Tech Science Press

Sentiment Analytics: Extraction of Challenging Influencing Factors from COVID-19 Pandemics

Mahmoud Oglah Al Hasan Baniata* and Sohail Asghar

Department of Computer Science, COMSAT University, Islamabad, Pakistan
*Corresponding Author: Mahmoud Oglah Al Hasan Baniata, Email: mbaniata2211@gmail.com
Received: 14 March 2021; Accepted: 04 May 2021

Abstract: The advancement in electronic devices and communication technologies in social media have introduced major changes in today's communication and people have accepted such communicational habits at a rapid pace. The changes involve the way people started interacting with each other, and modern mean of discovering new groups of people, and individuals with similar mindsets, mutual interests, and ideas to share with. As far as the communities are concerned, there are so many social drives (such as "Say No to Plastic") that need to be discussed on a certain platform for their promotion. Although, it's quit is challenging, but with the advancement of communication technologies propagations and the increasing engagement of the youth on social media towards social drives made this achievable. There exists a need to examine the influencing factors that peoples express when such drives are promoted on any social media forum in order to evaluate their sentiments. The objective of this research is to explore the influencing factors of pandemic Corona Virus Disease 2019 (COVID-19) when it has been promoted as a social drive on social media forums. In this research effort, sentiment analysis along with novel similarity measures such as THINKMAP, VADER++ and lexicons have been utilized to analyze the influencing factors of those peoples who are affected by COVID-19 when social media forums promote it as a common social drive. The significance of this research is that it would help to demonstrate that participation of youth of different countries could boost up the promotion of social drives and fasten the process which leads towards the positive awareness about any common social drive. Moreover, it would cover the gap of common issues on social media that are not in practice in past studies. It also strengthens the awareness phenomena in youth beyond boundaries. From the experimental results and discussion, it has been observed that the proposed research performed well in term of accuracy.

Keywords: Social media; social drives; covid-19; sentiment analysis; similarity measures

1 Introduction

As far as social networking is concerned, it is all about social actors and their settings like individuals, communities, and organizations comprised of social actors and the social interactions plus communication. In the inspection of culture and society, technological innovation has led in both positive and negative changes. Therefore, social media and networking sites provide both damaging and productive elements to the community and the effect of social media and social networking sites had a major influence on youth during their transition to adulthood [1] alike to any powerful platform or improvement in innovation.

As a way of promoting any social drive, searching for entertainment, distributing information and communication [2], social media is widely used by people particularly youth including graduate and undergraduate students. Fig. 1 depicts the age wise usage of social media forums. Their purpose of joining the social media is for "comments" and "likes" and their views that are being "shared" [3]. Social networks and micro-blogging sites like twitter, Facebook, Google views and flicker are more famous in public as compared to other media because of their extensive usage and simple processing [4]. Social media and social networks have dramatic influence on the community. With the passage of time different social media forums, continuously passing through the different social drive and allow people to write their views on it and promote them on social media. A very recent social drive that has been promoted on social media forum is COVID -19 that affects the living and all segments of peoples.
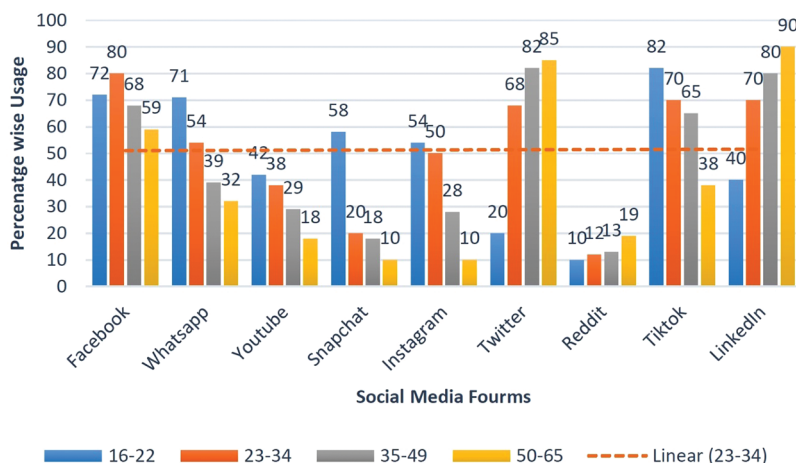

Figure 1: Age wise usage of popular social forums

COVID-19 the popular name of the corona-virus [5] most probably appeared from the city of Wuhan, China last year. In a very short interval of time, it has been designated as one of the most discussed topics on social media and on all other platforms as well. The reason behind this is definitely its wide spreading. Activities have been monitored in order to check the spread of COVID-19 through non-pharmaceutical interventions and preventive evaluations, for instance, social-expelling and self-isolation have induced no matter how you look at it end of the insightful world in excess of 100 nations [6]. Billions of people are allowed to publicly express their understanding about COVID-19. Over the past week, it has been observed that individuals, societies, and organizations are using the web-based lifestyle to spread the familiarity of COVID-19, as just an open-source discussion.

Probably the most unmistakable ways social media has impacted people in general since the infection arrived at scourge and pandemic levels [7], For instance half a month back, a large number of us hadn't knew about "social removing," which alludes to remaining in any event 6 feet from others to help forestall the spread of infection. Presently, social media clients, from loved ones to superstars and governments, have

normally called for social separating. Also, numerous individuals have been unnecessarily buying family unit merchandise, disinfection items, and food in dread that necessities will not, at this point be open—simply as they do when there is a tropical storm or some other cataclysmic event. This over-buying has become so typical that social media users have instituted an adage to portray it: alarm purchasing. Individuals are posting about their own frenzy purchasing, demonstrating pictures of trucks loaded up with bathroom tissue, water bottles, and solidified dinners; and they are likewise posting pictures of void racks or others' trucks as an approach to disgrace guessed alarm purchasers. From this discussion, we have arrived at new factors like social distancing, fear, serene etc. that leads to the conclusion that people have sentiment of panic against COVID-19 [8].

