# Measuring the impact of social drive across social media forums: a case study of COVID-19

Mahmoud Oglah Al Hasan Baniata [1] · Sohail Asghar [1]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract
Users considered Social media forums like Facebook, Twitter and blogs as the most prominent social networks in the present age, where users share their views quickly in words and respond to feedback from other users within no time. This study aims to: measure the impact of influencing factors on a particular community when social media forums promote it using a machine learning model. In this research work, we performed an association rule-based method to measure the impact of COVID-19 influencing factors on adolescence when they promoted it on social media. The proposed method gave a remarkable output when we compared it with the different existing approaches. It works well in all respective fields we observed. Last but not least, when compared with survey and official results, the proposed method predicts well, and the obtained results are pretty promising.

**Keywords** Social media · Facebook · Sentiments · Machine learning · COVID19

## 1 Introduction

Internet users from all over the world involved in the usage of social media platforms, which according to the distinct sources, are around 71%, which is quite huge in numbers [19]. Facebook, a renewed platform is, 60% of the worldwide social network users [34]. For the promotion of social drives, product and services, it is not surprising that students and childhood are widely using social media platforms [1, 25, 33]. Internet users employ social media platforms to trade and create material, and these platforms also help them stay linked to other users, according to Kaplan and Haenlein [2]. Especially in the case of any social drive's promotion, for the benefit of the community, the implementation of social media campaigns brings several advantages with them

✉ Mahmoud Oglah Al Hasan Baniata
 mbaniata2211@gmail.com

 Sohail Asghar
 sohail.asg@gmail.com

[1] Department of Computer Science, COMSAT University, Islamabad, Pakistan

specifically, but not limited to get connected with the people in a very short interval of time [42, 44]. Also, through the social network, the promotion of anything or product or service has a sincere impact on the users and help them to decide on certain things [22].

As far as the communities are concerning, users need to discuss so many social drives on a reliable platform for their promotion.

Although it's very challenging but with the advancement of technological propagations and the increasing engagement of childhood on social media towards social drives to promote more democratic forms of engagement between citizens and state made this achievable. There exist's many studies that perform the said by exploring different social initiatives on social media to get remarkable outcomes [1, 25, 33]. In their work, they try to show the importance of social media for promoting and social drive. However, no study measured the impact of a social Drive that users promoted on social media. This study aims to: measure the impact on young people of selective social drive-by influencing factors. This study evaluates the effects of leading determinants of Corona Virus Pandemic Disease 2019 (COVID-19) in social media forums when users promoted it.

In this research, efforts are to use association rule mining with novel similarity measures and formal concept representation to compute the impact of COVID-19 on Youth when social media forums promote it as a general social drive. From the experimental results and discussion, the performance of the proposed techniques is quite outstanding in term of accuracy. Researchers can improve current research on particular models of real-time applications to explore more social drives in the future. Another possible improvement could be the usage of extracted influencing factors towards computations and analysis.

The arrangement of the rest of the paper is as follows: Section 2 describes the literature review related to the proposed research work. Section 3 explains the proposed conceptual model.

Section 4 shows the experimental results and discussion.

Section 5 provides the conclusion and future work.

## 1.1 Research contribution

The contribution of this exploration effort is as under:

- This study tries to measure the impacts of various factors of any particular drive like COVID 19 on Youth when social media forums promote it.
- The proposed framework uses two essential collections of influencing factors: LF and RF onwards, with an advanced similarity measure and programming based interaction, to explore the new dimension of research compared to existing state-of-art techniques.
- Experimental results and evaluation depict that the proposed technique performs better analysis than the existing approaches.
- Last but not least, for future research work, this research in social media and social networks will provide a roadmap.

## 2 Literature review

There are so many applications where people try to discuss their issues and community problems. However, Facebook, Twitter and blogs are most widely used by most for

communication purpose [5]. People are using Snapchat and Instagram as communication app, through which they are sharing their ideas from day to day and have multiple purposes like sending messages and sharing pictures and videos with friends. There are so many options for teens on social media to browse and search for the content of their own choices, like video streaming and news sharing.

Presently and explicitly, the internet-based life has thought about a genuine component in the social and non-social crusades and how individuals view and handle the issues, talked about on this stage in subtleties free of existing requirements [21]. Beforehand, the vast majority of the work via social networking platforms, particularly informal groups locales, concentrated on individual uses while distributing a couple of studies on the distinctive social skirmishes. On the relationship between school and civic engagement, Yuen [18] performed research. They examined the uses of microblogs to advance community interest among Chinese undergrad and graduate understudies in their exploration study. Likewise, the yield of all preprocessed words practice string vectorization, splitter, and stemming has been unmistakably clarifying the preprocessing stage.

Like many other social drives, during the COVID-19 pandemic, numerous youngsters have been working distantly, schools have moved to online stages, and social separating keeps on being in actuality. In this regard, Youngster uses social media to raise awareness about COVID control, assists in the coordination of humanitarian programs for separated people, and increase assets for crisis professionals. It's also assisting doctors and attendants in saving lives by detecting infections early using clinical imaging and encouraging the total capacity of contact following to limit infection spread.

The initial examination via web-based media during a pandemic goes back to the 2009 H1N1 pandemic, following the pervasiveness of falsehood (decided as 4.5%), phrasing use ("H1N1" versus "pig influenza"), public opinions and dread, and connections between case frequency and public concern [41]. Past examinations used the web to gather information identified with sicknesses, for example, the inquiry recurrence of handwashing, hand sanitizing, and clean points [31]. The WHO proclaimed that they are presently battling a global pestilence in addition to an online media infodemic, with some media guaranteeing that the Covid is the main genuine web-based media infodemic because it has quickened data and falsehood worldwide and is filling frenzy and dread among individuals [17]. Here is a dubious yet testable theory since clients of web-based media use the stages to communicate their feelings, sentiments, and musings, which can be a significant wellspring of information for investigating emotional comfort [41].

ABC News revealed that Covid's unrest spread more quickly than infection, that is, because of social media. [31]. Also, social media is far much helpful in spreading awareness against such a viral disease [17]. Brewer on BBC News [3] sets that the ton of news about COVID-19the has influenced the general population, who made frenzy and building characters alive with strain. Essentially, Rothschild and Fischer [35] asserted that web-based media is spreading apprehension and furore among web-based media clients. Compatibly, through the media which is web-based, Cellan-Jones [4] revealed that the for picking up the data about COVID-19 the people believe on the statistics present on the social media or web, it has been observed that these factors are not much valid, but certain nations have used them as a channel to get the facts.

Millions of people have used online media to check the information about the disease after COVID-19 showed up and transferred to other nations after China. Molla indicated [30], there were 19 million notices of the particular disease in only 24 h across the online media. To

assume the liability of giving the data, the residents have used the comprehensive infrastruc-tures to check the reliability of the data. [15]. Frenkel [13] reported that when the WHO said that web-based media organizations are spreading the COVId-19 negativity, which created a worrying situation, some organizations tried to remove the false information online from their directory following that statement.

