MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: April 9, 2023
https://doi.org/10.31893/multiscience.2023021

Analyzing the misleading information on Covid-19 using MBCFWS4

A. Anbu[a]

[a]College of Management, SRM Institute of Science and Technology, India.

Abstract The COVID-19 pandemic had a major effect on almost every area of human lives in the majority of the world's countries. Misinformation spreads quickly in the initial phases of the COVID-19 pandemic in different forms, such as fear, distortion, contraindication assumption, and others. False and misleading advertisements harm millions of individuals. In recent research, there are more advanced techniques have been used to address the misinformation about the COVID-19 pandemic. But they are self-reported and probabilistic data during the lockdown period. So, it was difficult to find the respondents' perceptions based on sharing the COVID-19 misinformation. This proposed work analyzed and filtered some optimized factors to analyze the misinformation such as fake reports, fake remedies, conspiracy, susceptibility, vaccine rumors, social media, vitamin D prevents corona, political corona, socio-Economic, and more side effects after getting vaccinated. Collaborative filtering (CF) is the most efficient recommender system, and it is extensively utilized by a broad range of research institutions and enterprises, as well as being used in practice. It consists of two types of CF namely Memory-based CF and Model-based CF. In this work, Memory-based CF recommendation algorithm is combined with a similarity measure called MBCFWS4 to analyze the similarity measure between the factors to conclude the most impact factor. The Primary and secondary dataset helps to identify the respondents' perception based on the COVID-19 misinformation. The efficiency comparison of the proposed work is measured in terms of Precision, Recall & F-measure and found that this analysis using MBCFWS4 is outperforming well than the others as MBCFWS4 predicted accurately and revealed the conclusion based on the COVID-19 misinformation.

Keywords: corona virus, misinformation, memory-based CF, respondent's perception, similarity measure, ten factors

1. Introduction

The COVID-19 pandemic is a significant public health issue, as well as an impact on the world's financial sector. Disease mitigation measures implemented in several places result in increased unemployment, significant income reductions, and interruptions in the transportation, service, and production industries. The majority of governments across the country tend to have underestimated the dangers of the rapid COVID-19 spread and have been primarily reactive in their disaster response (Pak et al 2020). Nowadays, several websites have published information on COVID-19 and offered their users various instructions. In this regard, social media technologies are establishing a viable stage for timely interaction about health to specific users while simultaneously establishing a hotspot of disinformation generation, development, and transmission.

It is possible that individuals may be perplexed and will not make the correct option regarding COVID. Misinformation is often disseminated in a "florid" atmosphere overloaded with information overload, functional illiteracy, and confirmation bias. Misinformation will spread more quickly and widely, especially during a pandemic, since the public's response to pandemics and important social events appears to be extremely susceptible to error. The public typically gathers several types of knowledge about potential upcoming threats, creating the potential for disinformation to spread. The probabilistic decision always ends with disaster in the environment, and the right individual decision helps to secure health.

But, most people make the wrong decision in pandemic situations right from the sharing of misinformation. No analysis helps to make the right decisions about COVID and the causes of sharing misinformation dealing with various defects related to economically. (Roozenbeek et al 2020) Specific disinformation claims, when examined, are generally considered trustworthy by a significant portion of the population and constitute a serious public health concern. Importantly, this research shows a clear relationship, which has been repeated worldwide, between sensitivity to disinformation and vaccination reluctance and a lower probability of following public health advice. This study emphasized the essential part that scientists play as trusted and consistent information disseminators, as well as the possible relevance of strengthening capabilities in numeracy and logical reasoning to minimize exposure to misinformation.

 Received: December 16, 2023 | Accepted: March 2, 2023

Elhadad et al (2020) addressed the issue of identifying false information about COVID-19. We present a methodology for detecting false information that uses information from the UNICEF, World Health Organization, and the United Nations, as well as epidemiological data gathered from various internet sites for fact-checking. Collecting information from credible resources must ensure its accuracy. The acquired ground-truth value was utilized to develop a recognition mechanism that employs machine learning to recognize false information in this study. (Al-Zaman 2021) To get well knowledge of the misinformation issues of the pandemic in India, researchers underlined a new paradigm in the country's communication organization, a deficit of digital literacy, insufficient anti-misinformation efforts, and the political situation.

The previous analysis is based on self-reported data from self-selected individuals, and the lockdown time was a limitation in gathering more representative data. It was challenging to locate individuals who wanted to take part in their research. Among all the others, the most successful technique for analyzing the effects of COVID is factor analysis, but there are no real factors to measure the pandemic effects, and there are no right decisions taken under the factor analysis. Our proposed work studied and analyzed the real-time factors that are collected to measure the COVID pandemic causes. The contribution of this proposed work is to overcome the decision-based problems of COVID-19 and conclude the most influencing factor based on COVID-19 using MBCFWS4.