Sentiment analysis is a part of full effective computing research [9] that intends to characterize text in natural language processing. In this research work, the sentiment analysis process has been utilized in a more granular form to extract the influencing factors of pandemic COVID-19. Sentiment analysis along with advanced similarity measures and lexicons has been utilized to analyze the influencing factors of those peoples who are affected by COVID-19 when social media forums promote it as a common social drive. The proposed sentiment analytics model demonstrates that participation of youth of different countries could boost up the promotion of social drives and fasten the process which leads towards positive awareness about any common social drive. The purpose of this research is to utilize the sentiment analysis process for the extraction of influencing factors from any drive, when it promotes across social media. This research work helps to demonstrate the participation of youth of different countries for promoting social drives and fasten the process which leads towards the positive awareness about any common social drive. Moreover, it would cover the gap of common issues on social media that are not in practice in past studies. It also strengthens the awareness phenomena in youth beyond boundaries. From the experimental results and discussion, it has been observed that proposed research performed well in term of accuracy.

The rest of the paper is organized as follows: Section 2 describes the past literature. Section 3 illustrates the proposed conceptual model. Section 4 shows the experimental results and discussion. Conclusion and future work are provided in Section 5.

This research discusses the influencing factors of any social drive on social media forums and social networks. The main contribution of this research work is as follows:

- The key contribution of this study is to describe and evaluate the factor involved in the identification of COVID-19.
- The proposed Sentiment Analytics framework that is composed of influencing factor extraction, influencing factor reduction and factor score calculation utilizes advanced similarity measure and programming-based interaction that explore the new dimension of research in the field sentiment analysis as compare to existing state-of-art techniques.
- Experimental results and evaluation depict that the technique which we have adopted to perform analysis on opinion and reviews performs better than the existing approaches.
- Last but not the least, this study provides a roadmap and empirical proof for future research work in the social media and social networks.

2 Literature Review

The improvement on the internet and its acknowledgment at open level and a short time later it's utilization by ordinary citizens especially youth through online networking which helped in building up a social stage in an interesting space, has assumed a major job in this change [10,11]. The advancements in web-based life, and the cooperation in social exercises has completely changed the current scene into new

measurements in the most recent decade. Presently explicitly the internet-based life has thought about a genuine component in the social and non-social crusades and the way individuals view and handle the issues talked about on this stage in subtleties free of existing requirements. Before-hand, most of the work via social networking platforms particularly on informal organization locales concentrated on the individual uses while a couple of studies have been distributed on the distinctive social skirmishes.

On the relationship between school and civic engagement, Yuan [12] performed research. They analyzed the utilization of microblogs to advance community interest among Chinese undergrad and graduate understudies in his exploration study. An adolescent open stage and video site in China, named as Bilibili with a projectile remarking innovation can help in boosting interest culture is the additional investigation in their work investigated. Likewise, the yield of all preprocessed words utilizing string vectorization, splitter, and stemming has been unmistakably clarifying the preprocessing stage. With the utilization of Stanford Core NLP [13], the stemming procedure is finished, and the content is divided into sentences. For this reason, Porter stemmer and word library have been utilized. Tweets and reviews that have fewer than three words are wiped out. Besides, all the images and accentuation in text are likewise evacuated.

The study performed by Zaheer [14] highlights the importance of social media networks in political campaigns in Pakistani's democratic elections. Their research study is based on a systematic approach for which online Facebook ID's of various persons and parties have been used for the collection of data. They analyzed the data provided by the Facebook about the specific persons and parties. Their findings show the importance of the usage of Facebook. From their research, they revealed that President Asif Ali Zardari has fewer fans than President Parvez Mushraf. Among other parties Pakistan Tehreek-e-Insaf is on top in Pakistan on Facebook is also concluded from this study.

Elsayed [15] explained that social networks hit the Arab world firmly over the most recent couple of months, where youngsters conveyed, teamed up, and shared a wide range of data and records through Facebook and different sorts of SNs. It is not clear how youngsters see this kind of action and how they are affected in their everyday livings. It is critical to see how youngsters see Facebook and comprehend its preferences and hindrances. Their examination used 206 reactions from various classifications basically college understudies (paper study) and Facebook clients (electronic overview) and reasoned that the most astounding apparent bit of leeway of informal community is "I can look and discover new and old companions (cohorts and relatives)" and the most noteworthy saw inconvenience is "Extreme compulsion".

The cited literature has clearly demonstrated the effect of social media and its role in the awareness of social drives. All kinds of issues like pandemic situations, tourism, politics, and terrorism have been promoted by the peoples particularly youth on social media. However, there is a lack of such approaches which demonstrate the conclusion and after effect of any social drive that has been promoted on social media. No machine learning based model along with statistical measures to extract and analyze the influencing factors has been discussed so far. Therefore, there exists a need to explore the research dimension towards the machine learning based analysis of influencing factors when a particular social drive (such as "COVID-19") is promoted on social media forums.

As far as social media discussions and opinions are concerned, sentiment analysis and opinion mining has been well explored by many researchers. Sentiment analysis is a technique that is used to extract opinionated words expressed in reviews and comments. The research study in Samuel et al. [16] has tended to issues encompassing open conclusion reflecting profound worries about Coronavirus and COVID-19, prompting the recognizable proof of development in dread supposition and negative notion. They additionally exhibited the utilization of exploratory and expressive literary investigation and printed information representation techniques, to find beginning phase bits of knowledge. At long last, their work gave an examination of literary order instruments utilized in computerized reasoning applications and

exhibited their helpfulness for differing lengths of Tweets. Along these lines, they guarantee that current examination has given techniques important educational and open notion experiences age potential, which can be utilized to grow truly necessary persuasive arrangements and methodologies to counter the quick spread of "the trio of dread frenzy despair" related with Coronavirus and COVID-19 [17].

Dubey [18] portrayed the conduct of people groups of various nations during the lockdown days due to Coronavirus. His exploration work expounded that individuals have taken informal organizations to communicate their emotions and figure out how to quiet them down. Also, in his examination work he gathered tweets from twelve nations from eleventh March 2020 to 31st March 2020 and are identified with COVID19 in a few or the other way. The Examination has been done to dissect how the residents of various nations are managing the circumstance. In his examination technique, tweets have been gathered, pre-handled, and afterward utilized for text mining and conclusion investigation. The test results presumed that while the dominant part of the individuals all through the world are adopting a positive and cheerful strategy, there are occasions of dread, bitterness and disturb displayed around the world.

