Fuzzy Optimization and Decision Making (2023) 22:669–696
https://doi.org/10.1007/s10700-023-09407-5

Analyzing society anti-vaccination attitudes towards COVID-19: combining latent dirichlet allocation and fuzzy association rule mining with a fuzzy cognitive map

Nazmiye Eligüzel[1]

Accepted: 2 January 2023 / Published online: 17 January 2023

Abstract

COVID-19 has been declared a pandemic and countries are tackling this disease either through preventative measures such as lockdown and sanitization or through curative ones such as medication, isolation, and so on. Some people believe that vaccination is the best way to prevent this disease, while others disagree. Society's attitudes toward vaccination can be influenced by a variety of factors such as misunderstanding, ambiguity, lack of knowledge. The proposed study's goal is to better understand people's attitudes regarding vaccination by focusing on key topics related to COVID-19 anti-vaccine tweets. Tweets are obtained over a period based on the number of COVID-19 cases by utilizing the "anti-vaccine" keyword rather than the "vaccine" keyword. Furthermore, in addition to people perceptions and attitudes toward anti-vaccination, the causal relationship between each topic is investigated. As a result, latent dirichlet allocation (LDA), fuzzy association rule mining (FARM), fuzzy cognitive map (FCM), and fuzzy c-means are used to conduct a complete study. Topics are analyzed independently using clustering and scenario analysis. The findings demonstrate the most common topics in anti-vaccination tweets, as well as the influence of each topic on the others.

Keywords Anti-vaccination · COVID-19 · Fuzzy cognitive map · Fuzzy association rule · Fuzzy c-means · Scenario analysis

✉ Nazmiye Eligüzel
nazmiye.eliguzel@gibtu.edu.tr

[1] Industrial Engineering, Gaziantep Islam Science and Technology University, 27010 Gaziantep, Turkey

1 Introduction

COVID-19 disease was officially declared a pandemic by the World Health Organization in March 2020[1]. The COVID-19 pandemic caused massive harm on to lives, and affected many nations' health and economies. Vaccination, in combination with sanitary and behavioral controls, is the most effective technique for reducing or eradicating viral infection and dissemination (Pogue et al., 2020). Some people have a favorable attitude toward the vaccination campaign, while others lack awareness of the vaccination campaign and its benefits. Therefore, it is known that people have different opinions and perspectives about the vaccination. Some studies have highlighted potential COVID-19 vaccination obstacles, such as questioning the need for vaccines and choosing to benefit from the immunity conferred by COVID-19 survivors (Liu & Liu, 2021a). In comparison to surveys, social media, especially Twitter, can acquire timely information about COVID-19 vaccination behavioral intentions and determine public attitudes towards anti-vaccination. The world's top pharmaceutical companies are racing to develop vaccines. As a result, the arrival of vaccinations and people's opinions is a fascinating research topic (Rahul et al., 2021). So, analyzing people's opinions and perspectives toward vaccination via Twitter is one of the significant research areas.

In the proposed study, tweets have been gathered in certain time periods according to the number of COVID-19 cases by using the "anti-vaccine" keyword rather than the "vaccine" keyword. Furthermore, besides the opinions and attitudes of the public about anti-vaccination, causal relationship between each topic is examined. Therefore, a detailed analysis is conducted by using latent dirichlet allocation (LDA), fuzzy association rule mining (FARM), fuzzy cognitive map (FCM), and fuzzy c-means. With the help of clustering and scenario analysis, topics are evaluated separately. To the best of the author's knowledge, there is no other study in the literature that combines these methodologies and addresses the public's attitude toward vaccination in a pandemic in this way.

The rest of this paper is organized as follows: Sect. 2 demonstrates the literature review that is divided into two parts: vaccine-related studies during the coronavirus pandemic and studies that applied FCM to text mining. Section 3 presents the preliminaries of the study. Section 4 presents the research methodology. Section 5 represents the results and discussions. Finally, the paper is concluded in Sect. 6.

2 Literature review

COVID-19 has recently emerged as one of the most pressing public health challenges, and along with the COVID-19 disease, vaccination has also become a hot topic. In terms of research content and applicable models, the literature is classified into two categories.

[1] https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

2.1 Analysis of public opinion against vaccination on twitter

In this part of the literature, some of the studies that analyze public opinion against vaccination by using Twitter are considered.

Rahul et al. (2021) proposed a study that analyzed COVID-19 vaccine related tweets to generate a report. They used the latest statistical topic modeling and LDA to identify popular topics. Sentiment analysis was also utilized in the paper. Sentiment analysis was implemented by using two different models, namely, Valence Aware Dictionary and Sentiment Reasoner (VADER) and TextBlob, and a comparison was made between them. Furthermore, after the application of topic modeling, seven dominant topics were extracted. The limitation of the study is that it does not extract search keywords during the analyses, such as COVID, vaccine, and coronavirus. Liew et al. (2021) proposed a study that aimed to utilize Twitter data to understand in close-to-real time public sentiments and perspectives about COVID-19 vaccines. They want to know about the major concerns that have captured the public's attention, as well as the obstacles and enablers to successful COVID-19 vaccination. An unsupervised machine learning approach, which is structural topic modeling, was utilized to determine topics. Furthermore, by using VADER, the rule-based machine learning model was applied to conduct sentiment analysis. The limitation of the study is that it uses Twitter data that were posted only in English and from a specific geographic region. Liu et al. (2021) developed a system for automatically analyzing public perceptions of COVID-19 vaccines using real-time data from social media, which can be used to modify educational programs and other interventions to increase the public acceptability of COVID-19 vaccines. For this purpose, a leveraging transfer learning model was developed, followed by temporal analysis and topic modeling. One of the disadvantages of the study is that the Twitter users do not represent the entire public. Liu and Liu (2021b) proposed a study that aimed to find thematic and temporal trends in COVID-19 vaccine-related tweets, as well as to investigate variances in sentiment at the worldwide, national, and state levels in the United States. English-language COVID-19 vaccine related tweets were collected and the VADER tool was applied to determine the sentiments of tweets such as positive, neutral, and negative. Although the VADER sentiment score was able to reliably discern the sentiment in the text, it was unable to determine whether the sentiment was directed towards the COVID-19 vaccination or not. Roe et al. (2021) proposed a sentiment analysis and discovered that this sentiment-based approach was beneficial for determining levels of vaccine hesitancy in the general public and that it, in conjunction with the questionnaire, proposes solutions for addressing specific concerns and misinformation. The limitation of the study is that the sources (Twitter accounts) were manually classified as "personal", "accredited medical", "news", or "government/public health". However, given the vast dataset in the main study, this was impossible. Guntuku et al. (2021) utilized Twitter to obtain a random sample of vaccine-related tweets to determine the geographical and temporal variation in COVID-19 vaccination discourse. They inferred an insight that in the United States, Twitter discourse on COVID-19 vaccines differed greatly amongst communities and changed over time. They used LDA topic modeling.

The study's limitations include the fact that Twitter did not represent the general population in the United States, and tweets were provided as a random sample.

2.2 FCM applications in text mining

The choice of concepts in the FCM and the assessment of causal relationships between these concepts are the most important issues to tackle when creating an FCM. Recently, text mining techniques and topic extraction methods have started to be used in concept determination. In the second part of the literature, some of the studies that include FCM application in text mining are demonstrated.

Son et al. (2020) used scenario-based technology roadmaps to evaluate future uncertainties in the technology planning stage. They analyzed the textual big data to determine the casual relationships between factors that may influence future uncertainty. Therefore, the FCM technique was incorporated into the technology roadmap. Liang et al. (2020) proposed a study that examined the assessment of web celebrity shops by analyzing online reviews in depth. In addition, the competitive analysis was discussed, and suggestions for improvement were offered. For the topic extraction, LDA was utilized and attributes that customers care about were determined. Furthermore, long short term memory (LSTM) and linguistic term sets (PLTSs) were applied to portray the sentiments of customers towards various attributes. Lastly, interrelationships among attributes were investigated by using the FCM and association rule mining. Han et al. (2019) proposed a study to collect policy elements by using text mining and latent semantic analysis. An FCM was built to deduce the evolution of elements using a soft computing method, and a FARM technique and partial association test were utilized to identify the causal relationships and impact degrees between policy elements.

It is seen that there are some studies that consider COVID-19 vaccine-related tweets. In these studies, the sentiment of the public and vaccine hesitancy were taken into consideration through sentiment analysis or topic extraction. Some studies analyzed public attitudes towards vaccination according to demographic regions. The proposed study focuses on analyzing and evaluating the important topics corresponding to the COVID-19 anti-vaccine-related tweets.

