dheerajparashara@gmail.com

ABSTRACT

The pandemic corona virus so called COVID-19 outbreak, which was recognized in late 2019, requires special consciousness because of its future epidemics and possible worldwide menace. Besides its clinical approaches and treatments. As Artificial Intelligence (AI) assures a new prototype for healthcare. There are various AI tools that are built upon Machine Learning (ML) algorithms are included to examine the data and decision-making processes. People spend many hours every-day on social media websites to share their views, ideas, opinions, and expressions with others, so in this paper, I have analysed the sentiments on reddit news regarding coronavirus disease(COVID-19) jeopardy, because most of the peoples from various countries are affected by coronavirus that is very condemnatory issue at present days, so analyze the sentiments of various people's opinion for this disease, we are fetching the reddit streaming articles connected to coronavirus using reddit API and scrutinized these articles using machine learning methods and tools as positive, negative and neutral. In this paper, I have run experiments through Python programming on various articles related to pandemic corona virus using reddit API and NLTK library is used for pre-processing of articles and then scrutinizing the articles dataset by using Textblob and it pointers to stimulating outcomes as positive, negative, neutral sentiments over different conceptions. The outcomes show that on reddit related to COVID-19 articles about 50% percentage articles are neutral, 22% articles are positive and about 28 % was negative.

Key words : COVID-19 Articles, Machine Learning, Reddit News, Social Media, Sentimental Analysis**.**

1. INTRODUCTION

The present-day menace to global health is the extant outbreak of the respiratory disease that was recently discovered and given the name Coronavirus Disease 2019 (Covid-19). As it was recognized in December 2019 [1]. The upshot was abruptly shown to worldwide and be triggered by a COVID-19 that is structurally included to the virus that results capricious parlous affecting organs of respiration like SARS. Corona virus is related to family of both MERS and SARS virus these two preceding illustrations of emergence of coronavirus disease. In addition to this, SARS virus i.e., dire perilous relating to affecting the respiration between the year 2002 and 2003 and MERS virus related to affecting the respiratory organs in Middle East , still active virus from the year 2012 to the present .The pervasive Covid-19 spate has posed dire challenges for the public health, medical and research communities [49].The paper illustrates, Section 1 shows the introduction related to coronavirus contagion and works correlated to sentimental analysis and section 2 shows the social structure study for sentimental mining, section 3 shows the methods included for COVID 19 social media aggregation sentimental analysis, section 4 shows the Data classification for sentimental analysis, in what way the data from the articles are further classified, section 5 about experimental results, Section 6 Conclusion and future related works and final section 7 about the references.

COVID-19 virus was extremely sundry, found in assorted animal species, and are substantially confront in medical enactment during the cold and flu season, yet many paramount care medical practitioners are not commodious with these respiratory pathogens. These are sporadically tested, since medical practitioners are also newfangled for corona like new virus. Moreover, there has been no particular ministration for these kind of pervasive viruses [50]. The Sentimental mining and Opinion mining are pivotal research areas because due to the massive number of daily posts on social media networks, extracting people's opinions or behaviors is a challenging task. My paper analyses the sentiments regarding Coronaviruses and this virus was related to family of viruses. As we have seen there are many viruses including the new virus(COVID-19) effected to humans especially, as these viruses evolved from animals [24], but most just cause cold-like symptoms, Around 200 countries in the world was effected with coronavirus so in this paper I have mentioned the results(in section 5) about sentiments of peoples for coronavirus analysed [23].

About the 90 % of present data has been included during the past two years and retrieving intuition into this large scale data is not trifling [16, 17]. In addition to, sentimental analysis has various applications for various domains for epitome, in businesses to get customer feedbacks for their products by which companies can learn customer's feedback or reviews from social media in-order to ameliorate their prowess of the product. Sentimental or Opinion mining have studied on various research fields has discussed [9]. There are unalike exertion that have been done on sentimental perusal on Facebook [18-22] however in this paper my focus is on the sentimental probe using social media aggregation especially on COVID-19. For sentiment analysis first of all understand how text was classified, and summarized as it gives weight to every statement whether classified as positive

and negative or neutral. For this we have two fundamental approaches to disentangle text categorization 1) Extraction approach 2) Abstraction approach.