The effort in Ahmed et al. [19], attempt to get bits of knowledge into the open response as the revive stage begins. There has been some investigation from web-based life information about how individuals are responding in the lockdown time. We put forth an attempt to comprehend whether there is an adjustment in open supposition from the lockdown stage to the revive stage. We have made our examination on Twitter information on reviving related issues during this COVID-19 flare-up. From our examination, in their work they have discovered that even though individuals are demonstrating an uplifting disposition to revive their territories to make the economy utilitarian, they are additionally encouraging the expert in worry to decide for keeping away from the exceptionally unsurprising second wave. From genuine information, their work has seen that the new cases are expanding as the revive stage begins. Given this circumstance, the coordination of reactions from on the web and from this present reality makes certain to uncover promising outcomes to address the current emergency.

In the investigation of Drias et al. [20] a cultural target was done on individuals trading on informal organizations and more especially on Twitter to watch their emotions on the COVID-19.

More than 600,000 tweets with hashtag #COVID and #coronavirus was compiled between February 27, 2020 and March 25, 2020. A comprehensive investigation into the expansion of the tweets has been completed to produce clear ratings. A follow-up study was followed with positive results. Information mining techniques were integrated with FP-Growth to provide more consistent and effective results. The resulting outcome shows a significant improvement at high rate for traditional media.

From the above discussion, it is concluded that there exist different techniques used for opinion mining including Machine Learning or supervised approaches, lexicon or Dictionary based approaches. Moreover, for prediction most of existing techniques are *corpus* or lexicon based which are unable to deploy in some other domain because they are domain specific i.e., *corpus* built for medical domain cannot be deployed in public analysis and vice versa. Therefore, there exists a need to provide a more refined model that produces better results in term of accuracy and prediction.

3 Proposed Conceptual Framework: Sentiment Analytics

In this research work a sentiment analytics framework based on sentiment analysis has been proposed to explore and analyze the influencing factors of peoples during COVID-19 pandemic situations when it has been promoted as a social drive on social media. The proposed conceptual model helps someone to select appropriate machine learning model for identifying the influencing factors from social media reviews rather than traditional methods like surveys and interviews. An overview of the proposed sentiment analytic based architecture is given in Fig. 2.
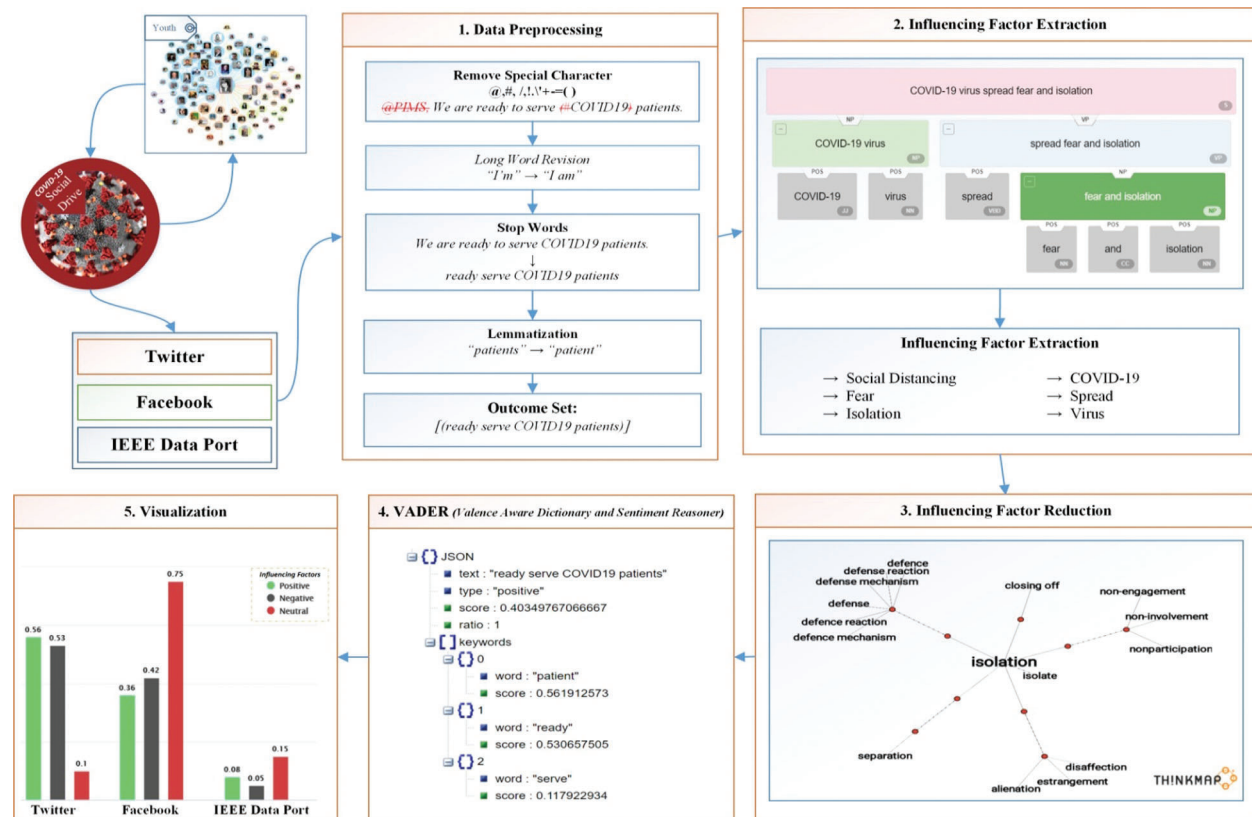
Figure 2: A sentiment analytic based proposed conceptual model

3.1 Social Drive (Collection of Data)

In this section the detail of social drive (dataset) along with extraction procedure has been discussed. As far as the popularity and awareness of this COVID-19 drive is concerned, we have received many discussions in the form of Tweets and post. The extraction and storage of such a huge bulk of data was very hectic task. However, various Application Programming Interface (API) has been utilized to obtain such bulks of data. The detail of each dataset has been described in below sub sections.