Victor [40] claims that in the present advanced age, the Chinese residents have not got realities about the COVID-19, and as a result, they relied on the online media due to this fact. Furthermore, the public authority has asked the social media platforms to stop floating the content about the diseases as it creates a fuss among the people in India. Emmott [12] noted that according to European Union reports, the Russian media spread large-scale false media campaigns on COVID-19 which, disturb the West's lives. In a contemporary conversation on the impacts of media, one analyst [20] has explained in a very diligent way that the web-based media is disrupting people's mind as when they want to buy anything from online stores, they already know about the shortage of the things and are going to buy the grocery items in the bulk which creates the lack of the products for the other people. The web-based media created this situation. As indicated by the paper The Star [39], web-based media is responsible for all the universal circumstances that destroying the foundations of the economies post COVID-19. In addition, Devlin [8] disclosed web-based shopping data identified as food weaknesses that have caused a frenzy.

Also, Kent [23] noticed that web-based media is allowing everyone to see and share information on the web that is the main reason due to which people can find and share everything on the web-based media, apart from certifying whether the information is correct or not and they heard and see about COVID-19 [16]. In the twenty-first century, with the help of innovations used for online communication, the open arena has changed. A significant wellspring of new media for the data wellbeing besides phase aimed at examining individual encounters, sentiments, and concerns for wellbeing, ailments, and treatment has been intro-duced [16].

Additionally, Dillon [9] find out that in this pandemic, people had wasted a lot of their energy and time on social media to make people frenzy and increased the purchasing between the different nations which, creates more fuss. Furthermore, El-Terk [11] revealed that, in this pandemic, everyone is trying to be the expert because everyone has the ability and has the right and potential as well to communicate in a voice or digital form to convey his message in COVID-19. Correspondingly, Garrett [14] made it clear that we provide the people to create the alarming situation on the web as we are the ones who have provided them with the latest data and information about the pandemic.

People today know many means for imparting and disseminating material, according to Merchant and Lurie [28]. that is owing to the advent of web-based media. These are fast and impressive and spread as quickly as false information does. Also, La et al. [24] said that there are so many nations that have limited access to the data and not on the data which, is not beneficial for them and in this way, they only access the information which is helpful for them and not for the harm or disaster because the general society has limited access over the COVID-19 flare-up. For the management of line data correctly, the Vietnamese case is a productive image. The nation's Ministry of Health's web-based broadcasting networks made the information available to the general public.

According to Mian and Khan [29], there has a significant increase in the distribution of fake news and incorrect information on COVID-19, i.e., the lab's hypothesis on the infection's origin began via the web. Correspondingly, Petric and others [32, 38] accept that "media

inclusion has featured COVID-19 as a one of a risk". Depoux and others [7, 36] established that web-based broadcasting supposed three fundamental parts in numerous countries. Users circulated the realities about the episodes via online media.

We learned from the preceding discussion that we need to measure the influence of influencing factors on Teens and conduct a detailed analysis [27, 37]. Therefore, it aimed to covert the influencing factors into the proper model, based on the reviewed literature, establishes which influence factors of COVID-19, and a huge impact on a particular group of Teens. This study tries to measure the effect of the influencing factor of COVID 19 on Youth when social media forum promoted it.

## 3 Proposed methodology

In this research work, we proposed a phase-wise analytics framework to explore and analyze the impact of influencing factors of COVID-19 on Youth when it promoted on social media. The following subsections explore the detailed description of each phase.

### 3.1 Parameter assumption

The improvement on the Internet, and its approval at the public level and thereafter, its usage by the common public, specifically young generation and students and communities through social media, this research study assumed three different categories i.e. *Y1*, *Y2* and *Y3* of Youth based on age, particular media usage and gender, Fig. 1 shows The parameter wise description of these.

### 3.2 Data collection

In this section, we elaborated a detailed description of influencing factors (data collection) and gathered Two different types (as described by equation1) of factors to evaluate the proposed model.

$$
\begin{aligned}
X &= \{x_1, x_2, x_3, ..., x_n\} &&\forall\ x_n \in LF \\
Y &= \{y_1, y_2, y_3, ..., y_n\} &&\forall\ y_n \in RF
\end{aligned}
\tag{1}
$$

Whereas $X$ are the factors that have obtained from past literature *(LF)*, $Y$ are those factors that are obtained from the previously proposed machine learning model *(RF)*.

Initially, we designed an inclusive search strategy to extract more relevant factors from LF: Scopus, Web of Science (all databases), and Google scholar for full text and abstracts. We searched for relevant articles published between December 2019 to November 2020 using various search terms: "COVID 19", "Influencing factors of COVID 19", "Risk factors of COVID 19, etc. combined with OR. Hits from all of the databases exported to an EndNote Library, with duplicates removed. Figure 2 depicted the search strategy for LF data collection. Whereas collecting only original research articles like conference abstracts, letters, and reviews. Favouring the studies in the English language that contain all COVID 19 factors and that influencing the community. Table 1 shows some of the resultant factors obtained from LF.

Another list of influencing factors may found in "Sentiment Analytics: Extraction of Challenging Influencing Elements from COVID-19 Pandemics," a machine learning-based
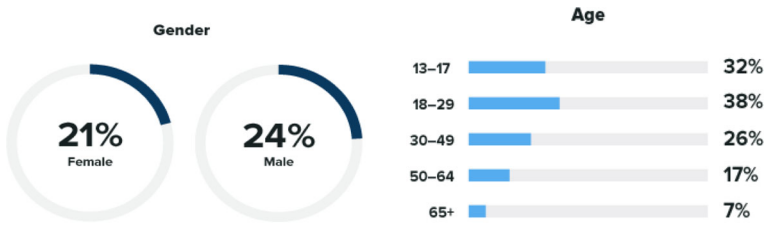
(a) Gender wise ratio of Y1, Y2 and Y3          (b) Age-wise ratio of Y1, Y2 and Y3
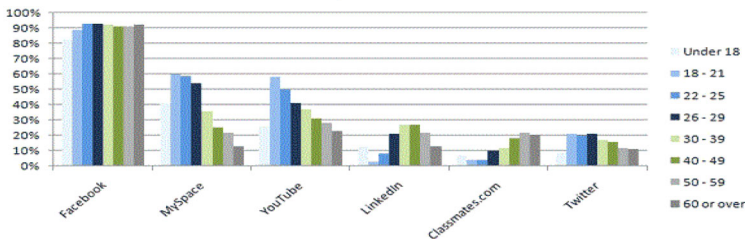


(c) Usage wise ratio of Y1, Y2 and Y3

**Fig. 1** **a** Gender wise ratio of Y1, Y2 and Y3, **b** Age-wise ratio of Y1, Y2 and Y3, **c** Usage wise ratio of Y1, Y2 and Y3

model that uses the sentiment analysis technique to extract various influencing factors from reviews and comments on various web portals. Table 2 shows some of the factors obtained by using the machine learning model.
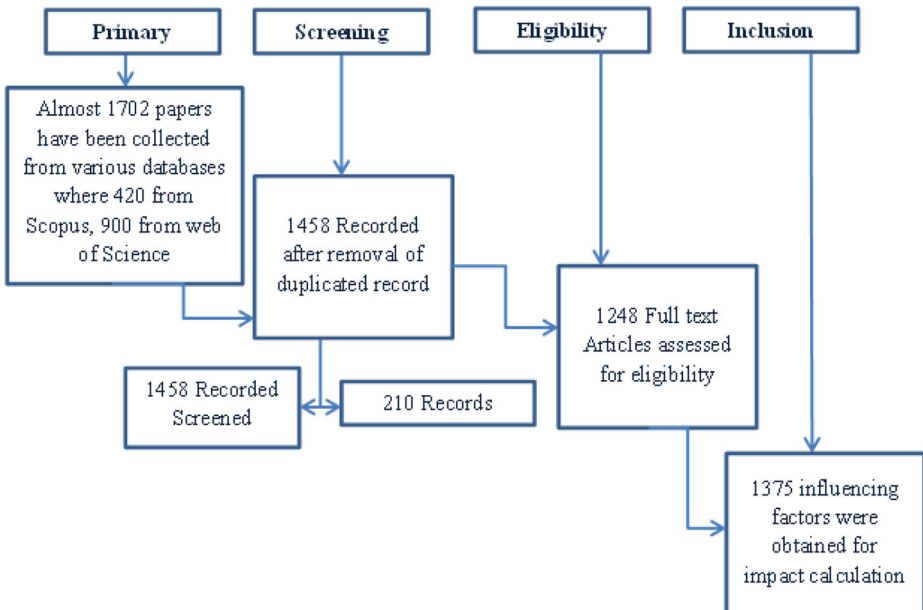