In the extraction approach, words and word expressions are extracted from the original text to generate an outline. Whereas abstraction approach tries to learn an intramural language representation and then generates an outline that is more similar to the outline done by the people. Understanding the text is a predominant task to solve. Moreover, some machine learning methods, including incompatible supervised and unsupervised algorithms, are being exploited. There are divers approaches to generate a contour. One of the approach could be to cavalcade the significance of sentences within the text and then commence an outline for the text based on the crucial numbers. Another approach so called end-to-end reproduction approach. In some domains likewise language conveyance, speech recognition and image recognition methods performs better [2]. Wilson, et.al [10] considered the phrase level in sentimental analysis system should be able to known the polarity of the phrase and facet -based representations and tree composition have been included for sentimental analysis in twitter [12]. SemEval-2017 [13] also shows the sentimental analysis in twitter tasks and similarly consider for reddit news articles. Since articles in reddit is a particular text not like a usual text there are some works similar to reddit, that address this problem like the work for short informal texts [14]. The Sentimental analysis has many applications in news [15]. In this paper, the social network analysis and the importance of it are discussed in the following section, then reddit is discussed as it is a rich resource for sentimental analysis. In the following sections, the model and its implementation was discussed. The sentiments are shown through positive, negative and neutral on coronavirus using various visualization methods disused in following section to better understand.

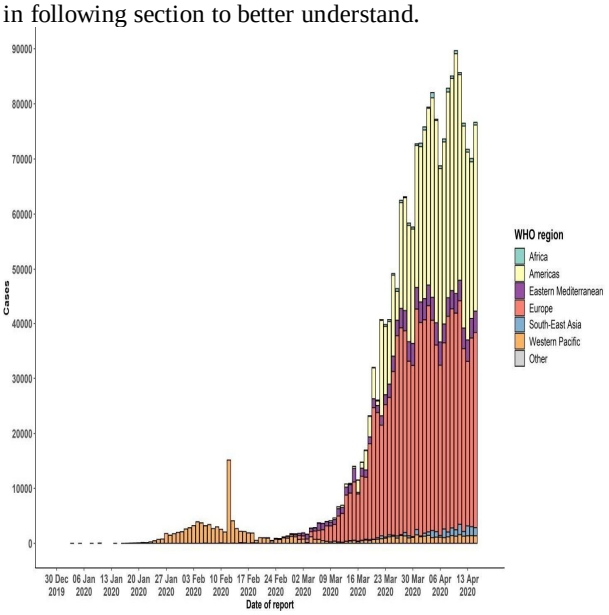

Figure 1: Epidemic curve of confirmed cases of COVID-19(source:https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200416-sitrep-87-covid19.pdf?sfvrsn=9523115a_2, April 16, 2020).

2. SOCIAL MEDIA STRUCTURE MEASURES

The Social media structure measures are included below based upon the feature selection, augury, machine learning, and sentiment analysis. In the following subsections contains related works to these methods.

2.1 Social Network Probe

Social network probe is the study of individual interactions and communications on various topics and yet-present-days it has received more attention. The millions of people around world, ceaselessly sharing their views on various topics on a routine basis on social network likewise Twitter, Facebook, Instagram, Reddit etc. It has various applications in various areas of research from social science to business [3]. Present-days, Reddit is one of the voguish social news aggregation which according to the [25]. At present it has over 330 million users also known as “redditors” as of 2018, it has accommodates more than 138,000 nimble communities [26]. It is a rich source to study about people’s opinions and sentimental analysis likewise twitter [4]. In this, each article was prominent to illustrates the sentimental perusal of the text whether is it positive, negative, or neutral [42].

2.2 Sentimental Probe

The sentimental probe can be included as the automatic uproot of quantitative perspectives from subjective text. Since many years, whacking methods indication, percipience and applications into sentiment or opinion analysis was published [32]. Since most the work has been done on precise domains, likewise colossal texts, movie reviews, [33], in some epitome indicating verisimilitude of sentimental classification. Go et al[32] introduced an interesting method towards the classification of sentiment in Twitter messages [7] and this I have taken an account for reddit COVID news. First, to perform sentiment analysis on Twitter data. As they propose the method called "distant supervision", and it was defined as training with noisy data. Since every Twitter message which contains a smiley character is considered a signal denoting the sentiment of that entire tweet on twitter. Accordingly, they automatically collect information from Twitter by filtering the common Twitter feed for smiley characters. Indistinguishable way, they have created an annotated training corpus, in they train machine learning approaches and create a sentiment polarity classifier [33].

2.3 Sentimental Prophecy

The sentimental prophecy can be indicated as the automatic prophecy of what the quantitative opinions of some viewers will be to some message, based on the contents of that message and the earlier inspection of a similar viewer’s response to similar messages. The Sentimental prophecy was most difficult task than sentiment Probe, due to the lack of emphatic opinion in source data that is classified and the dependency on viewer’s similitude for accurate prophecy [8]. Most of the research into sentimental prophecy for news has indicated on predicting the manoeuvre of the stock market [34] [35] [39]. In some recent work [38] [36] [37] authors are emergence to attempt to prophecy the sentiment polarity of