In nutshell, the proposed study provides insights into the importance of tweets in shaping societal anti-vaccine attitudes and presents us with a data-driven analysis. The following are the five contributions of this paper: To begin, LDA is utilized to extract characteristics on public opinion on vaccination via Twitter, as opposed to standard item construction regarding public questionnaires or current models. Second, FARM is used to discover probable relationships between items that are collected through the LDA process. Third, FCM is built to evaluate concepts acquired by the FARM method, which can explain the interrelationships of various items. Fourth, Fuzzy c-means is used to determine which concept belongs to which cluster. Finally, scenario analysis is used to show effect of one concept on others. Furthermore, instead of using the "vaccine" keyword, the "anti-vaccine" term is used to collect tweets over a period of time dependent on the number of COVID-19 cases. As a result, the influence of a change in the number of cases on the number of tweets posted on the anti-vaccine hashtag has been noticed. This research is significant because it explores the influence of various topics in distinct clusters on one another.

3 Preliminaries

3.1 Latent Dirichlet allocation (LDA)

LDA was proposed by Blei et al., (2003). It is a generative probabilistic topic extraction machine learning method where each single topic is demonstrated by a distribution of words. In LDA, documents are demonstrated as random mixtures of latent topics. The probability of a corpus can be computed by using Eq. 1.

$$p(D|\alpha,\beta) = \prod_{d=1}^{M} \int P(\theta_d|\alpha)\left(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn}|\theta_d)p\left(w_{dn}|z_{dn},\beta\right)\right)d\theta_d \tag{1}$$

α α is the Dirichlet-previous concentration parameter of each document topic distribution, β is the corpus level parameter, θ_d is the document-level variable, z_{dn} is the topic assignment for w_{dn}, w_{dn} is the n^{th} word in the d^{th}document, N is the number of words in the document, M is the number of documents to analyze, and D is the corpus of collection M documents.

3.2 Fuzzy association rule mining (FARM)

Association rule mining (ARM) is used to find the potential relationships between items by utilizing "if-then" rules to specify the correlations, frequency laws, and association structures among items to form and select association rules. It uses three main measures such as support, confidence, and lift (Xu et al., 2019). Fuzzy set theory is a solution to the fuzzy boundary problem, and it is implemented in the area of ARM called "fuzzy association rule mining" (FARM). In the FARM, the dependent relations among items are demonstrated by association rules (Wu, 2020). The rules are obtained by utilizing the minimum support and minimum confidence. Association rules are the dependent relationships between items. Let's consider, $A_n \rightarrow A_l$ as the association rules between items $A_n$and A_l(n, l = 1, 2, … ., n, landn ≠ l).Let $D = d_1, d_{2,\ldots\ldots,} d_L = numberofdocuments$ indicates the text dataset. v_{jl} represents the impact of document d_j towards item A_l. v_{jn} represents the impact of document $d_j$towards item A_n. Based on the results of the literature (Xu et al., 2019), support, confidence, and lift of the rule $A_n \rightarrow A_l$can be calculated as follows:

$$supp\left(A_n \rightarrow A_l\right) = \frac{\sum_{j=1}^{L} v_{jn} \otimes v_{jl}}{L} \tag{2}$$

$$conf(A_n \rightarrow A_l) = \frac{\sum_{j=1}^{L} v_{jn} \otimes v_{jl}}{\sum_{j=1}^{L} v_{jn}} \tag{3}$$

$$lift(A_n \rightarrow A_l) = \frac{conf(A_n \rightarrow A_l)}{supp(A_n \rightarrow A_l)} \tag{4}$$

where L is total number of the documents. Support represents the possibility of two items appearing together in a dataset. Confidence refers to the normalized impact of an item on a rule as well as the credibility that one item influences another (Liang et al., 2020). The polarity of association rules cannot be revealed by support and confidence. Therefore, to determine the polarity, it is required to calculate lift. Lift is calculated as the ratio between confidence and support.

3.3 Fuzzy cognitive map (FCM)

A fuzzy cognitive map (FCM) is used as a tool to formalize understanding of conceptual and casual relationships (Dickerson & Kosko, 1994). In an FCM, the casual relationship between concepts is demonstrated by fuzzy weights with positive and negative signs (Hajek et al., 2017). Each C_j concept can take values in the unit interval [0, 1], commonly known as the "activation level" (Papageorgiou & Kontogianni, 2012). The activation level can indicate membership in a fuzzy set describing the relative abundance of linguistic measures such as low, average, and high (Dickerson & Kosko, 1994).

$A_i^{(t)}$ is the state vector that represents the values of the concept C_i in time t. Furthermore, the state of the whole FCM can be defined by the state vector $A^{(t)} = [A_1^{(t)}, \ldots .., A_n^{(t)}]$ .The value A_i of each concept $C_{i\ \mathrm{i}}$ in a moment $t + 1$ is calculated by adding the previous value of $A_i$in a previous moment t with the product of the value A_iAof the cause node C_j in a previous moment t and the value of the cause-effect link wij (Hajek et al., 2017). The value $A_i$for each concept $C_i$is computed as follows:

$$A_i^{(t+1)} = f\left(A_i^{(t)} + \sum_{\substack{j \neq i \\ j = 1}}^{N} A_j^{(t)} . w_{ji} \right) \tag{5}$$

where t denotes time and $A_i^{(t+1)}$is the value of concept C_i at time $t + 1$, $A_j^{(t)}$ is the concept $C_j$value at time t, $w_{ji}$is the weight that refers to the interconnection between concept $C_j$and concept C_i. In addition, f is an activation function. In the literature, causality is indicated by some nonlinear edge functions, such as the commonly used sigmoid function. In this paper, a sigmoid activation function is considered that gives values of concepts in the range [0, 1]. As a result of its

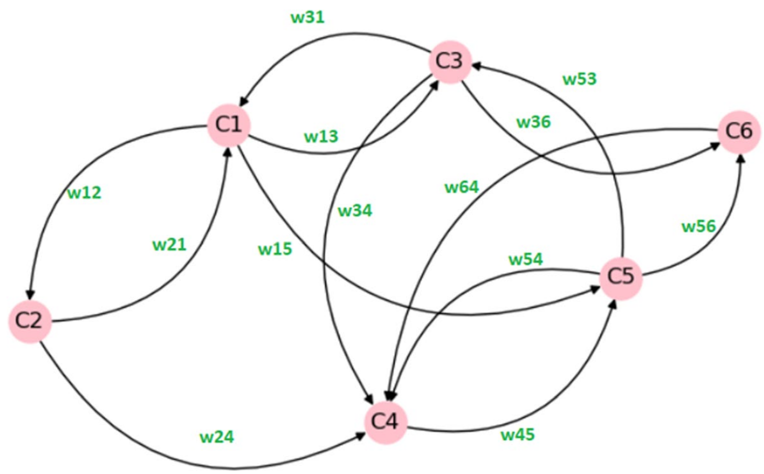

Fig. 1 Graphical representation of FCM

widespread application and positive outcomes (Kocabey Çiftçi & Unutmaz Durmuşoğlu, 2020), the sigmoid function is preferred. The mathematical formulation of the sigmoid function is given as follows:

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \tag{6}$$

where α is a real positive number that describes its steepness, and x is the value of $A_i^{(t)}$at the equilibrium point. The sigmoid activation function is utilized to constrain an unbounded weighted sum to a specific range, which limits quantitative analysis, but qualitative comparisons between concepts are allowed (Hajek et al., 2017). A graphical representation of an FCM is illustrated in Fig. 1. In Fig. 1, the representative FCM consists of six concepts (C1 to C6) and thirteen weights (cause and effect relationships between concepts-w_{ji}).

The weight describes the cause and effect relationships between two concepts, and it takes a value in the range of − 1 to 1 (Markinos et al., 2007). Three possible types of causal relationships are seen for weight. The first one is if w_{ji}>0, which represents positive causality between two concepts (C_j and C_i). The second one is if w_{ji}<0, which represents negative causality between two concepts (C_j and C_i), and the last one is if w_{ji}=0, which indicates no relationship between two concepts.

FCMs represent causal relationships. In the proposed study, association rules are obtained via FARM. Association rules are utilized input for FCM. That means, each association rule in the FCM describes the causal relationship between items. The effects of concepts on each other are investigated using the causal relationships between items.