**Fig. 2** Flow diagram of search result

**Table 1** shows the list of obtained factors along with extracted references

| F. ID | Influencing Factor | Referenced Literature | Impact on Community |
|---|---|---|---|
| Psychological Factors | | | |
| P.F1 | Fear cases | [18, 21] | • Human psychology |
| P.F2 | Stress in body | [21] | • Mental health disorder |
| P.F3 | Upset | [18] | |
| P.F4 | Stress | [17, 21, 41] | |
| P.F5 | Nervous situations | [41] | |
| P.F6 | Excited | [31, 41] | |
| P.F7 | Elated | [21] | |
| P.F8 | Relax | [31, 41] | |
| P.F9 | Serene | [41] | |
| P.F10 | Calm | [17, 31] | |
| Business Related Factors | | | |
| B.F1 | Isolation | [3] | • Business Degradation |
| B.F2 | Quarantine | [4, 5, 35] | • Low Economy |
| B.F3 | Home delivery | [4, 35] | • Currency decrease |
| B.F4 | Transport Restriction | [30, 35] | |
| B.F5 | Stay at home | [3] | |
| B.F6 | Lack of public event | [5, 30] | |
| Information Technology Factors | | | |
| I.F1 | Distance Learning | [21] | • Electronic devices get expensive |
| I.F2 | Surveillance | [18, 31, 41] | • Distance learning programs |
| I.F3 | Work from home | [18, 31] | • E-learning |
| I.F4 | Misinformation | [21, 41] | |
| I.F5 | Relax | [41] | |
| I.F6 | Playing Games | [18, 41] | |
| Sociological Factors | | | |
| S.F1 | Events banned | [17] | • Business Degradation |
| S.F2 | Social Gathering | [3, 4, 35] | • Sports Lack |
| S.F3 | Lack of Sports event | [4, 35] | |
| S.F4 | Isolation at home | [30, 35] | |
| S.F5 | Stay at home | [3] | |
| S.F6 | Self-Quarantine | [4, 30] | |

## 3.3 Computing impact of influencing factors

To investigate which factors more influence the Youth, we chose a two-step approach. First, we performed a systematic literature review and selected articles describing influencing factors. Second, we performed a qualitative synthesis of these factors. The following subsections describe the detailed description of each section of the proposed two-step approach as depicted in Figs. 3 and 4:

**Table 2** Influencing factors list using machine learning model

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

### 3.3.1 Data input

This research is to determine the impact of influencing factors of COVID 19 on a particular category of Teens when user promoted it as a social drive on social media forums. The first input dataset contains almost 1310 influencing factors, was extracted from past literature using a comprehensive literature review process as described in section II under the title "Data Collection". The second input dataset (almost 900 influencing factors) is from an early proposed sentiment analytics-based model that extracts influencing factors from a corpus-based on Tweets and Facebook reviews by using corresponding API (Application Programming Interface).

### 3.3.2 Formal representation

As far as, the inputs section is concerned, most of the obtained factors whether, obtained from past literature or sentiment analytic based model, may contain redundancy and irrelevant, which may lead to computational complexity. Therefore, there exists a need to represent it formally. To resolve this case, we proposed a Formal Concept Analysis (FCA). FCA can is a powerful tool like semantics, terms and metadata to structure and classify a set of resources represented through a set of attributes automatically. FCA is a mathematical theory of concept formation [43] that originates, through the kit of ordered arguments and lattice, with the help of
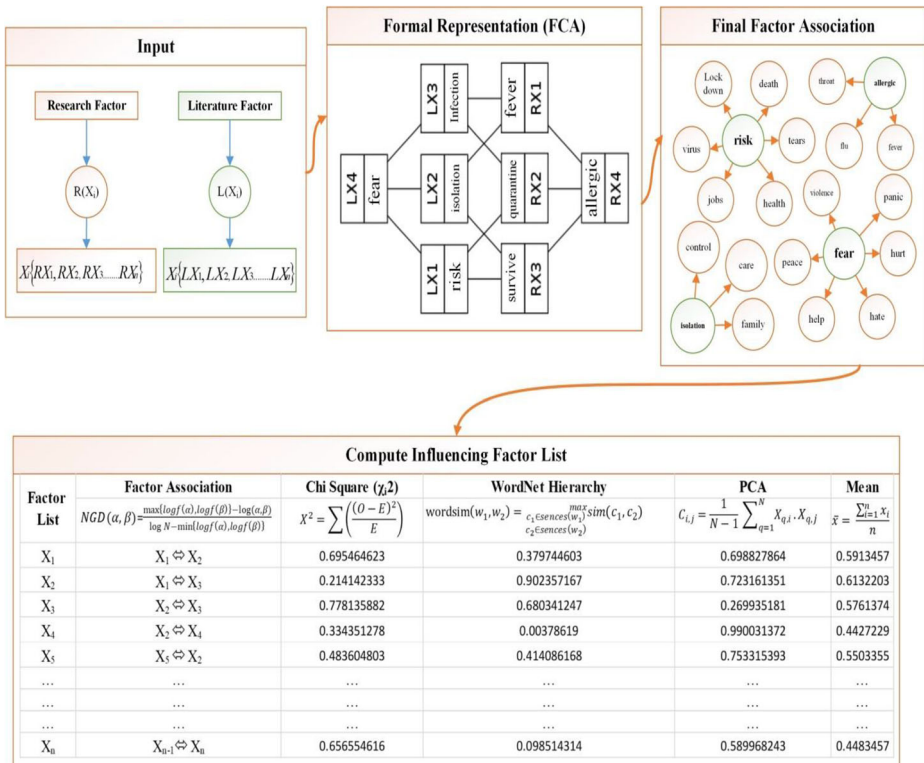