responses to news, that I have mentioned related results in section 5. In this section, some of the relevant works of sentimental prophecy was mentioned. Cheong Fung et al [34] present a system for stock prophecy based on the particulars contained within the news articles. They probe the effect of news articles on time series flow of stock signals, based on the coherent markets belief. As the maneuver of the stock price indicates a signal about the allures of a stock, which encodes the sentiment reaction of the investors. They use text mining and data mining techniques in a novel way [5]. Their system is circs -driven, it means that it can generate prophecy in real-time. This encompasses a novelty associated to former work on standard prophecy using news. Some researchers exploited perusal related on linear regression and huddling of stimulating methods. Related upon produced and apportioned tendency pointer and listing fascinating bulletin related articles comparative to the penchant labels by machine learning procedures. This includes that only the news related tendency was linked. Syndicating the each record was indicated as a standardized text with independent words as the structures of a feature records. Some researchers used the Tf-Idf procedure for feature assortment. By using the certain options, researchers considered the dividing their actual collections, since all are using related machine learning approaches. Including the many social media likeness approaches to contemplate whether any words related to its articles. To reach this approach, the grouping procedure primarily computes a centroid for every gathering. Ensuing, the similarity measure is pragmatic to each record in a group associating it to that cluster. Both records and groups in this circumstance are indicated as words achieved by Tf-Idf [42]. Since recently formed groups are associated using social media measures [43], and relinquish, for both upsurge and globule signals, the one group which is further identical to the two collections of the differing signal. It illustrates, upsurge clusters most indistinguishable to both globule clusters was severed. Therefore they achieve to sort the news articles which are less pertinent to their experiment. Moreover, representing the complex data to features included in the articles and which supports merely single tendency form. [42] They identified and measured many resemblance constants in social media. They amalgamated many similarity measure together in social media for each word in every record, in order to determine the value of that word for that article. The clairvoyance becomes, that the fewer a word in group of an article, the more predominant that it has an illustration feature for that article and groups. Throughout this inspection, they signified that this outline progresses the recall of their system.

2.4 Reddit Sentimental Probe

Social media or Social network analysis has a rich podium to learn about public opinions [11] and sentiment related to different topics, they can communicate and share their opinion or views on daily basis on social media like Reddit, Facebook ,Twitter[6], and Instagram, etc. The sentiment cognizant systems at present days have sundry applications from business to medical areas [43]. Meanwhile social networks, especially Reddit, comprehends texts and people may use different words and abbreviations which are strenuous to excerpt their sentiment by current Natural Language processing schemes easily. Also, some researchers have used deep learning and machine learning techniques to extract and excavation the polarity of the text [15]. Some are worked on sentiment extraction on facebook use text abbreviations [27], Therefore sentimental analysis on Reddit was most challenging task [28].

3. METHODOLOGY

This procedural probe reports the implementation of the reddit COVID-19 sentiment analysis, by using the Reddit API. PRAW is a wrapper of reddit API for fetching the articles straight from reddit that are posted by different people. The streaming articles are fetched and then saved into CSV files to sentiment analysis. Also, there are significant works and tools dealing with text-mining on social networks. In this research, the wealth of available libraries has been used. The approach to extract sentiment from articles is as follows: First, import PRAW for creating the connection with Reddit API Secondly, Fetch articles as dataset and then save into CSV file. Next, the Pre-processing of articles by removing the stop words, punctuations, #tags, etc. After that, tokenize each word in the dataset and save it into the dataset.

Moreover, comparing with positive, neutral and negative sentiments for each word in the dictionary. Then augment the sentimental analysis count. Eventually, based on this count, which results the percentage about sentiment to adopt the polarity. Most of the investigators have done various sentimental studies on reddit and also twitter for different tenacities for instance the work designed by [7], but in this I have used the real-time reddit sentimental analysis of pandemic coronavirus. In addition to, experimental results I have mentioned in section 5.

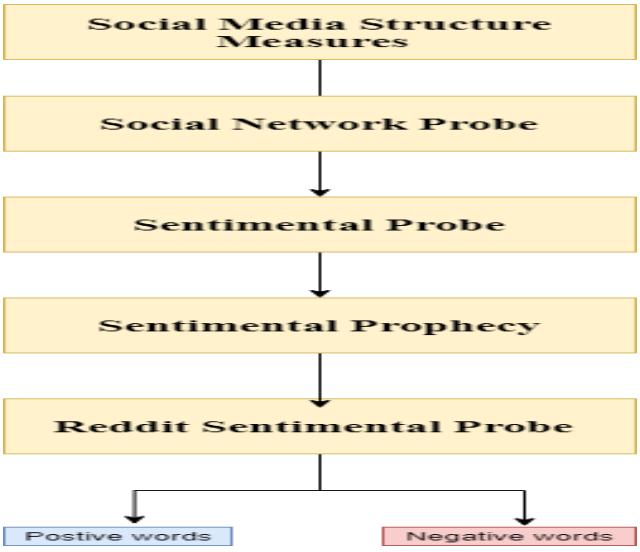

Figure 2: Shows the classification of Social Media structure, which results the percentage of positive and percentage of negative words.