3.4 Fuzzy C-means

Clustering is the task of assigning sets of elements to a number of groups. Clustering algorithms can be divided into two groups: the first is hard clustering, and the second is fuzzy (soft) clustering. In fuzzy clustering, elements in a dataset can belong to many clusters, and each element has a set of membership levels associated with it (Suganya & Shanthi, 2012). Fuzzy C-means is one of the most widely used fuzzy clustering algorithms. The fuzzy C-means uses fuzzy partitioning, which allows a data point to belong to any of the groups with membership grades ranging from 0 to 1 (Suganya & Shanthi, 2012). The algorithm is given as follows:

Algorithm 1 General process of the fuzzy c-means

1. ***Initialize*** $U = [u_{ij}]$ matrix, $U^{(0)}$
2. At t step: compute the centers vectors $C^{(t)} = [c_j]$ with $U^{(t)}$

$$C_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} xj}{\sum_{j=1}^{n} u_{ij}^{m}}$$

3. Update $U^{(t)}, U^{(t+1)}$
4.

$$d_{ij} = \sqrt{\sum_{i=1}^{n}(x_i - c_i)}$$

$$u_{ij} = \frac{1}{\sum_{t=1}^{c}(\frac{d_{ij}}{d_{tj}})^{2/(m-1)}}$$

5. İf $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$ ***then Stop; otherwise return to step 2.***

where '*m*' is a constant real number known as the fuzzifier that is greater than 1,u_{ij} is the membership degree of x_i in cluster j, x_i represents the ith of d-dimensional measured data, and c_j is referred to as the d-dimension center of the cluster. ϵ is the termination criteria between [0,1].

On the basis of the distance between the cluster center and the data point, this algorithm assigns membership to each data point corresponding to each cluster center.

4 Methodology

Application of LDA, FARM, Fuzzy C-means, and FCM together is one of the core ideas of the proposed paper to obtain public opinion about vaccination during the coronavirus pandemic. The other idea is to make a scenario analysis to show cause

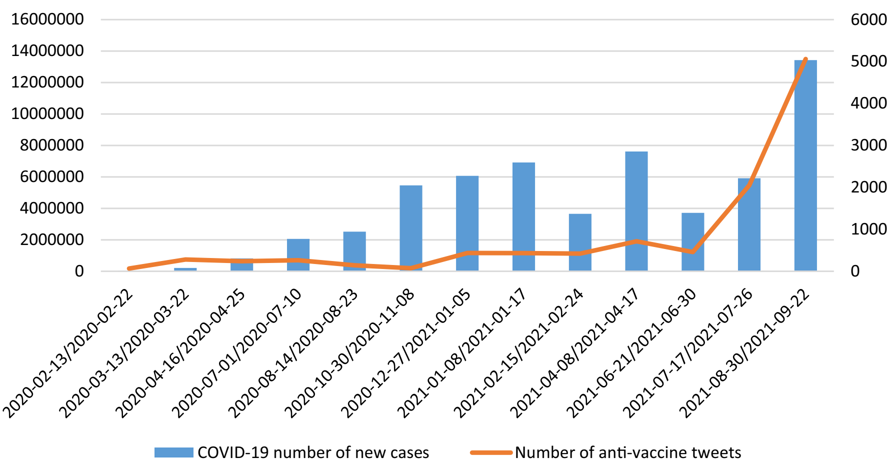

Fig. 2 The number of anti-vaccine tweets based on the number of COVID-19 cases

and effect relationships between proposed concepts. Depending on the COVID-19 case numbers, tweets are gathered within specific date intervals. A total of 10,670 tweets are gathered under "#anti-vaccine" and analyzed. That covers the pre-processing stage and topic extraction. In the pre-processing stage, punctuation and stop words are removed, and tokenization is applied, respectively. The output of this process has provided the input to the LDA process. In addition, some of the most frequently used words in the document, such as "covid", "covid-19", "vaccine", and "antivaccine" are removed in the pre-processing stage. Some of the tweets using the term "antivaccination" are likely to contain no mention of the COVID-19 vaccine. These tweets, however, are not deleted because they are the small part of the collected data-set and do not have significant impact on the results. The study's purpose is to figure out how society feels about COVID-19 vaccine in general. As a result, all tweets containing the hashtag "antivaccine" are grouped together, indicating how seriously people take this subject. According to the coronavirus cases in the world, tweets are collected by using the "#antivaccine" hashtag. The time periods considered in the proposed study are when the cases were at their highest and when the high case increases began to decline. As given in Fig. 2, at the beginning of the coronavirus pandemic, fewer anti-vaccine tweets are seen. Time intervals are taken in 10-day periods except for the last time period. It is possible to say that the number of tweets is related to the number of cases, but it's more likely that it is tied to the experience gained through time.

The collected tweet data belongs to a kind of unstructured data. After the pre-processing stage, it is required to make the data ready for analysis. Therefore, the LDA is used to translate textual data into structured data. Figure 3 illustrates the proposed tweet analysis, including sequential steps and techniques.

As shown in Fig. 3, the study considers tweets with the hashtag "antivaccine" using the Snscrape library in Python software. This study utilizes the LDA to put the FARM model to see the relations between keywords. FCM is considered with

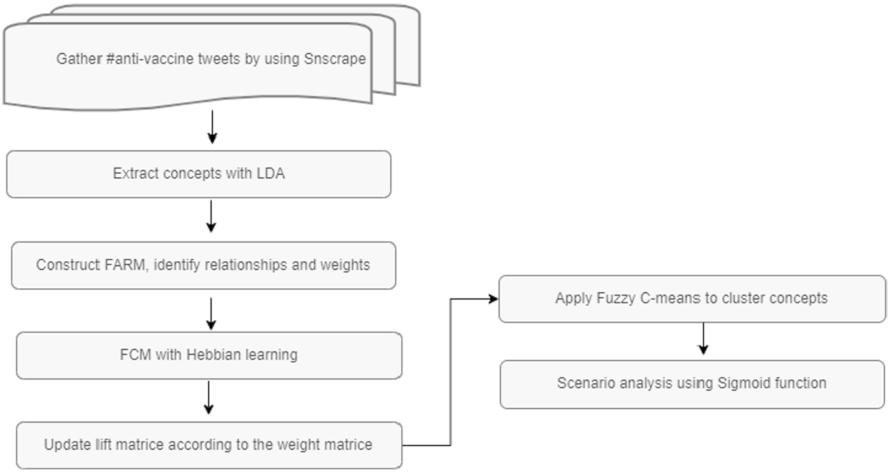

Fig. 3 Structure of the proposed analysis

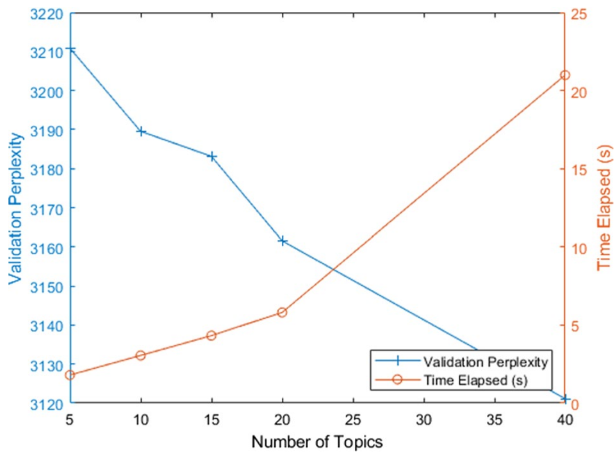

Fig. 4 The number of topics for LDA

Hebbian learning to update concept values and lift matrix. After the FCM operation, Fuzzy C-means is applied in order to cluster concepts. Finally, scenario analysis is used to understand the cause and effect relationships between concepts.

The LDA process is implemented using MATLAB software. According to Fig. 4, fitting the LDA model with 15–40 topics is a suitable choice. However, a better fit can be achieved by increasing the number of topics, but fitting the model takes longer. In order to reduce the processing time and gather results with sufficient perplexity, the number of topics is chosen as 15.

During the LDA analysis, all sentences are required to be divided into some words. In the process, document-to-topic probabilities are used. Each topic has its own related keywords. The document number is taken as the tweet number. Each topic is considered an FCM concept, and the probability of each topic in the document is taken into account as membership value. In Table 1, representative names of concepts, concept names, and their related keywords are given.