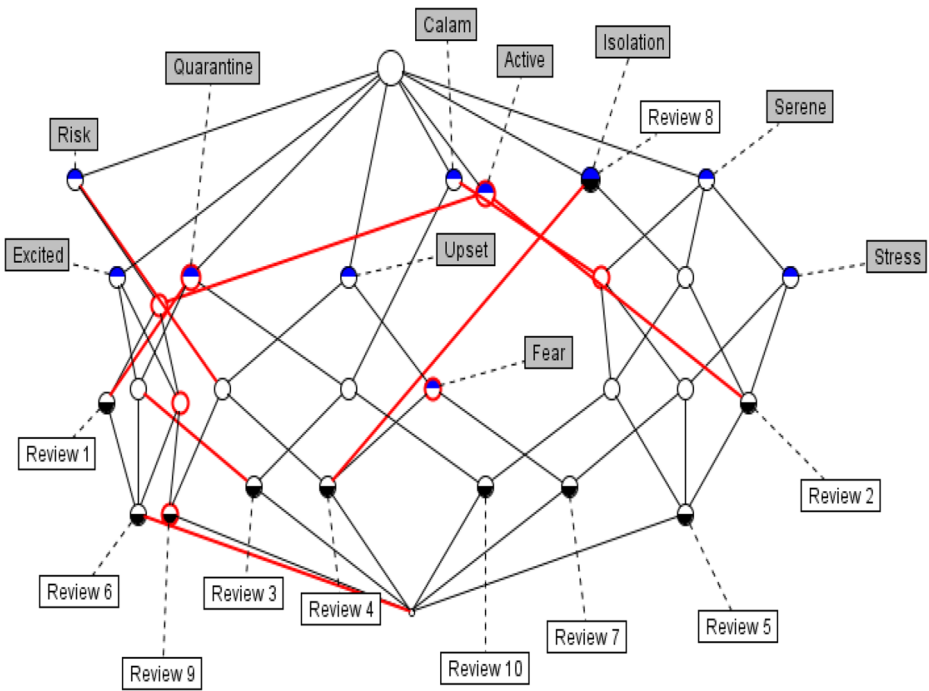| | Compute Influencing Factor List | | | | |
|---|---|---|---|---|---|
| Factor List | Factor Association $NGD(\alpha, \beta) = \frac{\max\{logf(\alpha), logf(\beta)\} - \log(\alpha, \beta)}{\log N - \min\{logf(\alpha), logf(\beta)\}}$ | Chi Square ($\chi_i 2$) $X^2 = \sum \left( \frac{(O-E)^2}{E} \right)$ | WordNet Hierarchy $wordsim(w_1, w_2) = \frac{max}{c_2 \in sences(w_2)} sim(c_1, c_2)$ | PCA $C_{i,j} = \frac{1}{N-1} \sum_{q=1}^{N} X_{q,i} . X_{q,j}$ | Mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ |
| $X_1$ | $X_1 \Leftrightarrow X_2$ | 0.695464623 | 0.379744603 | 0.698827864 | 0.5913457 |
| $X_2$ | $X_1 \Leftrightarrow X_3$ | 0.214142333 | 0.902357167 | 0.723161351 | 0.6132203 |
| $X_3$ | $X_2 \Leftrightarrow X_3$ | 0.778135882 | 0.680341247 | 0.269935181 | 0.5761374 |
| $X_4$ | $X_2 \Leftrightarrow X_4$ | 0.334351278 | 0.00378619 | 0.990031372 | 0.4427229 |
| $X_5$ | $X_5 \Leftrightarrow X_2$ | 0.483604803 | 0.414086168 | 0.753315393 | 0.5503355 |
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |
| $X_n$ | $X_{n-1} \Leftrightarrow X_n$ | 0.656554616 | 0.098514314 | 0.589968243 | 0.4483457 |

Fig. 3 Proposed conceptual Model Phase-1

**Fig. 4** FCA based Factor Lattice

a theoretical model for discovering relationships and knowledge organization. For this, two definitions are introduced based on Wang's paper [10].

1.  *A formal context based on factors association defined as a triple: F C = (S, T, I), where S is a set of factors obtained from past literature, T is a set of collected factors by using sentiment analytics model, and I is a binary relation between S and T. If s ∈ S, t ∈T and t occur in the document s, we say that s possessest expressed by s It or (s, t) ∈ I.*

Given a formal context F C = (S, T, I), if X⊂S, Y⊂T, two functions ↑and ↓ are defined as follows.

$$\uparrow: F(S) \to F(T), X^{\uparrow} = \{t \ \varepsilon \ T; \forall s \ \varepsilon \ X, (s,t) \ \varepsilon \ I\}$$
$$\downarrow: F(T) \to F(S), Y_{\downarrow} = \{s \ \varepsilon \ S; \forall t \ \varepsilon Y, (s,t) \ \varepsilon \ I\} \tag{2}$$

2.  *Given a formal context F C = (S, T, I) if an ordered pair (X, Y) ∈ F (S) × F (T) makes X↑ = Y and Y ↓ = X, (X, Y) is a formal concept of F C. X is the extent of this concept and Y is the intent.*

Consequently, only when several factors in the intent of a concept is two or three and to the extent, that the number of factors is more than the threshold, the notion can be retained otherwise excluded from the cluster. We considered selected factors as potential factors for further steps. Based on the proposed formal context, a unique representation known as concept

lattices represented a descriptive representation of all influencing factors. Based on the theory of FCA, Fig. 4 shows some of the derived lattices.

### 3.3.3 Computing influencing factor list

After the formation of the concept lattice, the obtained factors are needs to be refined. In this paper, we utilize four novel association based measure to choose the most becoming influencing factors. This section provides details of the proposed factor selection methods.

**Normalized Google distance (NGD)** When compared to the semantics of certain other expressions and phrases, the setting of words and expressions vary according to their use in everyday life. NGD consequently separate those pages relating to specific term affiliation, utilizing Google page includes appeared in Algorithm 2. [6] proposed Page tally affiliation estimated among various terms using Normalized Google Distance (NGD). Discovering the relationship amongst dual-distinct words utilizing NGD doesn't need some foundation learning contrarily, a specific investigation of the issue area. Preferably, it consequently examinations all highlights over Google search using World Wide Web. To get a particular event represented in a review, apply the social page count technique (SPC) based on NGD. NGD calculates the weights of all the aspects extracted from the LDA process. Table 4 shows the identified characters after using LDA. NGD works by checking the number and nature of connections to a page to decide an unpleasant gauge of how vital the aspects are for the event. Equation 3, 4 defines the Normalized Google Distance (NGD).

$$NGD(\alpha,\beta) = \frac{f(\alpha,\beta) - \min\Big((f(\alpha),f(\beta))\Big)}{\max\Big((f(\alpha),f(\beta))\Big)} \tag{3}$$

and

$$NGD(\alpha,\beta) = \frac{\max\{\log f(\alpha),\log f(\beta)\} - \log(\alpha,\beta)}{\log N - \min\{\log f(\alpha),\log f(\beta)\}} \tag{4}$$

Term $\alpha$ and $f(\alpha, \beta)$ indicate association of both LF and LR factors, $\alpha$ and $\beta$ reported by Google where $f(\alpha)$ represents the number of pages. We increase NGD as it seems that by decreasing N. The NGD has some properties which smeared in this experiment as below:

1.  Between 0 and ∞, the approximate value of the NGD lies but if the Google search count irrelevant score, it is a little bit negative when:

$$f(\alpha,\beta) > \max\{f(\alpha),f(\beta)\} \tag{5}$$

(a)  if the frequency $f(\alpha) = f(\beta) = f(\alpha, \beta) > 0$, then NGD($\alpha, \beta$) = 0, in all situations of $\alpha$ and $\beta$.