4. DATA CLASSIFICATION FOR SENTIMENTAL ANALYSIS

In this Paper focuses on predicting the general sentiment polarity of the reactions to COVID-19 news on Reddit before a news article is published. The research question regarding the influence of COVID-19 news and user group similarity on the accuracy of sentiment prediction, we require a news source that serves COVID-19 related news articles and has commentators to news who are divisible into various user groups. It is exigent that reactions of user group at a time are about the same COVID-19 related news article, to prevent source bigotry. For this cause, we need to acquire news and reactions for various groups from the same platform. Moreover, the volume of various topics recurring in news daily entail our training set to be sizable, as to collect sufficient training data on as various topics as possible. Probing the data needs to be acquire from a source that shows a high number and diversity of news topics, in multiple categorization, which has an active and somewhat assorted user community. For these reasons, we collect data from Reddit. The Reddit [40] is a social news website, where the users known as "Redditors", post links to online content or make text postings themselves. This content frequently consists of news, but also contains the links to images, videos, blog posts and other material [44]. For each submitted link, Reddit users can vote on how essential or relevant they find the associated content. They can give positive and negative votes. Allied on the votes and using a metric [45], the postings are ranked. This ranking is used to resolve the exhibit evince of the postings on Reddit front pages in the default setting. Moreover, articles can also be manually ranked by the number of responses they have collected, or chronologically. Where each user can react to every posting, or to the reactions of other users. The users can also vote on the standard of the responses. Same voting and ranking principles and options which are attainable for postings are used for the reactions as well. All postings on Reddit are determined by the poster into whatever classification they feel the material belongs so called as "subreddit". The independent users can subscribe to subreddits to customize what they see on Reddit homepage when they visit. “The year 2020 biggest subreddits are Announcements, Funny, Askreddit, Gaming, aww, Pics, Science, World news, Music Movies” [29].

Moreover, all subreddits are open to all users to post and comment. The Reddit's open culture [20], simple premise and general user interface entice a large audience. The site provides a very wide demographic, political and social range of users, with most of its users imminent from the United States [46]. As of April 2020 there are around 2,049,838 subreddits [30]. Also, the factual amount of subreddit users, plausible to be higher [41] [46] “and assessments includes that position of Reddit in the year 2020 as 6th most visited site in the United States” [46].

4.1 Data Collection

While the broader Reddit community and social mechanisms are an engrossing subject in itself, in this paper I focus on the subreddits which contain the comparatively largest amount of postings linking to entrenched online news sources. Manual survey of the content in most popular subreddits capitulated the following shortlist of news-content rich subreddits, which includes the following topics [31]: Roaring2020s, MisreadSprites, PetTheDamnKitty, Wholesomereddit, TheRedditSymphony.

Reddit runs a simple API to enable data assortment from Reddit, I have used Reddit API, scraper application in Python. The scraper polls the chronologically adjured posting listings for the shortlisted topics. New postings related to COVID-19 were added to an internal watch list. Every posting present in the watch list is gathered from Reddit in full exactly 24 hours later, to verify that the comments for every posting are always gathered over the same time period. We disregard the fact that the odds of a posting attracting comments could be determined by the posting time. We presume that, since redditors from similar geographical areas both posting the articles and commenting to them, the posting and commenting articles are correlated. For every posting, we gather both the posting link and all obtainable comments. In order to gather the full text of the linked news articles, we conceive and implement a heuristic system in Python 2.7. The system gathers the web-page from the link in the posting areas, and extracts candidates for article text from the webpage. For every candidate in the webpage, we compute the respective weight of text versus HTML tags. Beside this, we adjust the scores of the candidate segments with their Okapi BM25 score [47], using the webpage title as the search question. At last, we opt the highest scoring article text candidate and remove all HTML tags from it. Next, we add the extracted article text to the Reddit posting and preserve the posting to disk [48].

4.2 Data Contour

I have collected COVID-19 related postings from the shortlisted topics (subreddits) between the months of January to April 2020 which conform to the adherent rules. The posted link does not indicate back to Reddit like, not a self-post. The Posted link refers to text content, be made up of at least 50 words in the article text for instance, the content was not to a picture, video or other kind of post. The result shows that there are 868 related COVID-19 postings.