Concepts are described according to the term vector corresponding to each concept. The 15 topics are the 15 conceptual nodes of FCM, namely, the 15 attributes of the anti-vaccine tweets. After the determination of the concepts, the FARM process is applied, and the causal relationships between concepts are discovered. In order to obtain the weight matrix that is used to find interconnection between items, the general process of the FARM model is given as below:

Algorithm 2 General process of the FARM

Input: all items A_j $(j = 1, 2, \ldots\ldots\ldots, 15)$, document-item matrix

Output: weight matrix of association rules
    for all items $A_n \in [1, 15]$
        for all items $A_l \in [1, 15]$, $i \neq j$
            compute *support*($A_n \to A_l$)
            compute *confidence*($A_n \to A_l$)
            compute *lift* ($A_n \to A_l$)
        end
    end
for all association rules $A_n \to A_l$
    set the minsupp and minconf
    if association rule's confidence is smaller than the minconf ***then***
        remove this association rule
    end
    if association rule's support is smaller than the minsupp ***then***
        remove this association rule
    end
end
for all rest of the association rules $A_n \to A_l$
    if *lift* ($A_n \to A_l$) ≥ threshold lift value ***then***
        calculate the weight of association rule w_{nl}=*conf* ($A_n \to A_l$)
    else if *lift* ($A_n \to A_l$) < threshold lift value ***then***
            calculate the weight of association rule w_{nl}=−*conf* ($A_n \to A_l$)
    end
end

In the FCM application, Hebbian learning is used for concept learning. Hebbian learning is an unsupervised technique that was first used to train artificial neural networks (Haykin, 1999). The fundamental aspect of this learning rule is that the change of a synaptic is computed by taking into account the flow of presynaptic and postsynaptic signals towards each neural network processing unit, that is, a neuron (Papakostas et al., 2012). The reason behind using Hebbian learning drives from

Table 1 Concepts for FCM

Representative name of concepts	Concept name	Related keywords
C1	Anti-vaccine movement via celebrity	“anti-science”, “movement” “never”, “again”, “anything” “end” ,“left”, “political” “kanye-west” ,“called” “once”, “thought”, “stance” “follow”, “needs”
C2	Health care crisis	“protest” ,“passpo” “protests” ,“today”, “hospital” ,“protesters” “medical” ,“lockdown” “hospitals”, “rally” ,“outside”, “antilockdown” “care”, “doctors”, “full”
C3	Criticism on anti-vaccine idea	“know”, “like”, “why” “back”, “work”, “down”, “thats”, “let”, “take”, “idiots”, “children”, “shot” “taking”, “well”, “kids”
C4	Uprising	“people”, “like”, “really” “getting”, “shit”, “think” “fuck”, “yall”, “stupid”, “off” “mask”, “way”, “trying” “yet”, “keep”
C5	Against precaution-free behaviours	“people”, “why”, “against” “same”, “understand”, “still” “others”, “antimask”, “ones” “help”, “community”, “try” “views”, “fear”, “wonder”
C6	Conservative anti-vaccine people	“another”, “love”, “talk” “time”, “find”, “always”, “show” “talking”, “death” “conservative”, “put”, “call” “job”, “died”, “made”
C7	Social media fallacy	“need”, “stop”, “media” “social”, “health” “misinformation”, “public” “spread”, “passpos” “spreading”, “wrong”, “side” “thinking”, “hate” “government”
C8	Freedom on decisions	“pro”, “choice”, “life”, “right” “make”, “body”, “believe” “freedom”, “crowd”, “own” “lives”, “makes”, “care” “prolife”, “antimask”
C9	USA politic	“trump”, “news” “propaganda”, “right” “rhetoric”, “fox”, “republican” “literally”, “white”, “gop” “killing”, “americans” “republicans”, “deaths” “antimask”
C10	Falsified news in social media	“conspiracy”, “like”, “facebook” “time”, “every”, “post” “twitter”, “day”, “theories” “posts”, “stuff”, “posting” “seems”, “thing” ,“read”
C11	Misleading information for health	“amp”, “pay”, “antimask”, “themselves”, “health” ,“against”, “months” ,“new”, “fact”, “fake” “disinformation”, “clear” “found”, “guys”, “lies”
C12	Political decision in favor of provaccine	“mandate”, “everyone”, “point” “thing”, “mandates”, “group” “world”, “biden”, “provaccine” “healthcare”, “unvaccinated” “free”, “trust”, “around” “actually”
C13	Insultation	“die”, “person”, “fucking” “year”, “last”, “years”, “two” “lol”, “dead”, “told”, “old” “next”, “died”, “research” “morons”

Table 1 (continued)

Representative name of concepts	Concept name	Related keywords
C14	Anti-vaccine people	“people”, “many”, “say” “anyone”, “lot”, “dying”, “folks” “else”, “getting”, “today”, “first” “change”, “god”, “vote” “things”
C15	Make wearing mask significant	“know”, “mask”, “family”, “even” “please”, “hes”, “think”, “away” “bad”, “wear”, “friends”, “people” “give”, “ill”, “masks”

their popularity and fast convergence to desirable FCM states (Papakostas et al., 2012), Hebbian learning is chosen for training in the proposed study. That means the learning methods used by Hebbians are relatively quick. Their performance is determined on the initial weight matrix and FCM structure (Ren, 2012). Rather than using traditional approaches, a machine learning algorithm (LDA) is used to determine initial weights. In Eq. 7, the formulation of the Hebbian learning algorithm is given.

$$\Delta w_{ij} = \eta y_i(n)x_j \tag{7}$$

where η is the learning rate, which is taken as a positive constant, x_i and y_i are the signals of presynaptic and postsynaptic activation. Oja (1989) modified the above formulation to bring a solution to the stability problems. The generalized form of the Hebbian formulation is as follows:

$$\Delta w_{ij} = \eta y_i(n)(x_j(n) - y_i(n)w_{ij}(n)) \tag{8}$$

In the proposed study, η is taken as 0.5 by trial and error. After the 20 iterations, components of the weight matrix are stabilized.

5 Results and discussions

By comparing the support value of each association rule, a preliminary FARM model can be built with the association rules. First of all, it is required to set a minimum support value (minsupp). The user-specific threshold support value is referred to as the "minimal support" (Dave et al., 2014). Next, some association rules that are less than the minimum support value are removed. In this paper, the minimum support value is considered from different perspectives. The minimum support value is found by calculating the average of all components in the support matrix, ignoring the diagonal values. So, the minimum support value is set at 0.0041. Tables 2 and 3 show the obtained support matrix and the support matrix under the minimum support requirement. In the same way, the minimum confidence value (minconf) is calculated. After that, some association rules that are less than the minimum confidence value are removed. The minimum confidence value is found by calculating the average of all the components in the confidence matrix, ignoring the diagonal values. It is the user-specific threshold confidence value that is referred to as the "minimum confidence" (Dave et al., 2014). In Tables 4 and 5, the confidence matrix and the confidence matrix under the requirement of the minimum confidence are given. So, the minimum confidence value is set at 0.0613.

The degree of the lift value is the ratio of the confidence value to the support value, that is, the occurrence probability. Lift can be utilized to determine the polarity of weight. In this paper, the average of all the components in the lift matrix is considered the threshold lift value. Therefore, for each association rule $A_n \rightarrow A_l$, if lift value $(A_n \rightarrow A_l)$ = threshold lift value, then A_n and A_l are independent of each other. If lift value > threshold lift value, then A_n and A_l are positively correlated, that is, these two items appear together more frequently. In the same way, if lift

Table 2 Support matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1	0.05393	0.00328	0.00408	0.00470	0.00370	0.00317	0.00299	0.00394	0.00340	0.00353	0.00301	0.00347	0.00357	0.00361	0.00384
C2	0.00328	0.06167	0.00466	0.00537	0.00420	0.00360	0.00342	0.00450	0.00381	0.00401	0.00346	0.00398	0.00410	0.00411	0.00439
C3	0.00408	0.00466	0.07685	0.00677	0.00523	0.00450	0.00424	0.00564	0.00478	0.00502	0.00427	0.00493	0.00509	0.00515	0.00549
C4	0.00470	0.00537	0.00677	0.08881	0.00606	0.00521	0.00490	0.00652	0.00554	0.00582	0.00491	0.00570	0.00589	0.00593	0.00633
C5	0.00370	0.00420	0.00523	0.00606	0.06919	0.00405	0.00385	0.00510	0.00433	0.00451	0.00385	0.00446	0.00455	0.00465	0.00493
C6	0.00317	0.00360	0.00450	0.00521	0.00405	0.05946	0.00328	0.00433	0.00374	0.00389	0.00333	0.00382	0.00394	0.00398	0.00423
C7	0.00299	0.00342	0.00424	0.00490	0.00385	0.00328	0.05633	0.00411	0.00354	0.00368	0.00317	0.00360	0.00370	0.00376	0.00399
C8	0.00394	0.00450	0.00564	0.00652	0.00510	0.00433	0.00411	0.07451	0.00463	0.00483	0.00415	0.00478	0.00488	0.00498	0.00532
C9	0.00340	0.00381	0.00478	0.00554	0.00433	0.00374	0.00354	0.00463	0.06370	0.00415	0.00358	0.00412	0.00420	0.00428	0.00451
C10	0.00353	0.00401	0.00502	0.00582	0.00451	0.00389	0.00368	0.00483	0.00415	0.06635	0.00371	0.00425	0.00442	0.00446	0.00472
C11	0.00301	0.00346	0.00427	0.00491	0.00385	0.00333	0.00317	0.00415	0.00358	0.00371	0.05665	0.00364	0.00373	0.00381	0.00403
C12	0.00347	0.00398	0.00493	0.00570	0.00446	0.00382	0.00360	0.00478	0.00412	0.00425	0.00364	0.06517	0.00429	0.00437	0.00462
C13	0.00357	0.00410	0.00509	0.00589	0.00455	0.00394	0.00370	0.00488	0.00420	0.00442	0.00373	0.00429	0.06709	0.00449	0.00480
C14	0.00361	0.00411	0.00515	0.00593	0.00465	0.00398	0.00376	0.00498	0.00428	0.00446	0.00381	0.00437	0.00449	0.06800	0.00483
C15	0.00384	0.00439	0.00549	0.00633	0.00493	0.00423	0.00399	0.00532	0.00451	0.00472	0.00403	0.00462	0.00480	0.00483	0.07229