(b)  we have NGD($\alpha, \beta$) = ∞ /∞, if the frequency $f(\alpha) = 0$ then for every search term $\beta$

2.  For every $\alpha$, the value of NGD is almost non-negative and NGD($\alpha$, $\alpha$) = 0. We haveNGD($\alpha$, $\beta$) = NGD($\beta$, $\alpha$)for every pair of$\alpha$ and $\beta$. For example choose $\alpha \neq \beta$ with x = y, then f($\alpha$) = f($\beta$)= f($\alpha$, $\beta$)andNGD($\alpha$, $\beta$) = 0. Nor does the NGD satisfy triangular inequality:

$$\text{NGD}(\alpha, \beta) \leq \text{NGD}(\alpha, \eta) + \text{NGD}(\beta, \eta) \text{ for all } \alpha, \beta, \eta \qquad (6)$$

where $\alpha, \beta, \eta$ represents factors.

**Chi-Square test** A chi-square test for independence compares two influencing factors in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of categorical variables differ from each other. To evaluate Tests of Independence, using a cross-tabulation (also known as a bivariate table), we use the Chi-Square statistic. With the intersections of the categories of the variables appearing in the table columns, Cross tabulation presents the distributions of two categorical variables simultaneously. The pattern that would be expected, if the variables were truly independent of each other and in this case, the Test of Independence assesses whether an association exists between the two variables by comparing the observed pattern of responses in the cells. The comparison and calculation of this test and its comparison with the critical value gives the freedom to the researchers to observe the counting of the cells, which is quietly different from the expected cell count. It is quite straight-forward and intuitive when it comes to the statistics of the Chi-Square:

$$\chi\mathbf{2} = \sum \frac{(f_o - f_c)\mathbf{2}}{f_c} \qquad (7)$$

The abbreviation fo denotes the observed frequency, and fe represents the expected frequency if no relationships exist between the variables. Equation 7 clearly shows the difference between the observed and the predicted frequency on which the Chi-Square statistics based.

**WordNet hierarchy** WordNet is a huge lexical database of English. Nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms (synsets), separately expressing a distinct concept. Synsets are interlinked using conceptual-semantic and lexical relations. The resulting network of meaningfully related words and notions can navigate with the browser (link is external) along with its numerical values. WordNet superficially resembles a thesaurus, in that it groups words based on their meanings and numerical values. However, there are some significant distinctions.

–   WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found close to one another in the network are semantically disambiguated.
–   WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus do not follow any definite pattern other than meaning similarity.

Based on the above description, Table 3. depicted the numerical values along with influencing factors.

**Principle Component Analysis (PCA)** Principal components analysis (PCA, for short) is a variable-reduction technique that shares many similarities to exploratory factor analysis. It aims to reduce a larger set of variables into a smaller set of 'artificial' variables, called 'principal components', which account for most of the variance in the original variables. In this section, for PCA based influencing factor reduction, we construct a $d$ x $k$–dimensional transformation matrix $W$ that allows us to map a sample vector $x$ onto a new $k$–dimensional feature subspace that has fewer dimensions than the original $d$–dimensional feature space:

$$x = [x_1, x_2, .............., x_d], \quad x \ \varepsilon \ R^d \quad \downarrow xW, W \ \varepsilon \ R^{d \times k} \tag{8}$$

$$z = [z_1, z_2, .............., z_k], \quad z \ \varepsilon \ R^k \tag{9}$$

As a result of transforming the original $d$-dimensional data onto this new $k$-dimensional subspace (typically $k < < d$), the first principal component will have the largest possible variance, and all consequent principal components will have the largest variance given the constraint that these components are uncorrelated (orthogonal) to the other principal components — even if the input features are correlated-, the resulting principal components will be mutually orthogonal (uncorrelated). Based on the above description, Table 3 depicted the numerical values along with influencing factors.

### 3.4 Impact of influencing factor –Phase-II

This section provides a deep insight into our final impact calculation. We would like to construct an association rule model based on information generated from reviews. Finding the correct impact of influencing factor is just one concern of our work. In addition, we are

**Table 3** Computing score based on described measures

| Factor ID | NGD | Chi-Square | WordNet | PCA | Mean |
|-----------|-----|------------|---------|-----|------|
| P.F1 | 1.3903345780 | 1.6948789377 | 5.1914481541 | 1.2838623643 | 1.9894 |
| P.F2 | 2.0578876008 | 5.2173003153 | 3.6731725995 | 0.9207583282 | 6.2421 |
| P.F3 | 1.4817756697 | 2.9469864942 | 4.4118349316 | 1.1894695436 | 6.9543 |
| P.F4 | 1.0081729086 | 1.6819896242 | 2.9383204467 | 0.9486935333 | 3.8320 |
| P.F5 | 1.7196937524 | 4.8807995178 | 2.7366824462 | 4.0455121214 | 5.5121 |
| P.F6 | 1.3779735437 | 1.2055199135 | 1.6108231000 | 2.8908355369 | 9.0835 |
| P.F7 | 2.2924662569 | 2.8908355369 | 1.4457943846 | 1.2057559135 | 3.5437 |
| P.F8 | 2.4761474339 | 4.0455121214 | 2.1393965950 | 0.4880795178 | 3.9659 |
| P.F9 | 2.7028171305 | 3.8256285962 | 1.1907947332 | 1.6819467242 | 6.8198 |
| B.F1 | 2.9469864942 | 2.8910244628 | 0.2446239659 | 1.4108231011 | 3.6558 |
| B.F2 | 1.68198946242 | 1.4216946243 | 1.4261946466 | 0.7619896412 | 4.6232 |
| B.F3 | 4.2763449721 | 2.2014199444 | 0.2626231044 | 2.1428355142 | 5.1256 |
| B.F4 | 5.1157400000 | 1.8080355221 | 1.2314943900 | 1.1717556020 | 3.8471 |
| B.F5 | 5.5506365440 | 3.1221212001 | 2.2214965600 | 3.1423795222 | 3.5555 |
| B.F6 | 2.4407172609 | 1.7219896214 | 3.1414044123 | 2.1721935220 | 3.8741 |
| I.F1 | 2.4137937522 | 2.4242995232 | 1.1422822242 | 2.1720121122 | 2.5147 |
| I.F2 | 1.7474535441 | 1.4014199270 | 2.1108231011 | 2.2748355116 | 4.0417 |
| I.F3 | 2.4213662423 | 3.1320155450 | 1.7757942275 | 1.1412559707 | 3.4102 |
| I.F4 | 3.2143472323 | 2.1402121333 | 2.1423265214 | 0.2022791414 | 2.9659 |
| I.F5 | 1.6060171200 | 3.1414285700 | 3.1720941612 | 1.1419467122 | 2.5841 |

interested in identifying the relationship between the influencing factors and their impact on Youth. Thus, we propose a factor hierarchy based method to show the impact of influencing factors on Youth. The whole process consists of two main steps: impact hierarchy construction using Min-Max association rules and hierarchy expansion based on influencing factors. The input of our system is a collection of LF and RF factors list. The output is an impact hierarchy that contains not only LF and RF but also the impact on youth. Figure 5 shows a proposed model in Fig. 5.

To obtained better results, each of the influencing factors is stored in a bag of words where it represents a unique factor id. A particular factor may put a high impact on one category of youth while it may not affect other categories. Understanding these gathering patterns can help to increase the accuracy of work in several ways. If there is a pair of items, X and Y that are frequently bought together the association rules analysis help to uncover how items are associated with each other? There are two common ways to measure association.