3.1.1 Twitter Dataset

Based on terms and conditions for the usage and the security reservations, the greater part of the datasets was acquired from Medicare [21]. However, various programming libraries with application programming interface (API) have been deployed to extract data from Twitter. Furthermore, for appropriate tweet collections, most usable keywords i.e., COVID-19 and Coronavirus along with timeline, location has been applied [22]. As far as the duration is concerned, most of the Tweets were about pandemic season that has not targeted a particular continent, country. Approximately 490,000 Tweets have been collected in all over the world using Coronavirus and Covid-19 keywords. Collected Tweets are stored in. xl format and provided to the Sentiment Analysis library. The description of mentioned dataset is shown in Tab. 1.

3.1.2 Facebook and IEEE Data Port Dataset

Extracting and searching comments and discussion about COVID-19 on Facebook and IEEE data port manually is a hectic task in time perspective with so many limitations [23]. In addition, researchers will not be competent to understand public opinion or mood on a mass scale and their biases may influence their interpretations of what they read [24]. So, in this research, the Graph API has been used to extract data

from Facebook platform. It's a Hyper Text Transfer Protocol (HTTP)-based application that works using query-based programming, through which one can perform various tasks including data management, uploading, and photo processing. Tab. 1 demonstrates about the reviews and comments showing that there are 22,450 reviews of that were collected from groups using #tag of Coronavirus and COVID-19, while miscellaneous reviews are 20,400 in total describing the detail of the dataset used. Certain decomposition rules as described in below, of CoreNLP have been utilized to suit the text fragment into required tokens.

Table 1: Social drive description

#	Keywords	Vocabulary Size	Reviews/Tweets
	Twitter Data Description using		
	https://twitter.com/hashtag/COVID19		
1	COVID-19	88,481	122,716
2	Coronavirus	145,236	365,689
	Facebook Data Description using		
	https://web.facebook.com/hashtag/covid19		
3	#COVID19	275,432	457,287
4	#Coronavirus	24,390	35,800
	IEEE data port Data Description using		
	https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset		
5	SARS-CoV-2	17,521	20,400

- URL based post had removed as it was observed as spam, advertisement or headline and will not consider as a potential opinion.
- All Tweets with "RT" prefix were also discarded as it signifies as "ReTweet" and considered as duplicate.
- Opinion and Tweets that create ambiguity in sentiments like misspelled or digit as a postfix had also removed.
- Full stop, punctuation, articles, and commas were also discarded by using the functionality of Porter Stemmer algorithm.

3.2 Factor Extraction

In this section, the process of influencing factor extraction has been discussed and linked to each word in a strong sentence to think of as an influential or non-influential factor.

Initially, each tweet and opinionated sentence of a specific description identified with the keyword can be considered as a word document WDi. All highlights alongside opinions from the WDi are recognized utilizing the Modified POS tagger. POS tagging tool discovers grammatical form in a sentence alongside linguistic relationship present in a related sentence [25]. Each sentence breaks down into tokens that are labelled as verbs (VB), nouns (NN), numerical phrases (NNP), proper nouns (NNP), adjectives (JJ), articles (DT) and more. Based on the theory of Miyahira et al. [26], each of the entity particularly noun and proper noun could be considered as factor and adjectives to factor are the opinion. From Fig. 3 it has been observed that traditional POS considered and extract all noun, pronoun, adjective and adverb as influencing factor that led to ambiguous result. Although, such techniques work best when many different

factors are fully compatible with certain categories of words or phrases (e.g., nouns), but fail when using low-key words as factor. Tab. 2 shows the list of influencing factors obtained by using POS tagger.
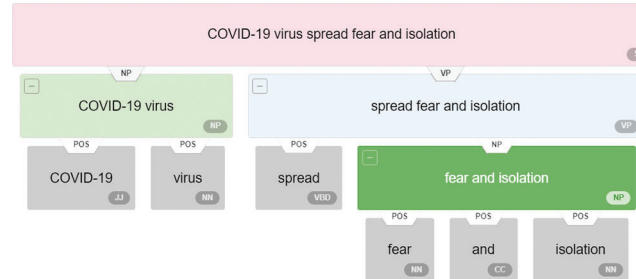

Figure 3: POS based factor extraction process

Table 2: Influencing factors list using modified POS tagging

Possible influencing Factors List			
COVID-19	Fear	Cases	Excellence
Coronavirus	Antibody	Reproductive	Symptomatic
SARS-COVID-19	Vaccine	Rate	Cough
Pandemic	Laboratories	USA	Fever
Wuhan	Epidemic	Jhon	Self
China	Outbreak	Albert	Throat
Incubation Period	Surgical Mask	Doctor	Injection
Sanitizer	Representation	Front line	Vaccine
Isolation	N95 Respirator	Soldiers	Mission
Mask	Spread	Global	Hand
Novel Strain	Threat	Anxiety	Hospital
Community	Death	Economic	Problems
Pathogen	Shortage	Prices	Retail
Advisor	Appearance	Education	Partner
Self-isolation	Diseases	Blogs	Agenda
Self-quarantine	Government	Insights	Reduce
Super Spreader	Test	Rubric	Products
Campaigns	PCR	–`	–

3.3 Factors Refinement

There exist various machine learning techniques that have already extracted and analysed features and factors from text datasets [27]. But they have certain limitations including increase in execution time due to irrelevant and redundant factor set of data, in addition, irrelevant and redundant factor participate in determining the sentiment of the given document, varying the complexity and accuracy of the algorithm.

Factor refinement process decreases the dimensionality of the factor space to obtain most optimal feature to reduce computational cost.