Table 3 Support matrix under the requirement of the minimum support

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1	0.05393	0	0.00408	0.00470	0	0	0	0	0	0	0	0	0	0	0
C2	0	0.06167	0.00466	0.00537	0.00420	0	0	0.00450	0	0	0.00346	0	0.00410	0.00411	0.00439
C3	0.00408	0.00466	0.07685	0.00677	0.00523	0.00450	0.00424	0.00564	0.00478	0.00502	0.00427	0.00493	0.00509	0.00515	0.00549
C4	0.00470	0.00537	0.00677	0.08881	0.00606	0.00521	0.00490	0.00652	0.00554	0.00582	0.00491	0.00570	0.00589	0.00593	0.00633
C5	0	0.00420	0.00523	0.00606	0.06919	0	0	0.00510	0.00433	0.00451	0.00385	0.00446	0.00455	0.00465	0.00493
C6	0	0	0.00450	0.00521	0	0.05946	0	0.00433	0	0	0.00333	0	0	0	0.00423
C7	0	0	0.00424	0.00490	0	0	0.05633	0.00411	0	0	0.00317	0	0	0	0
C8	0	0.00450	0.00564	0.00652	0.00510	0.00433	0.00411	0.07451	0.00463	0.00483	0.00415	0.00478	0.00488	0.00498	0.00532
C9	0	0	0.00478	0.00554	0.00433	0	0	0.00463	0.06370	0.00415	0.00358	0.00412	0.00420	0.00428	0.00451
C10	0	0	0.00502	0.00582	0.00451	0	0	0.00483	0.00415	0.06635	0.00371	0.00425	0.00442	0.00446	0.00472
C11	0	0	0.00427	0.00491	0	0	0	0.00415	0.00358	0	0.05665	0	0	0	0
C12	0	0	0.00493	0.00570	0.00446	0	0	0.00478	0.00412	0.00425	0.00364	0.06517	0.00429	0.00437	0.00462
C13	0	0.00410	0.00509	0.00589	0.00455	0	0	0.00488	0.00420	0.00442	0.00373	0.00429	0.06709	0.00449	0.00480
C14	0	0.00411	0.00515	0.00593	0.00465	0	0	0.00498	0.00428	0.00446	0.00381	0.00437	0.00449	0.06800	0.00483
C15	0	0.00439	0.00549	0.00633	0.00493	0.00423	0	0.00532	0.00451	0.00472	0.00403	0.00462	0.00480	0.00483	0.07229

Table 4 Confidence matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1	1	0.06074	0.07574	0.08724	0.06858	0.05884	0.05542	0.07303	0.06302	0.06546	0.05574	0.06428	0.06618	0.06701	0.07121
C2	0.05312	1	0.07556	0.08701	0.06818	0.05839	0.05542	0.07296	0.06180	0.06505	0.05605	0.06449	0.06647	0.06659	0.07118
C3	0.05315	0.06064	1	0.08805	0.06812	0.05856	0.05516	0.07345	0.06221	0.06533	0.05559	0.06418	0.06624	0.06702	0.07149
C4	0.05297	0.06042	0.07619	1	0.06826	0.05862	0.05521	0.07346	0.06237	0.06557	0.05531	0.06418	0.06637	0.06677	0.07132
C5	0.05345	0.06077	0.07566	0.08761	1	0.05850	0.05566	0.07365	0.06263	0.06516	0.05561	0.06442	0.06573	0.06718	0.07127
C6	0.05337	0.06056	0.07568	0.08756	0.06808	1	0.05518	0.07288	0.06287	0.06540	0.05604	0.06428	0.06633	0.06693	0.07117
C7	0.05305	0.06067	0.07525	0.08705	0.06837	0.05824	1	0.07300	0.06281	0.06538	0.05634	0.06399	0.06566	0.06679	0.07081
C8	0.05285	0.06038	0.07575	0.08756	0.06839	0.05816	0.05519	1	0.06220	0.06480	0.05565	0.06417	0.06548	0.06689	0.07139
C9	0.05335	0.05983	0.07506	0.08697	0.06803	0.05869	0.05555	0.07276	1	0.06509	0.05628	0.06475	0.06595	0.06725	0.07082
C10	0.05320	0.06046	0.07566	0.08776	0.06794	0.05861	0.05550	0.07277	0.06248	1	0.05590	0.06398	0.06665	0.06729	0.07115
C11	0.05306	0.06102	0.07542	0.08671	0.06793	0.05882	0.05603	0.07320	0.06328	0.06548	1	0.06422	0.06590	0.06718	0.07121
C12	0.05319	0.06102	0.07568	0.08745	0.06840	0.05865	0.05531	0.07337	0.06328	0.06514	0.05582	1	0.06590	0.06706	0.07094
C13	0.05320	0.06110	0.07587	0.08786	0.06779	0.05879	0.05513	0.07273	0.06261	0.06592	0.05564	0.06401	1	0.06697	0.07161
C14	0.05314	0.06039	0.07575	0.08721	0.06835	0.05852	0.05533	0.07329	0.06299	0.06566	0.05597	0.06427	0.06607	1	0.07110
C15	0.05312	0.06072	0.07599	0.08762	0.06821	0.05854	0.05518	0.07358	0.06240	0.06530	0.05580	0.06396	0.06646	0.06688	1

Table 5 Confidence matrix under the requirement of the minimum confidence

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1	1	0	0.07574	0.08724	0.06858	0	0	0.07303	0.06302	0.06546	0	0.06428	0.06618	0.06701	0.07121
C2	0	1	0.07556	0.08701	0.06818	0	0	0.07296	0.06180	0.06505	0	0.06449	0.06647	0.06659	0.07118
C3	0	0	1	0.08805	0.06812	0	0	0.07345	0.06221	0.06533	0	0.06418	0.06624	0.06702	0.07149
C4	0	0	0.07619	1	0.06826	0	0	0.07346	0.06237	0.06557	0	0.06418	0.06637	0.06677	0.07132
C5	0	0	0.07566	0.08761	1	0	0	0.07365	0.06263	0.06516	0	0.06442	0.06573	0.06718	0.07127
C6	0	0	0.07568	0.08756	0.06808	1	0	0.07288	0.06287	0.06540	0	0.06428	0.06633	0.06693	0.07117
C7	0	0	0.07525	0.08705	0.06837	0	1	0.07300	0.06281	0.06538	0	0.06399	0.06566	0.06679	0.07081
C8	0	0	0.07575	0.08756	0.06839	0	0	1	0.06220	0.06480	0	0.06417	0.06548	0.06689	0.07139
C9	0	0	0.07506	0.08697	0.06803	0	0	0.07276	1	0.06509	0	0.06475	0.06595	0.06725	0.07082
C10	0	0	0.07566	0.08776	0.06794	0	0	0.07277	0.06248	1	0	0.06398	0.06665	0.06729	0.07115
C11	0	0	0.07542	0.08671	0.06793	0	0	0.07320	0.06328	0.06548	1	0.06422	0.06590	0.06718	0.07121
C12	0	0	0.07568	0.08745	0.06840	0	0	0.07337	0.06328	0.06514	0	1	0.06590	0.06706	0.07094
C13	0	0	0.07587	0.08786	0.06779	0	0	0.07273	0.06261	0.06592	0	0.06401	1	0.06697	0.07161
C14	0	0	0.07575	0.08721	0.06835	0	0	0.07329	0.06299	0.06566	0	0.06427	0.06607	1	0.07110
C15	0	0	0.07599	0.08762	0.06821	0	0	0.07358	0.06240	0.06530	0	0.06396	0.06646	0.06688	1

value < threshold lift value, then A_n and A_l are negatively correlated, namely, these two items appear together less frequently. Table 6 demonstrates the lift matrix.

It is seen that there is no negative value in the lift matrix because of the scanning of tweets with specific hashtag on Twitter. However, the important point is the strength of the relationships among words. Therefore, the components in the lift matrix that are under the threshold value (average value) are considered infrequent words. The threshold lift value is set at 15.2458.