5. RESULTS AND DISCUSSION

Using the Reddit API we can get thousands of headlines related to different news subreddits and. I have used NLTK's vader analyzer, which computationally identifies and classifies text into three sentiments i.e., positive, negative, or neutral. Considering the "new" posts related to COVID-19, and I have received 868 headlines, and using PRAW wrapper, it use a really simple interface while it handles a lot of errand in the background, like rate limiting and classifying the JSON responses. This analyzer results lexicon of positive, neutral and negative words .We can use this tool by first implementing a Sentiment Intensity Analyzer to classify our corona virus related headlines, then we'll use the polarity scores method to get the sentiment. I have used the Dataframe encloses four sections from the sentiment scoring i.e., Pos, Neu, Neg then compound. The first three illustrate the sentiment score percentage of COVID-19 related content in our headline, also the compound single number that scores

the sentiment. The compound confines from -1 as immensely negative to 1 as immensely positive. I have considered the posts with a compound value significant than 0.2 as positive and less significant than -0.2 as negative. Everything else will be 0. In addition to, some testing and experimentation that goes with opting these ranges, and there is a trade-off to be made here. If I have chosen a higher value, I might get more dense results likewise less false positives and false negatives. Apparently, the size of the outcomes will reduce glaringly.

Table 1: Shows how the COVID-19 virus related text are classified and labelled.

SNo	Compound	News articles related to COVID-19
0	0.6260	I couldn't find no masks where I live, so I se...
1	0.4939	How do I deal with my friend not doing any cho...
2	0.0000	Image tagged in bugs bunny wide awake
3	-0.8176	Type one diabetic- covid is scaring the hell o...
4	-0.1779	CA doing well - plot of deaths/day normalized ...

We can now keep conjoin the csv, but without considering without getting the duplicates. For this, I have added a more advanced saving function that reads and eliminate duplicates before saving. In reddit dataset, I have categorized COVID-19 related articles as Neutral means 0 and positive as 1 and negative as -1. The following result shows the top 20 words distributed related to COVID-19.

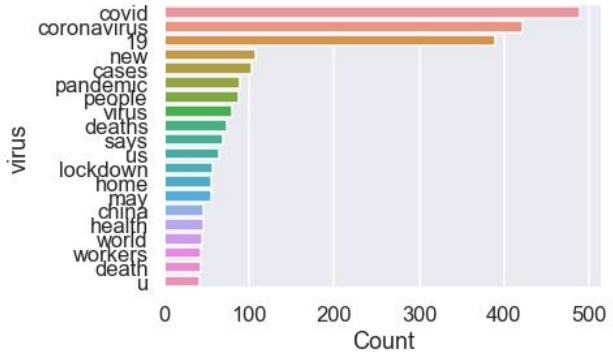

Figure 3: Shows the 10 most Corona virus related frequently used words for plotting the bar plot.

The words cloud below figure 4 shows the most of the words related to Pandemic virus: Coronavirus, COVID, Symptom etc. Some words are more related to the present situation facing thought world was Quarantine, Situation, Home, Patients, etc.

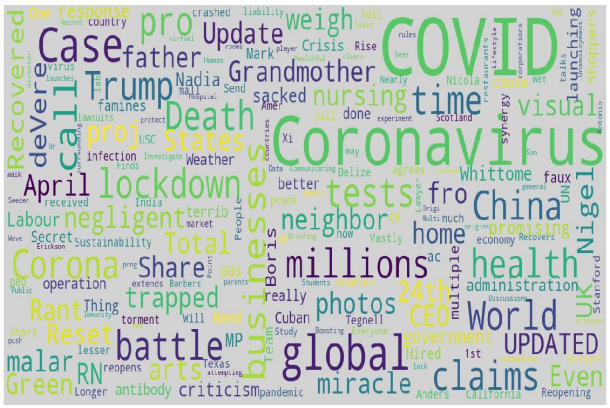

Figure 4: Word cloud related to COVID-19

In the subsequent figure, represents the percentage of words distributed as three categories, in which most of the articles I have observed in reddit related to COVID-19 articles as neutral.

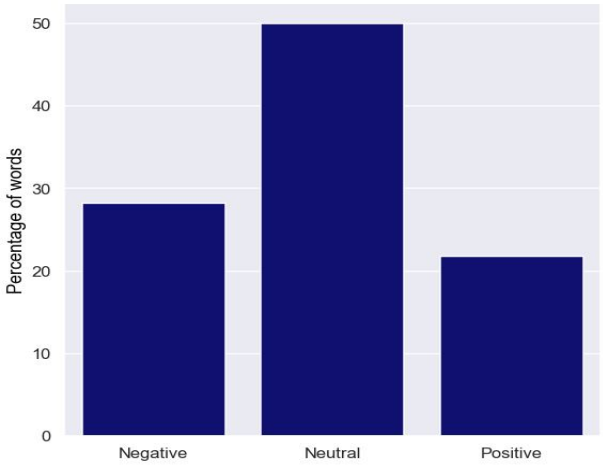

Figure 5: Chart showing the COVID articles occurrence outlines

The below figure displays the positive sentimental analysis for the word occurrence and also displays words ordered by their frequency.