The proposed work readies the primary source data, i.e., a structured questionnaire that was used to determine the individual responses toward the effects of COVID-19 misinformation. The secondary source data is also used and collected online. And more importantly, the proposed work considers several factors related to the COVID pandemic to analyze people's perceptions and reveal the right decision. This work used the (Ejegwa 2020) proposed work called a new similarity measure which is used to combine in memory-based collaborative filtering technique as MBCFWS4 to determine the similarity between proposed factors. Figure 1 represents the workflow diagram.

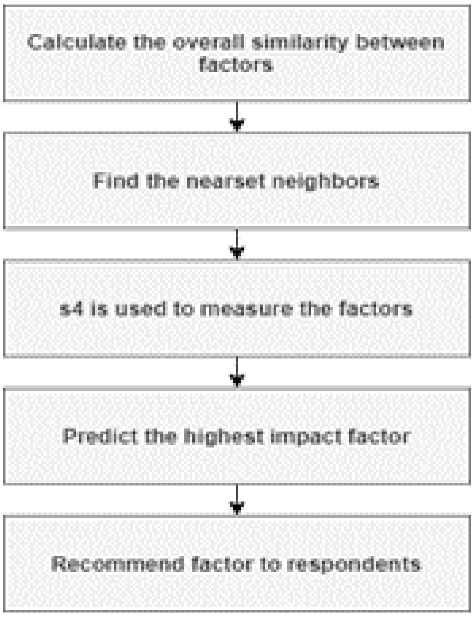

Figure 1 Flow diagram.

1.1. Literature survey

Peng (2018) presented a new Pythagorean Fuzzy distance and similarity metric. Some intriguing features of distance and similarity measurements were demonstrated. To demonstrate the availability of similarity metrics among (Pythagorean Fuzzy Set) PFS, several remarkable scenarios are provided.

Chen et al (2018) reviewed the conventional CF-based methods and approaches utilized in (Recommender System) RS and investigated several new hybrid CF-based recommendation systems and methods, comprising the most recent hybrid memory-based and model-based CF recommendation techniques. Finally, this workshop examined the possible influence that might enhance the RS as well as the future path. This paper proposed current hybrid CF-based recommendation approaches integrating social networks to tackle high dimensionality and data sparsity, as well as give a new point of view to increase RS performance.

Hatmal et al (2021) suggested that authorized COVID-19 vaccinations are safer, as well as being vaccinated makes individuals feel safer. Most of the adverse responses after vaccination are usually mild, representing that the innate immune system is boosting its resistance. In addition, machine learning techniques are developed to assess the intensity of side effects, and it was discovered that XGBoost, RF, and MLP offered excellent prediction accuracy in predicting the level of adverse effects depending upon the input data.

Roy and Ghosh (2020) processed multiple open-access datasets on US states to generate a unified database of probable pandemic spread variables. Then, use a variety of supervised machine learning methods to achieve an opinion and list the essential criteria. The analysis of regression is used to identify the important pre-lockdown variables that influence post-lockdown infection and death, influencing imminent lockdown policies.

Strength-opportunity (SO), weakness-threat (WT), strength-threat (ST), and weakness-opportunity (WO) methods for COVID-19 prevention and management were evaluated and developed by Wang and Wang (2020). This paper performed a thorough investigation and selected the most significant policies, and they are as follows: reconfiguring the emergency system (SO1), incorporating health emergency departments into universities and other institutions (WO2), modifying the financial structure and enhancing domestic and international connections (ST2), and improving public involvement in contributing to public health emergencies (WT1).

Almost 18 million tweets about coronavirus were gathered between March 1, 2020, and May 31, 2020, and were examined by Tariq Soomro et al (2020). This study examines the link concerning public attitude and the number of COVID-19 cases using VADER, a rule-based supervised machine learning model. This research also examines the frequency with which a nation is mentioned in tweets, as well as the increase in its daily sum of COVID-19 instances.

The macroeconomic impact of the COVID-19 pandemic on the UK and the government's response activities were assessed by comparing actual information and building regression models with STATA (Junlan and Chengke 2020). Before the pandemic, the British economy was in great condition. The emergence of COVID-19, on the other hand, has sent the British economy into a tailspin. The British government and the Bank of England have devised a lot of techniques to mitigate the pandemic's economic impact.