Based on the general FARM process, the weight matrix is given in Table 7.

After removing negative correlations (under threshold value) and zero components, the main rules are accessed. As shown in Tables 7 and 109 rules remained to be processed. Next, indegree, outdegree, degree centrality, and centrality concept values are calculated as seen in Table 8. In addition, Fig. 5 demonstrates the relationship between concepts by considering weights.

The degree centrality of the item A_n is calculated as follows (Liang et al., 2020):

$$\text{Degree centrality}\left(A_n\right) = \text{Outdegree}\left(A_n\right) + \text{Indegree}\left(A_n\right) \tag{9}$$

The indegree value refers to the sum of the weight of all association rules point to this item. The outdegree value refers to the sum of the weight of all association rules point out from this item. The centrality of concept values is defined as follows:

$$c_i = \frac{\text{Degree centrality}\left(A_n\right)}{\sum_{i=1}^{15}\text{Degree centrality}\left(A_i\right)} \tag{10}$$

The network includes 15 concepts and 109 edges with their weights that demonstrate the negative and positive effects of each concept on another. For example, there is no direct effect of C1 on C2, and no link is demonstrated between these two concepts. Another example related to the network is that C6 has a negative effect on C3. After the determination of concept values, FCM can be used to find updated concept values and a lift matrix to use for clustering. In Table 9, the updated weight matrix after the FCM implementation is given.

The updated weight matrix is taken as the distance matrix and used for Fuzzy C-means clustering. In addition, concept values are updated as follows:

C=(0.6231, 0.6148, 0.2963, 0.3378, 0.01401, 0.6197, 0.6214, 0.2238, 0.1486, 0.1637, 0.6214, 0.1473, 0.0136, 0.0134, 0.085).

Updated concept values demonstrate that “C1-Anti-vaccine movement via celebrity”, “C2-Health care crisis”, “C6-Conservative anti-vaccine people”, “C7-Social media fallacy“, and “C11-Misleading information for health” are the most powerful factors in the document. Following, it is said that “C3”, “C4”, and “C8” are powerful, but they are not as effective as the above.

After the application of Fuzzy C-means, obtained 3 clusters are demonstrated in Fig. 6.

The weight matrix that is obtained from the FCM application is used to aggregate clusters. As a result, the proposed study’s clustering approach allows for the derivation of clusters with strong concepts and which clusters are more effective than the

Table 6 Lift matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1	18.54256	18.51855	18.56455	18.56079	18.53541	18.56285	18.53464	18.53552	18.53549	18.54469	18.51881	18.52428	18.53817	18.56334	18.54455
C2	16.19368	16.21534	16.21523	16.20339	16.23443	16.22061	16.20502	16.21351	16.21963	16.22296	16.20041	16.20278	16.21263	16.20229	16.21391
C3	13.02702	13.01203	13.01236	13.00664	13.02421	13.01294	13.00951	13.02267	13.01475	13.01299	13.01909	13.01885	13.01312	13.01449	13.02121
C4	11.27006	11.25115	11.2547	11.25999	11.26334	11.25181	11.26761	11.26684	11.2588	11.26601	11.2638	11.25897	11.26853	11.26004	11.2674
C5	14.44622	14.46939	14.46581	14.45737	14.45296	14.44547	14.45752	14.44155	14.46334	14.44717	14.44527	14.44447	14.44614	14.44641	14.45644
C6	16.83457	16.82233	16.81791	16.80546	16.80883	16.81803	16.82234	16.83229	16.81033	16.8133	16.82768	16.82689	16.83607	16.81578	16.82538
C7	17.74349	17.7405	17.74823	17.76467	17.75811	17.75758	17.75253	17.76095	17.74383	17.76628	17.77393	17.77617	17.74522	17.76319	17.74753
C8	13.41469	13.41882	13.43121	13.42915	13.41028	13.43264	13.42725	13.42102	13.43417	13.41626	13.40984	13.42497	13.41894	13.4308	13.41824
C9	15.69264	15.70342	15.70243	15.69836	15.71117	15.69317	15.69222	15.71545	15.69859	15.685	15.7202	15.71522	15.70217	15.71157	15.70244
C10	15.0712	15.07716	15.07108	15.07887	15.06465	15.06689	15.08238	15.06551	15.05639	15.07159	15.06667	15.05349	15.0802	15.08664	15.07342
C11	17.62898	17.63605	17.66176	17.65913	17.64363	17.66366	17.67431	17.6385	17.67584	17.6483	17.65225	17.64364	17.66768	17.6338	17.67077
C12	15.32785	15.33172	15.35153	15.34296	15.3352	15.35275	15.36467	15.3489	15.35919	15.32671	15.33607	15.34448	15.36023	15.34628	15.35519
C13	14.90082	14.90246	14.90609	14.91698	14.89851	14.92197	14.89942	14.9034	14.90771	14.91495	14.91793	14.9211	14.90535	14.91453	14.91918
C14	14.72147	14.69376	14.70827	14.70637	14.69952	14.70465	14.71503	14.71706	14.71712	14.72175	14.69017	14.70816	14.71505	14.70588	14.72072
C15	13.83361	13.83146	13.84234	13.84245	13.83657	13.83969	13.82936	13.83053	13.83548	13.83575	13.84716	13.84313	13.84589	13.84691	13.83317

Table 7 Weight matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1			0.076	0.087											
C2			0.076	0.087	0.068			0.073					0.066	0.067	0.071
C3				− 0.088	− 0.068			− 0.073	− 0.062	− 0.065		− 0.064	− 0.066	− 0.067	− 0.071
C4			− 0.076		− 0.068			− 0.073	− 0.062	− 0.066		− 0.064	− 0.066	− 0.067	− 0.071
C5			− 0.076	− 0.088				− 0.074	− 0.063	− 0.065		− 0.064	− 0.066	− 0.067	− 0.071
C6			0.076	0.088				0.073							0.071
C7			0.075	0.087				0.073							
C8			− 0.076	− 0.088	− 0.068				− 0.062	− 0.065		− 0.064	− 0.065	− 0.067	− 0.071
C9			0.075	0.087	0.068			0.073		0.065		0.065	0.066	0.067	0.071
C10			− 0.076	− 0.088	− 0.068			− 0.073	− 0.062			− 0.064	− 0.067	− 0.067	− 0.071
C11			0.075	0.087				0.073							
C12			0.076	0.087	0.068			0.073	0.063	0.065			0.066	0.067	0.071
C13			− 0.076	− 0.088	− 0.068			− 0.073	− 0.063	− 0.066		− 0.064		− 0.067	− 0.072
C14			− 0.076	− 0.087	− 0.068			− 0.073	− 0.063	− 0.066		− 0.064	− 0.066		− 0.071
C15			− 0.076	− 0.088	− 0.068			− 0.074	− 0.062	− 0.065		− 0.064	− 0.066	− 0.067	

Table 8 Indices of the FCM

Concept	Indegree	Outdegree	Degree Centrality	Centrality of Concept Values (c)
C1	0.000	0.163	0.163	0.011
C2	0.000	0.508	0.508	0.033
C3	1.061	0.624	1.685	0.109
C4	1.225	0.613	1.838	0.118
C5	0.680	0.634	1.314	0.085
C6	0.000	0.308	0.308	0.020
C7	0.000	0.235	0.235	0.015
C8	0.951	0.626	1.577	0.102
C9	0.562	0.637	1.199	0.077
C10	0.588	0.636	1.224	0.079
C11	0.000	0.235	0.235	0.015
C12	0.577	0.636	1.213	0.078
C13	0.660	0.637	1.297	0.084
C14	0.670	0.634	1.304	0.084
C15	0.782	0.630	1.412	0.091

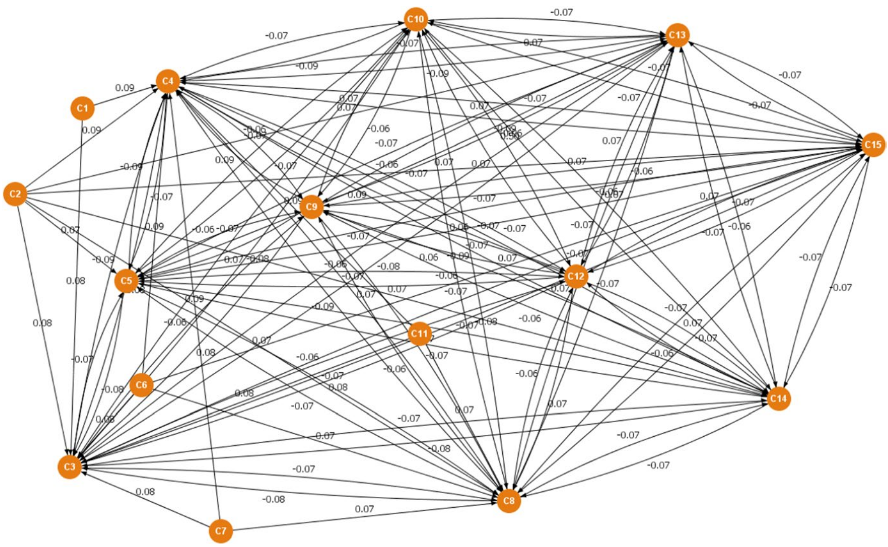

Fig. 5 The network of the proposed FCM

others. Clustering is additionally used to show related topics. As a result, the topics with the greatest overlap are looked at.