**Measure 1: support** This says how popular influencing factors set are, as measured by the proportion of reviews in which factor set appears. Mathematically, support is the fraction of the total number of review in which the factor set occurs.

$$Support = \frac{Reviews\ containing\ both\ LF\ and\ RF}{Total\ Number\ of\ Reviews} \tag{10}$$

If someone discovers that the impact of a particular factor is beyond a certain proportion tend to have a significant impact on youth, then this might consider using that proportion as your support threshold. You may then identify factor sets with support values above this threshold as significant factor sets.

**Measure 2: confidence** This says how likely a particular factor like *LF* is considered when any other factor such as *RF* is selected, expressed as {*LF- > RF*}. This is measured by the proportion of review with item *LF*, in which item *RF* also appears. One drawback of the confidence measure is that it might misrepresent the importance of an association. This is because it only accounts for how popular *LF* is, but not *RF*. If *LF* is also very popular in general, there will be a higher chance that a review containing *LF* will also contain *RF*, thus inflating the confidence measure. Mathematically:
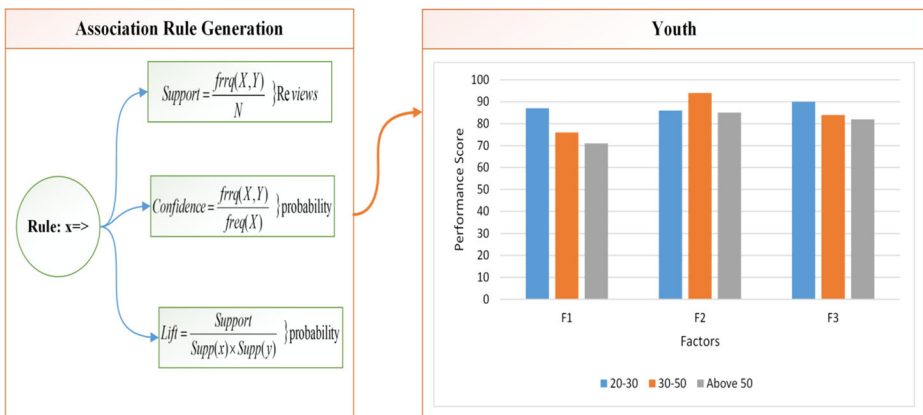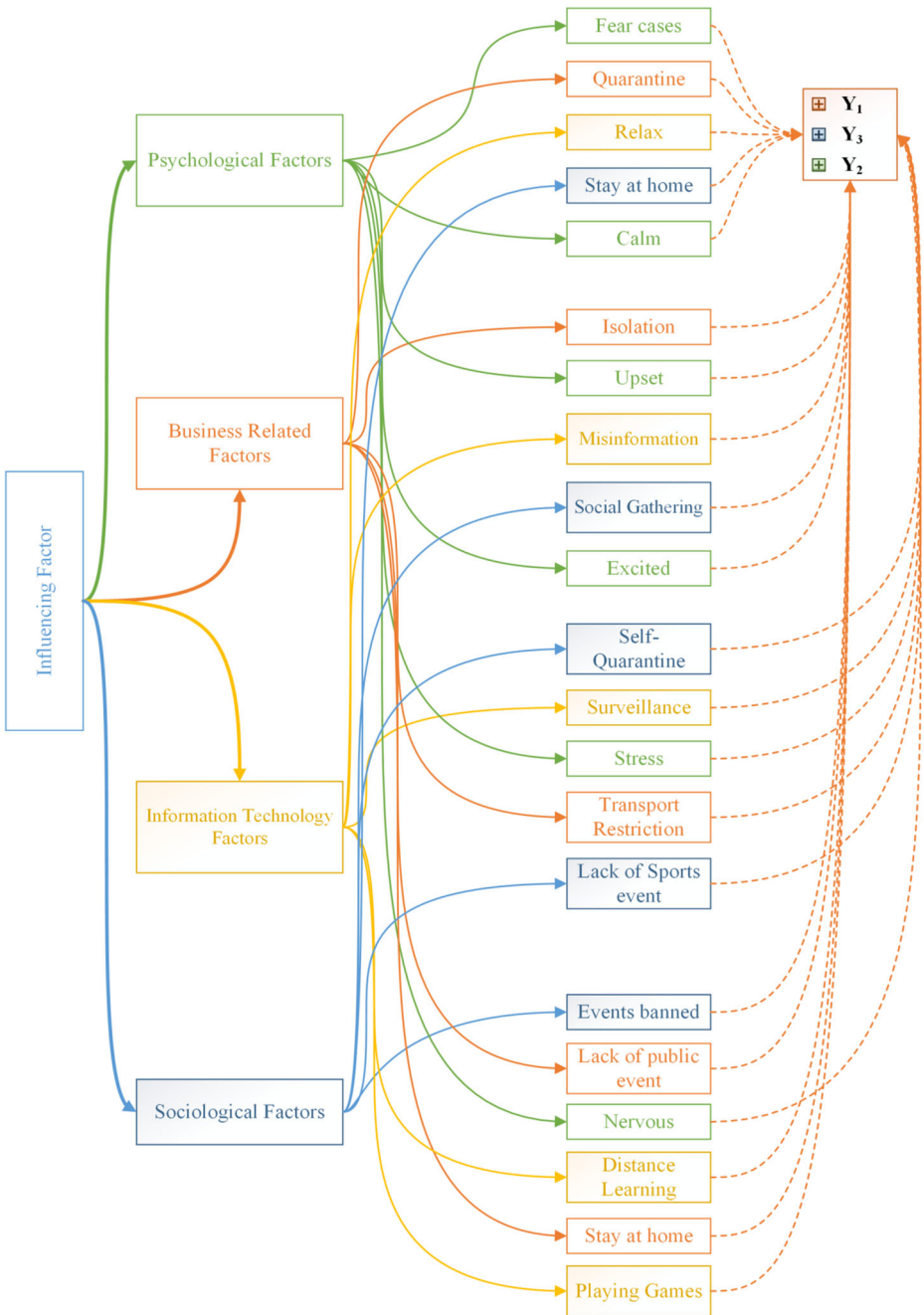


**Fig. 5** Proposed conceptual Model Phase-II

**Fig. 6** Visualization of COVID-19 factor impact on Youth

$$\text{Confidence } (\{\text{LF} \leftrightarrow \text{RF}\}) = \frac{Review\ containing\ both\ LF\ and\ RF}{Review\ containing\ RF} \qquad (11)$$

**Table 4** Computing score based on described measures

| # | Dataset | F-measure | Precision | Recall |
|---|---------|-----------|-----------|--------|
| 1 | extracted Influencing factors from past literature (dataset 1) | 85.89 | 87.56 | 90.23 |
| 2 | extracted Influencing factors Social media (dataset 2) | 86.23 | 88.72 | 91.48 |

Associations between selected items. Visualized using the rules VizR library and is depicted in Fig. 6.

## 4 Experimental results and discussion

In this section, we evaluated the working of the proposed conceptual model and performed Various experiments in term of precision, recall and f measure to check the efficiency of the proposed model. We used Two different datasets: extracted influencing factors from past literature (Dataset 1) and extracted influencing factors from tweets and Facebook posts (Dataset 2). We performed Separate experiments on both datasets. Table 1 shows Results.