Ma et al (2020) analyzed the mental health level of the public, like anxiety, nervousness, fear, and panic. This work found that positive emotions stood at 14.64%, neutral emotion blog posts still accounted for 66.64%, and negative emotions accounted for 18.72%. In the negative segment, the general negative emotions accounted for 12.485%, the moderate negative emotions accounted for 4.9%, and the users who showed a high degree of negative emotions accounted for 0.455% and then roughly given the relevant strategies and suggestions for some groups which have intense negative emotions. The limitation is that we can't give them specific suggestions for different users for the time being, which needs further research in the future.

Vrindavanam et al (2021) analyzed cough audio recordings; they were able to discover COVID-19 patients without using any contact. The article presents three machine learning techniques for classification and finds the best classifier among these three models. This work used a way of picking features that depended on ranking different scores generated by feature selection techniques. The preliminary findings will be a part of a broader study to build appropriate interfaces since these systems can help frontline employees by reducing stress and offering an effective approach to handling healthcare professionals' time and resources.

A new prediction approach for a collaborative filtering recommender system was developed by Alhijawi et al (2021). This prediction system is made up of a unique adaptive predictive model known as inheritance-based prediction (INH-BP) and an appropriate heuristic search method. INH-BP allows the predictor to be tailored to the user's specific needs. It uses an optimization technique, including a genetic algorithm, to form a user interest print (UIP) matrix.

Ghazarian and Nematbakhsh (2015) resolved data sparsity problems to improve memory-based approaches for group recommendation systems. The suggested technique is built around a support vector machine learning algorithm that calculates similarities between objects. This method makes use of computed similarities and improves on simple memory-based strategies.

Bhalse and Thakur (2021) proposed a movie recommendation system with the primary goal of generating a suggested list using singular value decomposed cosine similarity and collaborative filtering. The framework using a factorization form that substantially decreases the number of parameters while maintaining a regulated complexity has been improved.

1.2. Objective

The justification of the work is analyzing misconceptions about the Coronavirus and determining the influential element. In this age of global connectivity, where one erroneous thought can travel immediately to too many sensitive ears, misinformation is a profoundly harmful force. The impacts of misinformation are especially important to comprehend in the setting of a pandemic when people require trustworthy details to make key decisions that influence their own and others' well-being. During the pandemic, health emergencies, and humanitarian crises, people experience emotional, social, political, and/or economic hardship as a result of incorrect and deceptive health-related information on social media, according to systematic research.

It also has an impact on the individual who is prepared to get a Covid19 vaccination. Furthermore, it is critical to study how disinformation affects different socio-demographic categories and if populations at high risk of having serious COVID-19 problems are more prone to misinformation. Taking these issues into account, the proposed work's purpose is to thoroughly examine the disinformation of covid19 and its sources. The objective of this study is to investigate the causes that are propagating misconceptions about Covid19 and its treatment, as well as side effects, based on people's opinions. After completing the examination, identify the most influential cause that is responsible for the widespread misunderstanding. It assists individuals in making clear decisions and strengthens the mentality of those who are afraid about the Covid19 vaccinations and their treatment.

2. Materials and Methods

According to the factor-factor perception matrix, the memory-based CF recommendation system finds comparable links between variables and then recommends factors to active factors that have a high perception based on data obtained from respondents. In memory-based CF, the perception of factors from respondents is directly utilized to forecast which one has the highest value using the proposed calculation and find one factor among the COVID-19-based factors. There are two types of memory-based recommendation methods: misinformation factor-based CF Recommendation and affected Factor by misinformation-based CF Algorithm. The rationale of misinformation-based CF and affected by misinformation-based CF is presented in Figure 2.