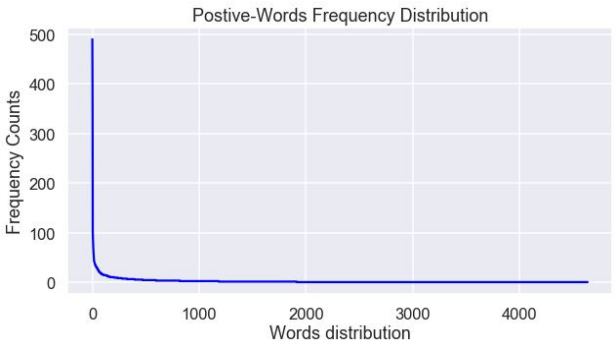

Figure 6: Shows positive word occurrence distribution

The above figures displays the results about nearly straight line with a ponderous edge means like deafening edge. Moreover, the above figure illustrates that, word distribution

a huge marginal of the words emerge at most, whereas the majority of words appear less.

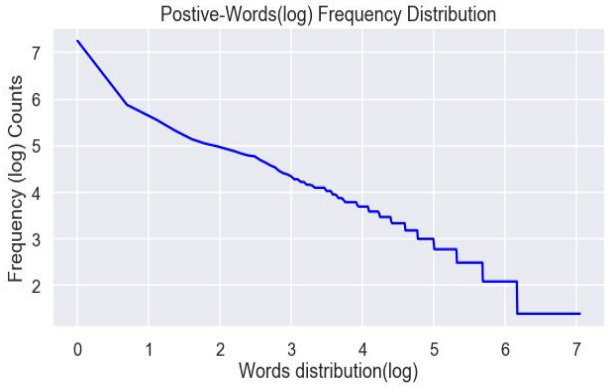

Figure 7: Shows the positive words (log) frequency distribution

Above I have examined the positive related words, apparently towards the negative ones. The below figure 8 shows how word frequency distributed and figure 9 shows the frequency word(log) for the negative text data.

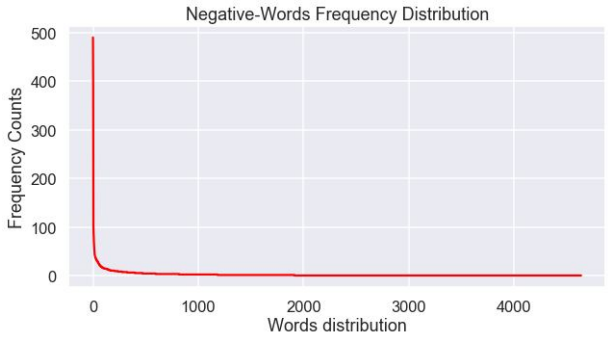

Figure 8: Shows the Negative words occurrence distribution

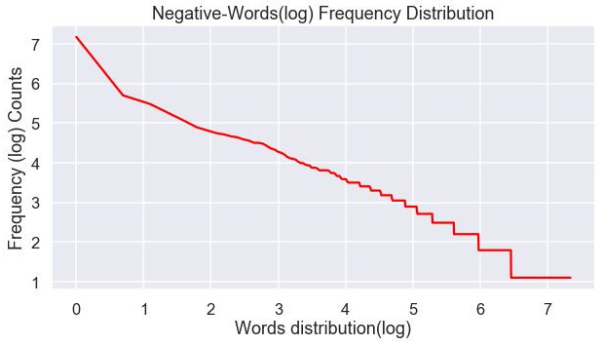

Figure 9: Shows the negative words (log) frequency distribution

I have observed that negative distribution fits under a bit of more satiny slope, but the heavy edge is certainly there. The peroration here, imposes the same as the previous one shown in positive distribution.

6. CONCLUSION AND FUTURE SCOPE

In this research paper, the different people opinion on effect of the pandemic COVID-19 was discussed. The main focus of this paper was to extract the sentiment and opinion mining about COVID-19 related articles on Reddit, Using the various machine learning techniques on Jupternotebook to implement the sentimental analysis as positive, negative and neutral. The outcomes are exposed by using Matplotlib library. The Reddit API sorts it extremely tranquil to accumulate a lot of news related data to COVID-19 impartially quickly. Certainly, by enhancing the data assortment process to get thousands of rows related to COVID-19 headlines, we can use for further analysis and prediction. In this paper, to accomplish the prophecy of broad sentiment polarity in the reactions to COVID-19 news on social media aggregation. Precisely, for sentiment prediction I have used a dataset of COVID -19 news articles and comments collected from Reddit. To train our sentiment prediction systems, I have used the size of the Reddit corpus that requires automatic annotation of the sentiment polarity of Reddit comments. Moreover, I have used a bootstrapping approach, by domain-knowledge comportment procedures using a Twitter corpus and to amend the sentiment polarity knowledge from tweets to Reddit observations. After computing several approaches of domain-knowledge conduct, I have concluded that our instinct that the groups of Reddit users more homogenous in their virus related topics have more predictable sentimental responses to COVID-19. The future work I will continue my perusal to construct and train an emotional classifier. It has been perceived that the neutral sentiments are remarkably high which results there is a need to better Reddit sentiment analysis. Imminent system must be able to ascertain clusters of which users reactions have the identical comment polarities towards the COVID-19 related topics, the topics themselves being related. Still a lot that could be engineered with regarding to data mining, and exploration to be done with the data retrieved. This system must be based on modelling the stability of user emotions related to COVID-19 topics from the news articles.