Table 4 demonstrate the performance of the proposed method in terms of precision, recall and F-measure. Initially, the experiment runs on Dataset 1 (Influencing factors extracted from past literature) produced a recorded precision of 87.56%, recall 90.23% and F-measure of 85.89%. When the experiment runs on Dataset 2 (Social media dataset), the values for precision, recall and, F-score are 88.72%, 91.48% and 86.23%, respectively. Thus the results obtained prove the optimality of the proposed approach. Figure 7 shows a Graphical representation.

### 4.1 Evaluation of proposed technique with COVID youth mental health impact [26]

In the first experiment, we compared the new methodology to previous research on teenage mental health. The research conducted a cross-sectional study in which Teen of 584 enrolled in the study and about the General Health Questionnaire (GHQ-12), the cognitive status of COVID-19, the Negative coping styles scale and the PTSD Checklist-Civilian Version (PCL-C), they have completed the questions about them.
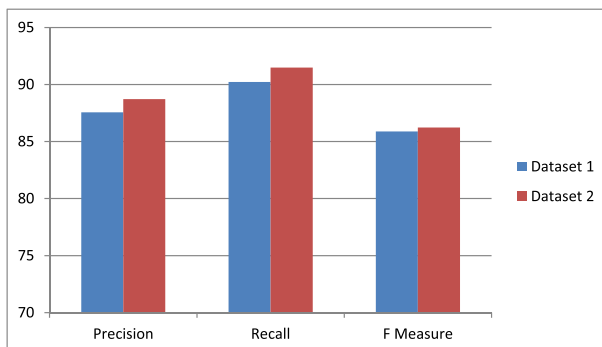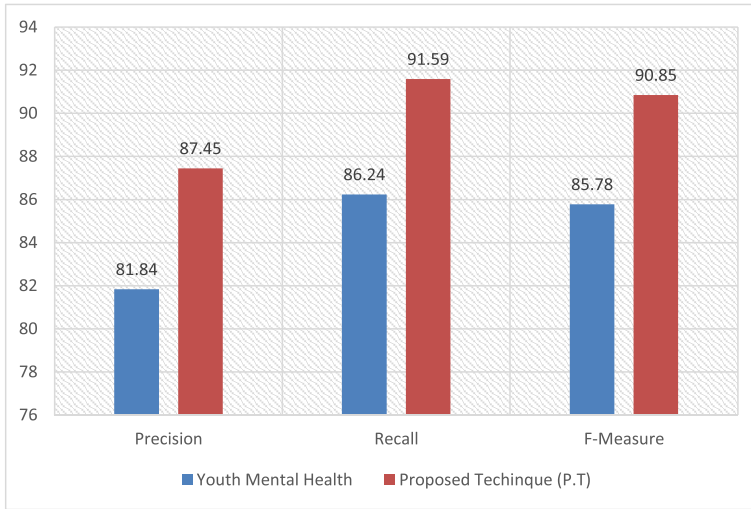


**Fig. 7** Proposed method results

**Fig. 8** Comparison of the proposed technique with Youth mental health

We observed from the details that the performance of the proposed approach, when compared with the actual published results of the proposed method, is up to the mark. The precision, recall and F-measure with almost 97%, 91% and 90% show the optimality of the proposed approach. Figure 8 shows the graphical representation of both results.

### 4.2 Evaluation of proposed technique with expert and survey

We have also had our proposed system evaluated by a group of professionals working at an NGO 'X' that provides healthcare solutions because no such study found in the past literature. Their expert group comprises two Medical Specialists, one Analyst and two Quality Affirmation Specialists. We have illustrated our proposed structure, created a knowledge base and its user interface. Our illustration covers the procedure of investigation, organization and administrations to their expert group. They have applied the rating running from 0 to 10. We have labelled rating like 0–3 as Not
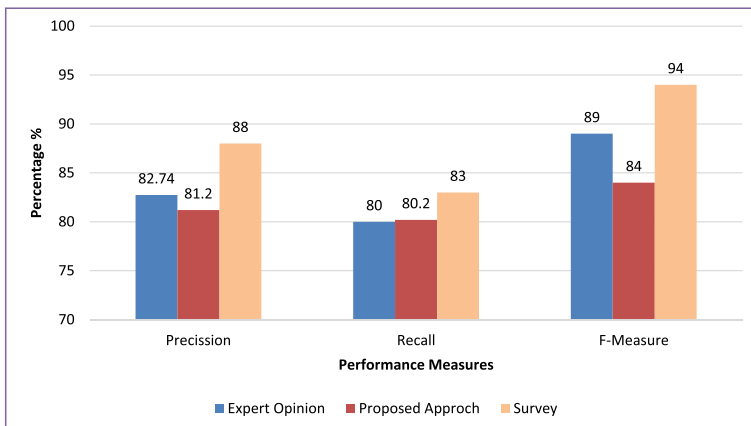


**Fig. 9** Comparison of proposed techniques with Expert Opinion and Surveys

Agreed, 4–6 as Partially Agreed and 7–10 as Agreed. We have condensed the outcomes dependent on agreed, partially agreed and Not Agreed by applying the mean on all the resultant estimations of every parameter. The criteria of expert opinions depended on three parameters: Agreed, Not Agreed and Partially Agreed. In light of the input from the specialists, they commonly concurred that the proposed methodology is suitable to deal with the finding of the impact of influencing factor. The proposed method for finding effect is easy to use and help in abridging comments and reviews (Fig. 9).

We compared the achieved findings and performed interviews with casually visiting patients in the hospital after calculating our anticipated technique. This process, which is weak in nature, applied to the persons who, are randomly visiting the hospitals. This interviewing procedure covered almost 30 hospitals containing both general and particular inquiries about COVID-19. In these meetings, we only pick appropriate information related to computing factors. There are distinctive choices reviewed from "poor" to "excellent". we compared the patient ranking with the proposed influence based model. The experimental results portray that the achieved results are optimistic when comparing with the recommendation of expert opinions. The performance measure of the proposed approach is encouraging in terms of accuracy. Figure 8 shows a Graphical representation.

## 5 Conclusion and future work

Microblogging sites and online forums are getting more attention in today's world, as it allows users to share opinions and views with other users. Peoples have a set of impact on social media sites in their daily life. These sites enhance the behaviour of peoples related to any opinion. Our paper presented a novel association rule-based method that used social media sites and forums to measure the impact of COVID-19 influencing factors on Teens.

When we compared the proposed method to some old state-of-the-art approaches, it produces exceptional outcomes when compared to some high-quality technique. In contrast with survey and official results, the proposed method predicts well. The obtained results are quite promising for the proposed method. In future, researchers can further enhance the proposed work in term of novel similarity measures.