For factors and their association with a particular entity, a semantic similarity based novel similarity measure THINKMAP [28] is used. THINKMAP is an online dictionary used for looking up words to find their meanings and associations with other words and concepts used for factors and opinion association. THINKMAP is source free and available on web with almost 10 million queries ant its solution. The match results against search term are automatically highlighted for synonyms and bold for rhymed. At the top of result list, there are links to near synonyms, homophones, definitions, rhymes, similar sounding words, and most related words. THINKMAP creates word maps that blossom with meanings and branch to associate words. Its innovative display encourages exploration and learning. Fig. 4 shows the interactive visualization of an influencing factor "Isolation" with different synonyms and associative factors using the "rel_jjb" and "rel_jja" constraints following the concerned API, and the D3 visualization library. After the successful utilization of THINKMAP based factor reduction process some of the influencing factors are shown in Tab. 3.
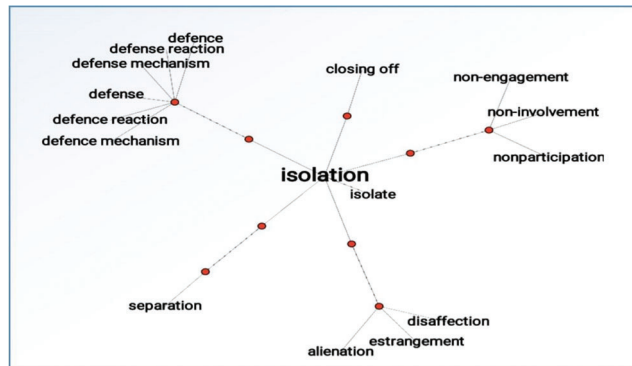

Figure 4: THINKMAP visualization showing factor reduction process

Table 3: Influencing factors list using modified POS tagging

No.	Influencing Factor
1	Isolation
2	Fear
3	Active
4	Nervous
5	Upset
6	Stress
7	Relaxed
8	Calm
9	Excited
.	Alert
.	Serene
.	Elated

3.4 Computing Influencing Score

In this section a modified version of programming base RVADER (Refined Valence Aware Dictionary for Sentiment Reasoning) has been proposed to detect the polarity (e.g., a *positive* or *negative*) of each factor. The traditional version of VADER [29] is fully open sourced under the MIT License. VADER developers have utilized Amazon Mechanical Turk to obtain most of their measurements used for text analysis in the polarity (positive / negative) and the intensity (strength) of emotions. It performs polarity estimation at sentence level that may increase the complexity when the sentence contains wide range of factor or extra symbols as depicted in Fig. 5.


```
#Baseline sentence
sentiment_analyzer_scores('The food here is good')

The food here is good------------------ {'neg': 0.0, 'neu': 0.58, 'pos': 0.42,
'compound': 0.4404}

#Punctuation
print(sentiment_analyzer_scores('The food here is good!'))
print(sentiment_analyzer_scores('The food here is good!!'))
print(sentiment_analyzer_scores('The food here is good!!!'))

The food here is good!----------------- {'neg': 0.0, 'neu': 0.556, 'pos': 0.44
4, 'compound': 0.4926}
None
The food here is good!!---------------- {'neg': 0.0, 'neu': 0.534, 'pos': 0.46
6, 'compound': 0.5399}
None
The food here is good!!!--------------- {'neg': 0.0, 'neu': 0.514, 'pos': 0.48
6, 'compound': 0.5826}
None
```

Figure 5: The impact of increasing the number of (!)

RVADER is a programming-based modification that introduce keywords separator (a python programming function) to extract the relevant keywords with sentiment scores from traditional VADER that relies on a dictionary that maps lexical features. The sentiment score of factors then further is plotted on Vector Space (VS) to show the overall impact of factor set. RVADER as it utilizes the services of VADER is intelligent enough to understand the basic context of these words, such as *"Low Fear"* as a positive statement. It also understands the emphasis of capitalization and punctuation, such as *"ENJOY"*. Fig. 6 shows the programming-based snapshots to calculate the score of different influencing factors of a particular review.


```
text : "COVID-19 creates fear and nervousness."
type : "negative"
score : -0.36685831566667
ratio : -0.45744928877047
keywords
   0
      word : "create"
      score : 0.652660016
   1
      word : "nervous"
      score : -0.938141058
   2
      word : "fear"
      score : -0.815093905
```

Figure 6: Programming based snapshots to calculate the score of influencing factors

The VS is based on the notion of score obtained in the previous step along with polarity. The model assumes that the relevance of a factor to factor is roughly equal to the factor-to-factor similarity. Both the associated factors are representing same influence using the bag-of-words model [30]. A vector is a number that has both a magnitude and a direction. Both magnitude and direction need to be measured with respect to the space in which the vector is defined. Each dimension of the space represents an influencing factor, and a vector represents the extent to which the object of the model has those factors. Thus, a vector is a list of numbers: one for each factor that is part of the model space. The direction of the vector is the one that from the origin of the space to the point defined by those numbers. Fig. 7 shows the VS diagram of influencing factors.
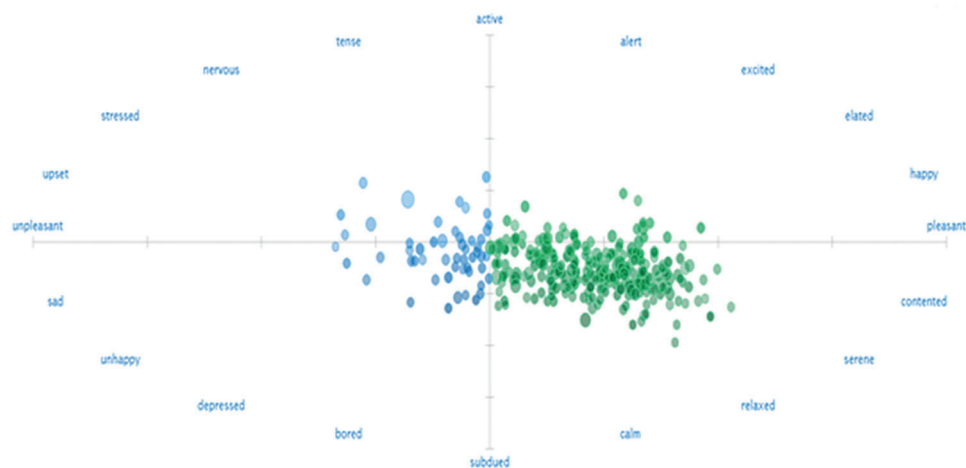

Figure 7: VS diagram of influencing factors

The VS diagram in Fig. 7 depicts the positive (may also be portrayed as pleasant) influencing factors in green colour, while negative (may also be portrayed as unpleasant) in blue colour respectively. Moreover, the mores confidential factors of each category are represented by opaque in the figure, while transparent factors represent less confident estimates.