REFERENCES

1. ttps://www.who.int/csr/don/05-january-2020-**pneumoni-of-unkown-cause-china**/en/
2. https://www.elsevier.com/connect/**coronavirus-information**-center
3. Boguslavsky, I. **Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies.** Papers from the Annual International Conference "Dialogue"(2017) (pp. 14-28) "(2017).
4. Pak, A., & Paroubek, P. **Twitter as a corpus for sentiment analysis and opinion mining**. In LREC (Vol. 10) (2010, May).
5. Scott, J. **Social network analysis: developments, advances, and prospects.** Social network analysis and mining, 1(1), 21-26(2011). https://doi.org/10.1007/s13278-010-0012-6

6. https://www.statista.com/statistics/282087/**number-of-m onthly-active-twitter-users/**
7. Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S.. **A system for real time twitter sentiment analysis of 2012 us presidential election cycle.** In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics(2012, July).
8. Pang, B., & Lee, L. O**pinion mining and sentiment analysis.** Foundations and Trends® in Information Retrieval, 2(1–2), 1- 135(2008). https://doi.org/10.1561/1500000011
9. Dos Santos, C. N., & Gatti, M. **Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts** (2014, August).
10. Wilson, T., Wiebe, J., & Hoffmann, P. **Recognizing contextual polarity in phrase level sentiment analysis**. In Proceedings of the conference on human language technology and empirical methods in natural language processing(pp. 347-354). Association for Computational Linguistics(2005, October). https://doi.org/10.3115/1220575.1220619
11. Liu, B. **Sentiment analysis and opinion mining.** Synthesis lectures on human language technologies, 5(1), 1-167(2012).
12. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. **Sentiment analysis of twitter data.** In Proceedings of the workshop on languages in social media (pp. 30-38). Association for Computational Linguistics(2011, June).
13. Rosenthal, S., Farra, N., & Nakov, P. SemEval-2017 task 4: **Sentiment analysis on Twitter**. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)(pp. 502-518) (2017). https://doi.org/10.18653/v1/S17-2088
14. Kiritchenko, S., Zhu, X., & Mohammad, S. M. **Sentiment analysis of short informal texts**. Journal of Artificial Intelligence Research, 50, 723-762(2014).
15. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen ,B., & Belyaeva, J. **Sentiment analysis in the news.** arXiv preprint arXiv:1309.6202(2013).
16. Ortigosa, A., Martín, J. M., & Carro, R. M. **Sentiment analysis in Facebook and its application to e-learning.** Computers in Human Behavior, 31, 527-541(2014). https://doi.org/10.1016/j.chb.2013.05.024
17. Bagheri, Hamid, and Abdusalam Abdullah Shaltooki. **"Big Data: challenges, opportunities, and Cloud-based solutions."** International Journal of Electrical and Computer Engineering 5.2: 340(2015).
18. Bagheri, Hamid, Mohammad Ali Torkamani, and Zhaleh Ghaffari. **"Multi-Agent Approach for facing challenges in Ultra-Large Scale systems."** International Journal of Electrical and Computer Engineering 4.2: 151(2014). https://doi.org/10.11591/ijece.v4i2.4102
19. Sushil Kumar Trisal and Ajay Kaul **, "Dynamic Behavior Extraction from Social Interactions Using Machine Learning and Study of Over Fitting Problem"** , International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.5, September - October 2019. https://doi.org/10.30534/ijatcse/2019/54852019
20. Feldman, R. **Techniques and applications for sentiment analysis.** Communications of the ACM, 56(4), 82-89(2013).
21. Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacherjee, S. K., & Viswanathan, **A.Sentiment analysis of Facebook data using Hadoop based open source technologies.** In Data Science and Advanced Analytics (DSAA), 2015. 36678 . IEEE International Conference on (pp. 1-3) IEEE(2015). https://doi.org/10.1109/DSAA.2015.7344883
22. Trinh, S., Nguyen, L., Vo, M., & Do, P. Lexicon-based **sentiment analysis of Facebook comments in the Vietnamese language.** Recent developments in intelligent information and database systems (pp. 263 276). Springer International Publishing(2016).
23. Haddi, E., Liu, X., & Shi, Y.The **role of text pre-processing in sentiment analysis.** Procedia Computer Science, 17, 26-32(2013). https://doi.org/10.1016/j.procs.2013.05.005
24. Wolfe ND, Dunavan CP, Diamond J. **ORIGINS OF MAJOR HUMAN INFECTIOUS DISEASES.** In: Institute of Medicine (US). Improving Food Safety Through a One Health Approach: Workshop Summary. Washington (DC): National Academies Press (US); 2012. A16.
25. "Singer, Philipp and Flöck, Fabian and Meinhart, Clemens and Zeitfogel, Elias and Strohmaier, Markus on **Evolution of reddit: from the front page of the internet to a self-referential community?"**
26. https://en.wikipedia.org/wiki/Reddit
27. **Sentiment analysis in Facebook and its application to e-learning** by Alvaro Ortigosa, José M. Martín, Rosa M. Carro.
28. **The Impact of Crowds on News Engagement: A Reddit Case Study** by Benjamin D. Horne, Sibel Adalı.
29. https://**redditmetrics**.com/top
30. https://**redditmetrics**.com/history
31. https://www.**reddit**.com
32. A. Go, L. Huang and R. Bhayani, **"Twitter Sentiment Classification using Distant Supervision,"** The Stanford Natural Language Processing Group, 2008/2009.
33. A. Montoyo, P. Martínez-Barco and A. Balahur, **"Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments,"** Elsevier, 2012. https://doi.org/10.1016/j.dss.2012.05.022
34. G. P. C. Fung, J. X. Yu and W. Lam**, " News Sensitive Stock Trend prediction," in Advances in Knowledge Discovery and Data Mining** : 6th Pacific-Asia Conference, Taipel, Taiwan, 2002.
35. G. P. C. Fung, J. X. Yu and W. Lam, **"Stock prediction: Integrating text mining approach using**