Most of the concepts in cluster 1 are the strongest, and cluster 2 comes after cluster 1 in that aspect. Lastly, the weakest concepts take their place in cluster 3.

Table 9 Updated weight matrix after the FCM application

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
C1	0.000	0.241	0.078	0.090	0.001	0.245	0.246	0.002	0.006	0.006	0.246	0.006	0.001	0.001	0.001
C2	0.241	0.000	0.078	0.090	0.069	0.239	0.240	0.075	0.006	0.006	0.240	0.006	0.067	0.067	0.072
C3	0.002	0.002	0.000	− 0.006	− 0.051	0.002	0.002	− 0.017	0.041	0.043	0.002	0.039	− 0.050	− 0.051	− 0.041
C4	0.002	0.003	0.006	0.000	− 0.049	0.003	0.002	− 0.009	0.056	0.059	0.002	0.055	− 0.048	− 0.048	− 0.036
C5	0.001	0.001	− 0.059	− 0.068	0.000	0.001	0.001	− 0.060	− 0.041	− 0.040	0.001	− 0.042	− 0.059	− 0.061	− 0.062
C6	0.245	0.239	0.078	0.090	0.001	0.000	0.243	0.075	0.006	0.006	0.243	0.006	0.001	0.001	0.072
C7	0.246	0.240	0.077	0.090	0.001	0.243	0.000	0.075	0.006	0.006	0.245	0.006	0.001	0.001	0.001
C8	0.002	0.002	− 0.020	− 0.023	− 0.054	0.002	0.002	0.000	0.018	0.020	0.002	0.016	− 0.052	− 0.053	− 0.047
C9	0.006	0.006	0.178	0.206	0.090	0.006	0.006	0.153	0.000	0.366	0.006	0.378	0.087	0.089	0.113
C10	0.006	0.006	0.032	0.037	− 0.043	0.006	0.006	0.012	0.239	0.000	0.006	0.237	− 0.042	− 0.043	− 0.026
C11	0.246	0.240	0.078	0.089	0.001	0.243	0.245	0.075	0.006	0.006	0.000	0.006	0.001	0.001	0.001
C12	0.006	0.006	0.179	0.206	0.091	0.006	0.006	0.154	0.377	0.367	0.006	0.000	0.087	0.089	0.113
C13	0.001	0.001	− 0.060	− 0.069	− 0.061	0.001	0.001	− 0.059	− 0.041	− 0.042	0.001	− 0.042	0.000	− 0.061	− 0.062
C14	0.001	0.001	− 0.059	− 0.068	− 0.062	0.001	0.001	− 0.060	− 0.041	− 0.041	0.001	− 0.043	− 0.060	0.000	− 0.062
C15	0.001	0.001	− 0.046	− 0.053	− 0.059	0.001	0.001	− 0.049	− 0.021	− 0.020	0.001	− 0.022	− 0.057	− 0.058	0.000

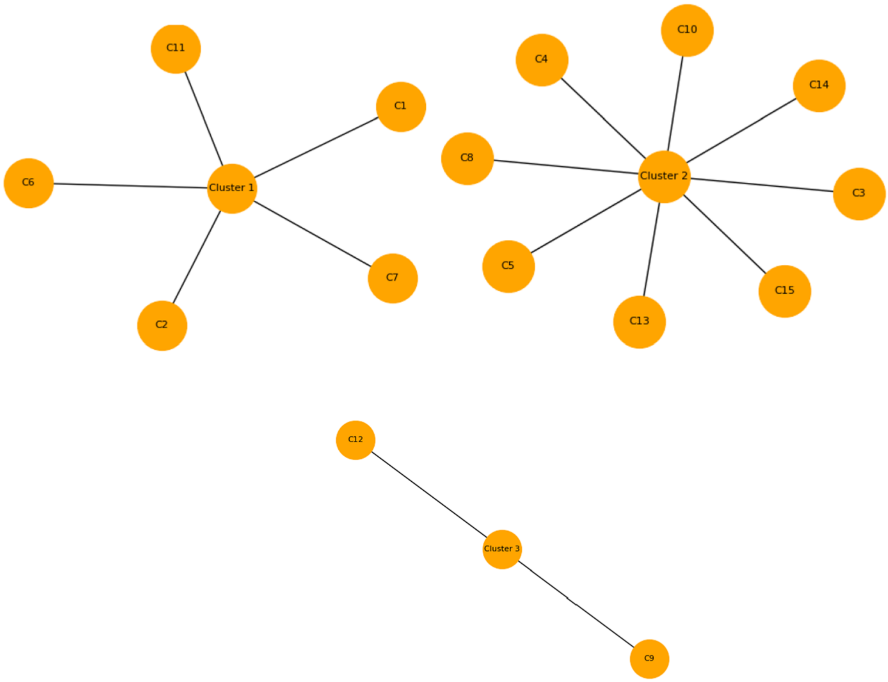

Fig. 6 Representation of the clusters after the fuzzy c-means

Therefore, cluster 1 is the most effective cluster, which is followed by clusters 2 and 3 in order.Cluster 1 includes concepts such as "Anti-vaccine movement via celebirity", "Healthcare Crisis", "Conservative anti-vaccine people", "Social media fallacy', and "Misleading information for health". Therefore, cluster 1 indicates the problems related to issues arising from the pandemic. Namely, the effects of the pandemic on the health system, society, and misleading information related to the pandemic.

Cluster 2 contains the concepts "Criticism on anti-vaccine idea," "Uprising", "Against precaution-free behavior", "Freedom of decision", "Falsified news in social media", "Insultation", "Anti-vaccine people", and "Make wearing mask significant". When it comes to cluster 2, it can be said that tweets mostly related to the reaction of society against emerging issues are included.

Cluster 3 covers the concepts of "USA politic" and "Political decision in favor of provaccine". In cluster 3, tweets with the focus of politics are observed.

5.1 Scenario analyses

For the scenario analysis, the updated concept values are considered. The sigmoid function is used to demonstrate the equilibrium point for concept values. By considering Eq. 6, α, namely, steepness is considered to be between 0.01 and 0.9.

In the proposed study, six scenarios are analyzed and illustrated.

5.2 Scenario 1

An increase of the strongest concept of C1, that is, “Anti-vaccine movement via celebrity”, causes to an increase on the “Health care crisis”, “Criticism on anti-vaccine idea” and “Uprising”, “Conservative anti-vaccine people”, “Social media fallacy”, and “Misleading information for health“. Namely, as the anti-vaccine movement among celebrities rises, so does public protest and criticism.

5.3 Scenario 2

C2 is one of the most powerful concepts, that is, “Health care crisis “. An increase in this concept leads to positive changes in many other concepts including “Anti-vaccine movement via celebirity “, “Criticism on anti-vaccine idea”, “Uprising”, “Against precaution-free behaviours”, “Conservative anti-vaccine people”, “Social media fallacy”, “Freedom on decisions”, “Misleading information for health “, and “Make wearing mask significant”.

5.4 Scenario 3

If the concept of C3, that is, “Criticism on anti-vaccine idea” decreases, many concepts are affected both positively and negatively. That means concepts like “Against precaution-free behaviours”, “Insultation”, and “ Anti-vaccine people are increased by the decrease of C3. Furthermore, as the C3 value decreases, the concepts of “USA politics,“ “Falsified news in social media,“ and “Political decision in favor of provaccine” decrease.

5.5 Scenario 4

Concept C6 refers to “Conservative anti-vaccine people”. With the increase of this concept, many other concepts such as “Anti-vaccine movement via celebirity”, “Health care crisis”, “Criticism on anti-vaccine idea”, “Uprising”, “Social media fallacy”, “Freedom on decisions”, “Misleading information for health”, and “Make wearing mask significant” are increased.