4 Experimental Results & Discussion

This section describes the experimental results and evaluation of the proposed Sentiment Analytics model using different dataset as discussed in Sections 3.1.1 and 3.1.2 in term of precision, recall and F-measure. The discussion is divided into the following phases of assessment: the initial assessment is an analysis of the factor extraction and factor reduction process. The very next session describes the evaluation of RVADER for the final score and its recommendation. The details of each phase are as follows.

4.1 Experimental Evaluation of Factor Reduction

Numerous experiments have been performed to show the effects of the proposed conceptual model in term of accuracy and performance. In the very first experiment the traditional POS based factor extraction process has been compared with THINKMAP based extraction. Factor reduction reduces the irrelevant and redundant factors that boosted the process of sentiment analysis up to 25%–35%. The comparison of our proposed technique for aspect findings with POS is shown in Fig. 8

Fig. 8 demonstrate the performance of THINKMAP with traditional POS in terms of precision, recall and F-measure. Initially, on Twitter dataset the standard POS produced a recorded precision of 85.04%,

recall 77.45% and F-measure of 81.07%. With the integration of THINKMAP to POS, the value for precision, recall and F-score is 88.72%, 84.51% and 86.56% respectively. When the same has been applied on Facebook dataset, the values for precision, recall and F-score has lied on 89.95%, 84.60% and 82.21% respectively. Thus, the results obtained prove the optimality of the proposed approach. Last but not the least, when the proposed approach is applied on IEEE Data Port dataset, the value for precision is 93.78%, 90.55% recall and 92.14% for F-score. It is clearly observed that the values of precision, recall and F-measure are more promising than the previous two.
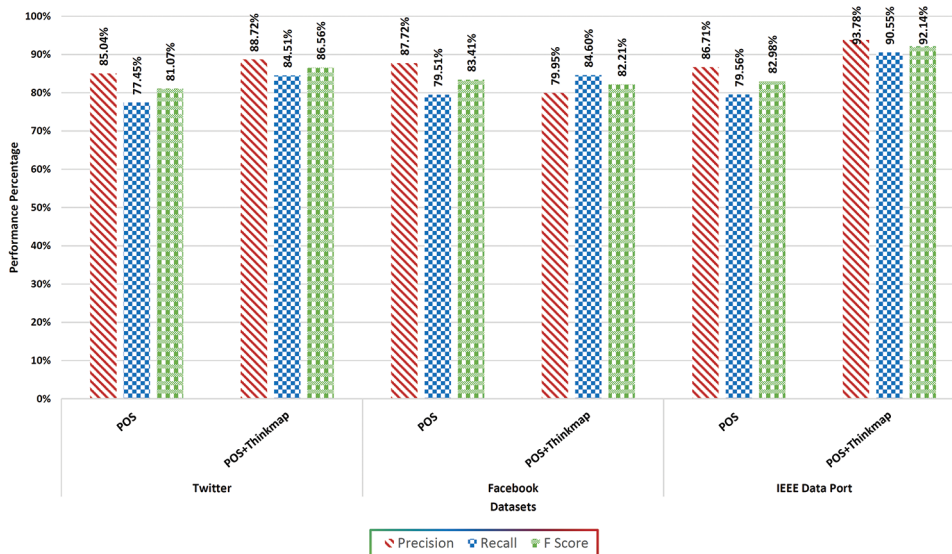

Figure 8: Comparison of aspect extraction and reduction techniques

4.2 Experimental Evaluation of influencing Factor Recommendation

In the next experiment, the performance of the proposed approach with an existing state-of-art method for calculating the influencing score leading to the final recommendation has been proposed. In the comparison of LDA based factor identification and proposed technique on Tweet dataset, it is strongly suggested that the proposed approach achieved 88.34% accuracy compared to the LDA of 82.22%. As far as Facebook dataset, the accuracy of the proposed method is 90.45% compared to the LDA-based production technique of 87.44%. Finally, for IEEE Data Port, again the proposed approach is leading with 88.37% while the LDA scored 78.45%. The graphical representation is shown in Fig. 9. Since the LDA uses POS patterns for detecting single and multiword factor, therefore some of the factors are not covered and thus may lead to lower accuracy results

In comparison with the working of Kleinberg et al. [31], it has been observed that the performance of proposed techniques is quite remarkable. Fig. 10 shows the actual difference that has been observed. In comparison of Sentiment Analytics (Proposed Model) with COVID-19 (NSW) for Tweet dataset [31], the Sentiment Analytics Model achieved 86.24% accuracy in comparison with recorded accuracy i.e., 83.24%. In case of Facebook dataset, the proposed technique produced 89.45% which is almost 4.25% better than COVID-19 (NSW). In IEEE Data Port its performance is approximately 12% better than COVID-19 (NSW).

Finally, the performance of the proposed technique with existing models has been measured in terms of precision, Recall and F-score. The graphical demonstration in Fig. 11 describes the evaluation of the proposed technique when applied on Tweets, Facebook and IEEE Data Port datasets. From this, it has

been clearly demonstrated that Sentiment Analytics produce primitive results on all datasets. On the tweets dataset, the results obtained showed almost 7%–9% betterment in precision, 6.5% to 9.2% in recall and 9.5%–10.5% in F-Score than existing research. The sub figure for Facebook dataset the proposed technique shows 8.5%–9.0% better performance in precision, 6.5–8.5 in recall and almost 6.5% to 9.2% in F-score which is quite impressive. The subfigure for IEEE Data Port shows 86.84% precision score, 85.79% recall and 86.31% F-Score which is quite notable.
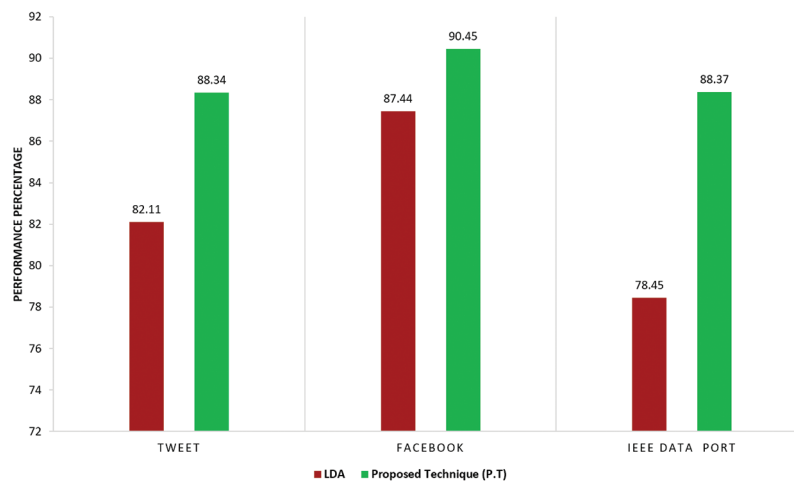