real-time news," in Computational Intelligence for Financial Engineering, Hong Kong, 2003.

36. R. Balasubramanyan, W. W. Cohen, D. Pierce and D. P. Redlawsk, **"What pushes their buttons? Predicting comment polarity from the content of political blog posts,"** in Workshop on Language in Social Media, Portland, Oregon, USA, 2011.
37. R. Balasubramanyan, W. Cohen, D. Pierce and D. Redlawsk, "Modeling Polarizing Topics: **When Do Different Political Communities Respond Differently to the Same News?**," in 6th International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 2012.
38. K. Lerman, A. Gilder, M. Dredze and F. Pereira, "Reading the Markets: **Forecasting Public Opinion of Political Candidates by News Analysis**," in 22nd International Conference on Computational Linguistics, Manchester, UK, 2008.
39. V. Seghal and C. Song, **"SOPS: Stock Prediction using Web Sentiment,"** in Seventh IEEE 77 International Conference on Data Mining Workshops, Washington, DC, USA, 2007. https://doi.org/10.1109/ICDMW.2007.100
40. **"Reddit," [Online].** Available: http://www.reddit.com. [Accessed 22 6 2012].
41. **"Could Reddit be the world's most influential website?,"**[Online].Available:http://www.blueglass.com/blog/could-reddit-be-the-worlds-most-influential-website/. [Accessed 26 4 2012].
42. W. Dai, G. Xue, Q. Yang and Y. Yu, "**Transferring Naive Bayes Classifiers for Text Classification,"** in 22nd national conference on Artificial intelligence - Volume 1, Vancouver, British Columbia, Canada, 2007.
43. M. Blei, A. Y. Ng and M. I. Jordan, **"Latent dirichlet allocation,"** The Journal of Machine Learning Research 3, p. 993–1022, 2003 .
44. K. K. Bun and M. Ishizuka, **"Topic Extraction from News Archive Using TF*PDF Algorithm,"** in Proceedings of the 3rd International Conference on Web Information Systems Engineering, Washington, DC, USA, 2002.
45. **"Reddit vote weighting metric,"** [Online]. Available: http://www.seomoz.org/blog/reddit-stumbleupon-delicious-and-hacker-news-algorithms-exposed. [Accessed 28 6 2012].
46. **"Reddit visit statistics,"** [Online]. Available: http://www.alexa.com/siteinfo/reddit.com. [Accessed 28 6 2012].
47. "Okapi BM25," in **An Introduction to Information Retrieval**, Cambridge University Press, 2009, p. 233.
48. N. Cristianini and J. Shawe-Taylor, **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods**, Cambridge University Press, 2000. https://doi.org/10.1017/CBO9780511801389
49. https://en.wikipedia.org/wiki/Coronavirs
50. S. Ravi Kumar, Dr. M. Vamsi Krishna , Dr. Anurag **"A survey on prediction approaches for epidemic disease outbreaks based on social media data"** International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.3, May - June 2019. https://doi.org/10.30534/ijatcse/2019/86832019