5.6 Scenario 5

The other strongest concept of C7, which is “Social media fallacy”, is the important factor in order to manipulate of people idea. It can cause provocation. When this concept value is increased, “Anti-vaccine movement via celebirity “, “Health care crisis”, “Criticism on anti-vaccine idea”, “Uprising”, “Conservative anti-vaccine people”, “Freedom on decisions”, and “Misleading information for health”

Table 10 The results of the scenario analyses

Concepts	No change	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario5	Scenario 6
C1	0.62317	**0.98**	0.64317	0.62317	0.64317	0.64317	0.61317
C2	0.61483	0.63483	**0.98**	0.61483	0.63483	0.63483	0.60483
C3	0.29638	0.30638	0.30638	**0.1**	0.30638	0.30638	0.29638
C4	0.33781	0.34781	0.34781	0.33781	0.34781	0.34781	0.33781
C5	0.01401	0.01401	0.02401	0.02401	0.01401	0.01401	0.01401
C6	0.61973	0.63973	0.63973	0.61973	**0.98**	0.63973	0.60973
C7	0.62145	0.64145	0.64145	0.62145	0.64145	**0.98**	0.61145
C8	0.22386	0.22386	0.23286	0.22386	0.23286	0.23286	0.22386
C9	0.14863	0.14863	0.14863	0.13863	0.14863	0.14863	0.14863
C10	0.16374	0.16374	0.16374	0.15374	0.16374	0.16374	0.16374
C11	0.62145	0.64145	0.64145	0.62145	0.64145	0.64145	**0.52**
C12	0.14739	0.14739	0.14739	0.13739	0.14739	0.14739	0.14739
C13	0.01366	0.01366	0.01366	0.02366	0.01366	0.01366	0.01366
C14	0.01342	0.01342	0.01342	0.02342	0.01342	0.01342	0.01342
C15	0.08512	0.08512	0.09512	0.08512	0.09512	0.08512	0.08512

increases. In the same way, these factors are negatively affected when the "Social media fallacy" is decreased.

5.7 Scenario 6

When the concept of C11 "Misleading information for health" is taken into consideration, a decreases in this concept causes decreases in the "Anti-vaccine movement via celebirity", "Health care crisis", "Conservative anti-vaccine people", and "Social media fallacy".

The results of the scenario analysis are provided in Table 10. In addition, the changed values of concepts are indicated in bold font for each scenario.

In Table 10, the summary of the scenario analysis is given. From Table 10, it can be concluded that changes in C9, C10, C12, C13, and C14 depend on only changes in C3. In addition, the change in C15 is driven by the change in the concepts of C2 and C4. In a nutshell, the strongest concepts have a widespread effect on the rest of the concepts.

6 Conclusion

In this study, anti-vaccine related tweets are deeply analyzed. Based on the analyzed data, the attributes (concepts) are extracted by using the LDA. Relationships between the concepts are investigated by constructing the FCM under the guidelines of the association rule mining. Concepts are clustered via fuzzy c-means in order to provide a new perspective on them. Furthermore, a scenario analysis is conducted to show the effects of each concept on the other. This method can provide a new viewpoint for decision-makers to understand reason behind the attitude of anti-vaccination and some of the factors that influence the concepts of anti-vaccine related tweets. Finally, the following are the implications of the proposed study:

- Cluster 1 represents the concerns resulting from the pandemic. Specifically, the pandemic's consequences on the healthcare system, society, and false information about the pandemic.
- Regarding cluster 2, it can be claimed that tweets mostly relating to society's response to emerging challenges are included.
- Politics-related tweets can be seen in cluster 3.
- The most strongest concept is "anti-vaccine movement via celebrity". Celebrity anti-vaccination activism is growing, and so are public protest and criticism.
- "Health Care Crisis" and "Social media fallacy" are the other strongest concepts.

Some limitations can be seen in the proposed study. First, tweets are gathered by searching "anti-vaccine" keyword and hashtag. However, it is possible that there are other tweets that do not contain this hashtag but are relevant to the topic. Second, tweets that were posted in English are only considered due to the difficulties in analyzing them in various languages together. Therefore, the results can be more typical of English-speaking populations.

For future studies, different languages other than English can be analyzed to have different perspectives on vaccination topics. In addition, heuristic approaches can be generated for large size datasets in order to construct an FCM model.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Dave, N., Potts, K., Dinh, V., & Asuncion, H. U. (2014). Combining association mining with topic modeling to discover more file relationships. *International Journal on Advances in Software*, *7*(3&4).

Dickerson, J. A., & Kosko, B. (1994). Virtual worlds as fuzzy cognitive maps. *Presence: Teleoperators & Virtual Environments*, *3*(2), 173–189.

Guntuku, S. C., Buttenheim, A. M., Sherman, G., & Merchant, R. M. (2021). Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. *Vaccine*, *39*(30), 4034–4038.

Hajek, P., Prochazka, O., & Pachura, P. (2017). Fuzzy cognitive maps based on text analysis for supporting strategic planning. In *International conference on research and innovation in information systems (ICRIIS)* (pp. 1–6).

Han, H., Wang, Q., & Chen, C. (2019). Policy text analysis based on text mining and fuzzy cognitive map. In *15th international conference on computational intelligence and security (CIS)* (pp. 142–146).

Kocabey Çiftçi, P., & Unutmaz Durmuşoğlu, Z. D. (2020). A multi-stage learning-based fuzzy cognitive maps for tobacco use. *Neural Computing and Applications*, *32*(18), 15101–15118.
Liang, D., Dai, Z., Wang, M., & Li, J. (2020). Web celebrity shop assessment and improvement based on online review with probabilistic linguistic term sets by using sentiment analysis and fuzzy cognitive map. *Fuzzy Optimization and Decision Making*, *19*(4), 561–586.
Liew, T. M., & Lee, C. S. (2021). Examining the utility of social media in COVID-19 vaccination: Unsupervised learning of 672,133 Twitter posts. *JMIR Public Health and Surveillance*, *7*(11).
Liu, S., Li, J., & Liu, J. (2021). Leveraging transfer learning to analyze opinions, attitudes, and behavioral intentions toward COVID-19 vaccines: Social media content and temporal analysis. *Journal of Medical Internet Research*, *23*(8).
Liu, S., & Liu, J. (2021a). Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. *Journal of Medical Internet Research*, *23*(5).
Liu, S., & Liu, J. (2021). Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine, 39*(39), 5499–5505.
Markinos, A., Papageorgiou, E., Stylios, C., & Gemtos, T. (2007). Introducing fuzzy cognitive maps for decision making in precision agriculture. *Precision Agriculture*, *7*, 223.
Oja, E. (1989). Neural networks, principal components and subspaces. *International Journal of Neural Systems*, 61–68.
Papageorgiou, E., & Kontogianni, A. (2012). Using fuzzy cognitive mapping in environmental decision making and management: a methodological primer and an application. *International Perspectives on Global Environmental Change*, 427–450.
Papakostas, G. A., Koulouriotis, D. E., Polydoros, A. S., & Tourassis, V. D. (2012). Towards hebbian learning of fuzzy cognitive maps in pattern classification problems. *Expert Systems with Applications*, *39*(12), 10620–10629.
Pogue, K., Jensen, J. L., Stancil, C. K., Ferguson, D. G., Hughes, S. J., Mello, E. J., Burgess, R., Berges, B. K., Quaye, A., & Poole, B. D. (2020). Influences on attitudes regarding potential COVID-19 vaccination in the United States. *Vaccines*, *8*(4), 582.
Rahul, K., Jindal, B. R., Singh, K., & Meel, P. (2021). Analysing public sentiments regarding COVID-19 vaccine on Twitter. In *7th international conference on advanced computing and communication systems (ICACCS)* (pp. 488–493).
Ren, Z. (2012). Learning fuzzy cognitive maps by a hybrid method using nonlinear hebbian learning and extended great deluge algorithm. In MAICS, 159–163.
Roe, C., Lowe, M., Williams, B., & Miller, C. (2021). Public perception of SARS-CoV-2 vaccinations on social media: Questionnaire and sentiment analysis. *International Journal of Environmental Research and Public Health, 18*(24), 13028.
Haykin, S. (1999). Neural networks: A comprehensive foundation. Pearson Education.
Son, C., Kim, J., & Kim, Y. (2020). Developing scenario-based technology roadmap in the big data era: An utilisation of fuzzy cognitive map and text mining techniques. *Technology Analysis & Strategic Management, 32*(3), 272–291.
Suganya, R., & Shanthi, R. (2012). Fuzzy c- means algorithm- a review. *International Journal of Scientific and Research Publications*, *2*(11).
Wu, S. (2020). A fuzzy association rules mining analysis of the influencing factors on the failure of oBike in Taiwan. *Mathematics, 8*(11), 1908.
Xu, X., Yin, X., & Chen, X. (2019). A large-group emergency risk decision method based on data mining of public attribute preferences. *Knowledge-Based Systems, 163*, 495–509.