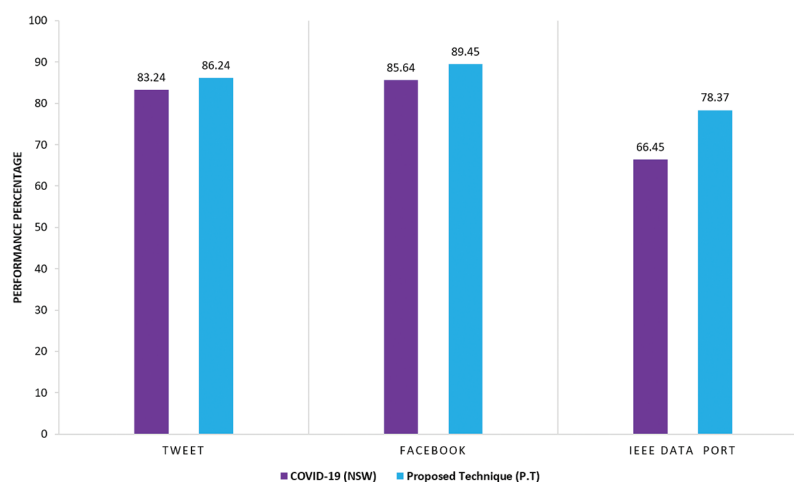**Figure 9:** Comparison of LDA and proposed techniques

Figure 10: Comparison of COVID-19 (NSW) and proposed techniques

From the above analysis and discussion, it has been observed that the proposed technique produced quiet remarkable results. This is because of the accumulation of novel similarity measures and possible improvements in existing work that leads to quite tremendous outcomes.
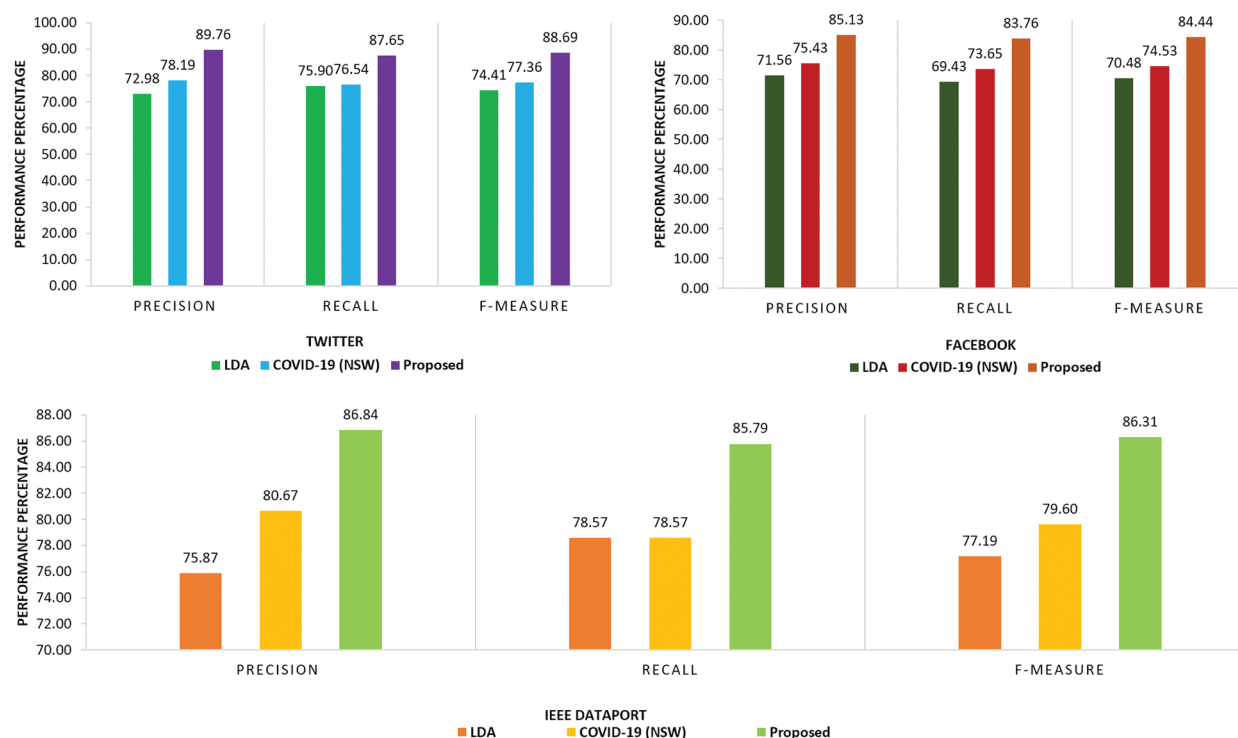
Figure 11: Comparison of proposed technique in term of precision, recall and f-score

5 Conclusions and Future Studies

As far as the communities are concerned, there are so many social drives that need to be discussed on a certain platform for their promotion. Although it's quite challenging but with the advancement of technological propagations and the increasing engagement of the youth on social media towards social drives to promote more democratic forms of engagement between citizens and state made this achievable. The objective of this study is to examine the influencing factors that peoples can express when such drives are promoted on any social media forum. This research has been conducted to explore the influencing factors of pandemic Corona Virus Disease 2019(COVID-19) when it has been promoted as social drive on social media forums. In this research effort, sentiment analysis along with novel similarity measures and lexicons have been utilized to analyse the influencing factors of those peoples who are affected by COVID-19 when social media forums promote it as a common social drive. From the experimental results and discussion, it has been concluded that the performance of proposed techniques is quite outstanding in term of accuracy. In future research direction the current research work could be enhanced to some real time applications models to explore more social drives.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References
[1]   B. Keles, N. McCrae and A. Grealish, "A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents," *International Journal of Adolescence and Youth*, vol. 25, no. 1, pp. 79–93, 2020.