# References

1. Ali AM Social Media and Its Impact on Arab Youth Identity. Review of European Studies 11(1). https://doi.org/10.5539/res.v11n1p1
2. Bozkurt A, Karadeniz A, Koçdar S (2017) Social networking sites as communication, interaction, and learning environments: perceptions and preferences of distance education students. Journal of Learning for Development 4(3):348–365
3. Brewer K (2020) Coronavirus: how to protect your mental health. URL: https://www.bbc.com/news/health-51873799
4. Cellan-Jones R (2020) Tech Tent: Is social media spreading the virus?. URL: https://www.bbc.com/news/technology-51510196
5. Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: a content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 5(11):e14118 [FREE Full text]. https://doi.org/10.1371/journal.pone.0014118 [Medline: 21124761]
6. Dai Y, Wang S, Chen X, Xu C, Guo W (2020) Generative adversarial networks based on Wasserstein distance for knowledge graph embeddings. Knowl-Based Syst 190:105165
7. Depoux A, Martin S, Karafillakis E, Bsd RP, Wilder-Smith A, Larson H (2020) The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 03:1–4. https://doi.org/10.1093/jtm/taaa031 [Medline: 32125413]
8. Devlin H (2020) Don't let coronavirus tip society into panic, say psychologists. URL: https://www.theguardian.com/science/2020/mar/06/dont-let-coronavirus-tip-society-into-panic-say-psychologists
9. Dillon C (2020) Coronavirus: the psychology behind panic-buying. URL: https://tinyurl.com/y9orfutq
10. Ejegwa PA (2020) Modified Zhang and Xu's distance measure for Pythagorean fuzzy sets and its application to pattern recognition problems. Neural Comput & Applic 32(14):10199–10208
11. El-Terk N (2020) Toilet paper, canned food: what explains coronavirus panic buying. URL: https://tinyurl.com/ww23bb3
12. Emmott R (2020) Russia deploying coronavirus disinformation to sow panic in West, EU document says. . URL: https://tinyurl.com/yx42jzyh
13. Frenkel S, Alba D, Zhong R (2020) Surge of Virus Misinformation Stumps Facebook Twitter. URL: https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html
14. Garrett O (2020) A case of modern mass hysteria? The coronavirus. URL: https://tinyurl.com/yarx6j7r
15. Gesser-Edelsburg A, Diamant A, Hijazi R, Mesch GS (2018) Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. PLoS One 13(12):e0209505 [FREE Full text]. https://doi.org/10.1371/journal.pone.0209505 [Medline: 30566485]
16. Gold H (2020) Inside the WHO's fight to stop false information about coronavirus from spreading. URL: https://www.wktv.com/content/news/568519502.html
17. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J et al (2017) Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages. JMIR Public Health Surveill 3(1):e14 [FREE Full text]. https://doi.org/10.2196/publichealth.6313 [Medline: 28336503]
18. Hao K, Basu T (2020) The coronavirus is the first true social-media infodemic. URL: https://www.technologyreview.com/s/615184/the-coronavirus-is-the-first-true-social-media-infodemic/
19. Hasnain H, Nasreen A, Ijaz H (2015) Impact of Social Media Usage on Academic Performance of University Students. In: 2nd International Research Management & Innovation Conference (irmic 2015) Langkawi, 26–27 August 2015
20. Hernandez S (2020) Panic spreads fear especially on social media, a local doctor says. URL: https://tinyurl.com/y736nqo7. Accessed 2020-03-20
21. JMIR E-collection (n.d.) Infoveillance, Infodemiology and Digital Disease Surveillance. URL: https://publichealth.jmir.org themes/304
22. Ke L, Starkey H (2014) Active citizens, good citizens, and insouciant bystanders: the educational implications of Chinese university students' civic participation via social networking. Lond Rev Educ 12(1):50–62
23. Kent S (2020) Fear spreads like a virus: How coronavirus panic is taking a toll on NJ's mental health. URL: https://tinyurl.com/ybwxk6o3. Accessed 2020-03-08
24. La V, Pham T, Ho M, Nguyen MP, . Nguyen K, Vuong T, et al. Policy Response, Social Media and Science Journalism for the Sustainability of the Public Health System Amid the COVID-19 Outbreak: The Vietnam Lessons. Sustainability 2020 ;12(7):2931. [doi:https://doi.org/10.3390/su12072931]
25. Lee AR, Suzanne HJ (2017) The role of social media on positive youth development: an analysis of 4-H Facebook page and 4-H'ers' positive development. Child Youth Serv Rev 77:127–138

26. Liang L, Ren H, Cao R, Hu Y, Qin Z, Li C, Mei S (2020) The effect of COVID-19 on youth mental health. Psychiatry Q 91:1–12
27. Mahmood QK, Saud M (2018) Civic Participation and Facebook: A Cross-Sectional Survey of Pakistani University Students. In: 2018 Presented at 4th international conference on contemporary social and political affairs (ICoCSPA) 2018; 13 August; Surabaya, pp 59–65
28. Merchant RM, Lurie N (2020) Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 23:29–31. https://doi.org/10.1001/jama.2020.4469 [Medline: 32202611]
29. Mian A, Khan S (2020) Coronavirus: the spread of misinformation. BMC Med 18(1):89–82 [FREE Full text]. https://doi.org/10.1186/s12916-020-01556-3 [Medline: 32188445]
30. Molla R (2020) How coronavirus took over social media. URL: https://tinyurl.com/ycwtmx3u
31. Muwahed J (n.d.) Coronavirus pandemic goes viral in the age of social media, sparking anxiety. URL: https://tinyurl.com/ybnms2se
32. Petric D (2020) Negative mental health effects of COVID-19 pandemic panic. URL: https://tinyurl.com/y8mqkpde
33. Raghavendra P, Hutchinson C, Grace E, Wood D, Newman L (2018) I like talking to people on the computer, Outcomes of a home-based intervention to develop social media skills in Youth with disabilities living in rural communities. Res Dev Disabil 76:110–123
34. Roblyer MD, McDaniel M, Webb W, Herman J, Witty JV (2010) Findings on Facebook in higher education: a comparison of college faculty and student uses and perceptions of social networking sites. Elsevier. Internet High Educ 13:34–140
35. Rothschild N (2020) Coronavirus panic sells as alarmist information flies on social media. URL: https://www.axios.com/coronavirus-social-media-b56326b6-ab16-4c8a-bc86-e29b06e5ab2b.html
36. Saud M (2015) Social Networking Sites and Deviance among Youth in Islamabad, Pakistan. EJBS:48–58. https://doi.org/10.33422/ejbs.2018.05.85
37. Saud M, Mashud M, Ida R. Persistence of Social Media on Political Activism and Engagement among Indonesian & Pakistani youths. IJWBC 2020;16(1):1 [FREE Full text] [doi: https://doi.org/10.1504/ijwbc.2020.10028448]
38. Shimizu K (2020 Feb) 2019-nCoV, fake news, and racism. Lancet 395(10225):685–686. https://doi.org/10.1016/s0140-6736(20)30357-3
39. The Star (2020) Coronavirus and how social media firms are combating the spread of misinformation about the disease. URL: https://tinyurl.com/y85jagfn
40. Victor D (2020) Panic and Criticism Spread on Chinese Social Media Over Coronavirus. URL: https://www.nytimes.com/2020/01/24/world/asia/china-social-media-coronavirus.html
41. Wongkoblap A, Vadillo MA, Curcin V (2017) Researching Mental Health Disorders in the Era of Social Media: Systematic Review. J Med Internet Res 19(6):e228 [FREE Full text]. https://doi.org/10.2196/jmir.7215 [Medline: 28663166]
42. Yang B, He Y, Long W (2016) Alienation of civic engagement in China? Case studies on social governance in Hangzhou. Volunt Int J Volunt Nonprofit Org 27(5):2150–2172
43. Yao Y (2020) Three-way granular computing, rough sets, and formal concept analysis. Int J Approx Reason 116:106–125
44. Yin Y, Fung A (2017) Youth online cultural participation and Bilibili: an alternative form of democracy in China? Digital Media Integration for Participatory Democracy:130–154