[2]   E. Enginkaya and H. Yılmaz, "What drives consumers to interact with brands through social media? A motivation scale development study," *Procedia-Social and Behavioral Sciences*, vol. 148, no. 1, pp. 219–226, 2014.

[3]   H. O. Witteman, A. Fagerlin, N. Exe, M. E. Trottier and B. J. Zikmund-Fisher, "One-sided social media comments influenced opinions and intentions about home birth: An experimental study," *Health Affairs*, vol. 35, no. 4, pp. 726–733, 2016.

[4]   H. Allcott, L. Braghieri, S. Eichmeyer and M. Gentzkow, "The welfare effects of social media," *American Economic Review*, vol. 110, no. 3, pp. 629–676, 2020.

[5]   C. D. C. Covid and R. Team, "Severe outcomes among patients with coronavirus disease 2019 (COVID-19)," *Morbidity and Mortality Weekly Report (MMWR)*, vol. 69, no. 12, pp. 343–346, 2020.

[6]   C. P. E. R. E. Novel, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases in China," *China CDC Weekly*, vol. 2, no. 1, pp. 113–122, 2020.

[7]   K. E. Riehm, K. A. Feder, K. N. Tormohlen, R. M. Crum, A. S. Young *et al.,* "Associations between time spent using social media and internalizing and externalizing problems among US youth," *JAMA Psychiatry*, vol. 76, no. 12, pp. 1266–1273, 2019.

[8]   G. Petrucci and M. Dragoni, "The irmudosa system at ESWC-2018 challenge on semantic sentiment analysis," in *5th Semwebeval Challenge at ESWC, Heraklion*, Greece, pp. 3–7, 2018.

[9]   S. Li, Y. Wang, J. Xue, N. Zhao and T. Zhu, "The impact of covid-19 epidemic declaration on psychological consequences: A study on active Weibo users," *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, pp. 1–9, 2020.

[10]   F. F. Lotan, "Making a positive internet through aocmed agawe guyub," *International Journal of Communication Systems*, vol. 1, no. 1, pp. 9–16, 2019.

[11]   B. Irawanto, "Making it personal: The campaign battle on social media in Indonesia's 2019 presidential election," *ISEAS Yusof Ishak Institute*, vol. 28, no. 1, pp. 1–11, 2019.

[12]   X. Yuan, "BRTs and investment fads: Civic engagement and fiscal discipline, " in *Fiscal Underpinnings for Sustainable Development in China*. Singapore: Springer, pp. 217–238, 2018.

[13]   L. R. Johnson, J. S. Johnson-Pynn and T. M. Pynn, "Youth civic engagement in China: Results from a program promoting environmental activism," *Journal of Adolescent Research*, vol. 22, no. 4, pp. 355–386, 2007.

[14]   L. Zaheer, "New media technologies and youth in Pakistan," *Journal of the Research Society of Pakistan*, vol. 55, no. 1, pp. 107–114, 2018.

[15]   A. M. Elsayed, "The use of academic social networks among Arab researchers: A survey," *Social Science Computer Review*, vol. 34, no. 3, pp. 378–391, 2016.

[16]   J. Samuel, G. G. Ali, M. Rahman, E. Esawi and Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification," *Information-an International Interdisciplinary Journal*, vol. 11, no. 6, pp. 1–22, 2020.

[17]   Francis Bacon triptych fetches $84 m during live-streamed auction, 2020. [Online]. Available: https://www.independent.co.uk/arts-entertainment/art/news/francis-bacon-triptych-sothebys-auction-how-much-a9594521.html.

[18]   A. D. Dubey, "Twitter sentiment analysis during covid19 outbreak," SSRN 3572023, pp. 1–9, 2020.

[19]   M. E. Ahmed, M. R. I. Rabin and F. N. Chowdhury, "COVID-19: Social media sentiment analysis on reopening. arXiv preprint arXiv: 2006.00804, pp. 1–8, 2020.

[20]   H. H. Drias and Y. Drias, "Mining Twitter data on covid-19 for sentiment analysis and frequent patterns discovery," *medRxiv,* 2020.

[21]   E. Chen, K. Lerman and E. Ferrara, "Covid-19: The first public coronavirus twitter dataset. arXiv preprint arXiv: 2003.07372, 2020.

[22]   L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins and T. P. Debray, "Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal," *BML* pp. 1–11, 2020.

[23]   A. Uribe-Tirado, "Open Science since COVID-19", Open Access+ Open Data, SSRN 3621047, 2020.

[24]   E. Nazaruka, J. Osis and V. Griberman, "Extracting core elements of TFM functional characteristics from stanford corenlp application outcomes," in *ENASE*, pp. 591–602, 2019.

[25] J. R. Bellegarda, U.S. Patent No. 8,719,006. Washington, DC: U.S. Patent and trademark office, 2014. [Online]. Available: https://patents.google.com/patent/US8719006B2/en.

[26] T. Miyahira and Y. Kamiyama, U.S. Patent No. 7,136,806. Washington, DC: U.S. Patent and Trademark Office, 2006. [Online]. Available: https://patents.google.com/patent/US7136806/en.

[27] Y. Li, X. Zhou, Y. Sun and H. Zhang, "Design and implementation of Weibo sentiment analysis based on LDA and dependency parsing," *China Communications*, vol. 13, no. 11, pp. 91–105, 2016.

[28] H. Steinmetz, "How effective are behavior change interventions based on the theory of planned behavior?," *Zeitschrift für Psychologi*, vol. 224, no. 3, pp. 216–233, 2016.

[29] C. J. Hutto, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Eighth Int. AAAI Conf. on Weblogs and Social Media*, vol. 8, no. 1, pp. 2016–2225, 2014.

[30] Y. Zhang, "Understanding bag-of-words model: A statistical framework," *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1, pp. 1–4, 2014.

[31] B. Kleinberg, I. van der Vegt and M. Mozes, "Measuring emotions in the covid-19 real world worry dataset. arXiv preprint arXiv: 2004.04225, 2020.