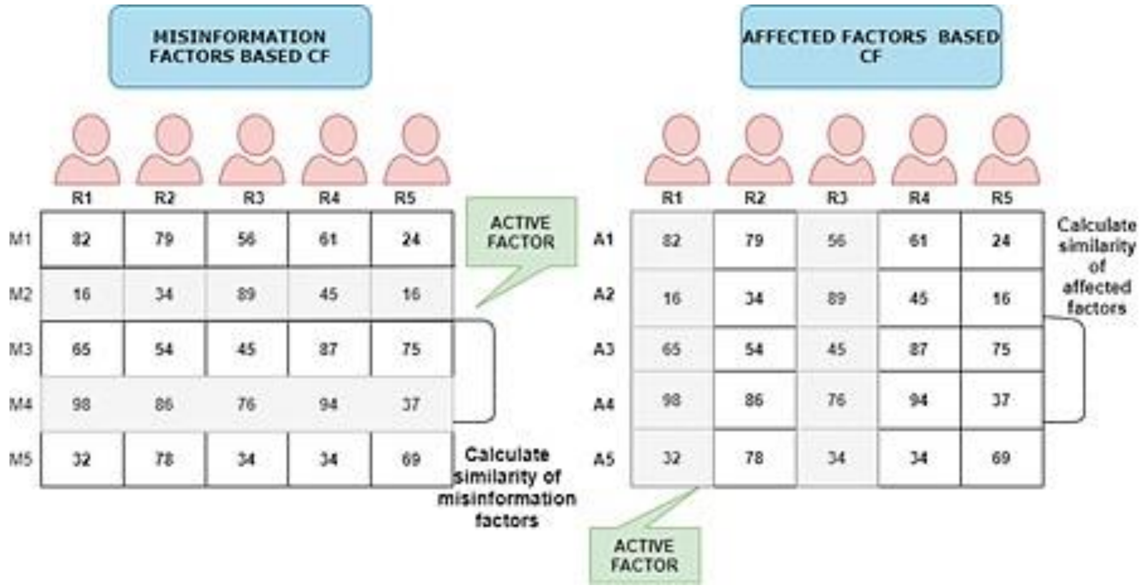

Figure 2 The rationale of misinformation-based CF and affected by misinformation-based CF.

Following Board 1 refers to the primary data based on the questionnaire type and contains 3 types of options. In this context, 100 respondents share their perceptions by choosing the options. Finally, this work calculates the primary data based on the respondents' count. It is represented in parenthesis near the option. For example, in the first question, 82 respondents have chosen option A, and 18 respondents have chosen option B.

Board 1 Primary data.

1)	Which spreads **fake reports** about corona?		
	A) Social media (0.82)	B) Video sites (0.18)	C) Search engines
2)	Which **fake remedy** is often recommended to treat corona?		
	A) (0.43) The keto diet can cure COVID-19	B) Vitamin D prevents corona (0.57)	C) Herbal remedies can help
3)	Who spreads the **conspiracy** information based on COVID-19?		
	A) (0.26) China Corona Conspiracy	B) Political Corona Conspiracy (0.72)	C) No idea
4)	Which type of **susceptibility** is caused fear in people?		
	A) Epidemic susceptibility (0.45)	B) Socio-Economic susceptibility (0.55)	C) Both A and B
5)	Which rumor spread against COVID **vaccination**?		
	A) food and drug allergies are carefully vaccinated (0.30)	B) More side effects after getting vaccinated (0.70)	C) Both A and B

Following Boards 2 and 3 refer to the secondary data based on the possibilities of respondent answers in percentage, and that is divided by the hundred and the resultant answer is taken as in the proposed analysis. For example, R1 is one of the respondents. He/she gave the perceptions based on the six affected factors. i.e., in Board 2, the first factor R1 gave their perception as 28%; after that, 28% is considered to be divided by hundred, that is 0.28. This type of data is retrieved from an online source (Nielsen et al 2021; Ilie et al 2020; Pummerer et al 2021; Augustine 2020).

Board 2 Secondary data for misinformation factors.

Sample number of Respondents	Fake reporting M1			Fake remedies M2			Conspiracy M3			Susceptibility M4			Vaccine rumors M5		
	Option %			Option %			Option %			Option %			Option %		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
R1	28	16	11	57	65	73.5	40	40	8	0.000	0.011	0.000	53	47	32
R2	35	24	23	67.7	65	47.4	6.4	6	5.75	0.099	0.000	0.249	68	38	62
R3	27	22	16	56.4	59.5	49.3	4.5	3.5	6.25	0.011	0.021	0.099	93	7	18
R4	30	21	16	50.1	60	46	6.2	5.2	7	0.130	0.249	0.390	82	80	20
R5	20	17	15	50	42.5	51	6	7	5	0.021	0.041	0.130	67	33	64

Board 3 Secondary data for affected factors by misinformation.

Sample number of Respondents	Social media A1			Vitamin D prevents corona A2			Political Corona Conspiracy A3			Socio-Economic susceptibility A4			More side effects after getting vaccinated A5		
	Option %			Option %			Option %			Option %			Option %		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
R1	26	21	16	62.5	81.5	60.6	3	7	4.75	0.153	0.390	0.525	36	95	5
R2	12	11	27	51.8	39	0	6.2	6.4	5.4	0.041	0.081	0.153	5	95	95.5
R3	27	20	13	68	19	18	5.8	4	5.8	0.178	0.525	0.666	4.5	45.7	41.3
R4	21	10	15	38	48	7	5.6	6.4	4	0.081	1.000	0.178	10.5	2.5	29.4
R5	35	21	27	25	18	9	6.4	7	6	1.000	0.666	1.000	44.3	6.3	11.1

The primary and secondary data are helping to identify the highest misinformation factor as well as to identify the affected factor by misinformation based on COVID-19 using MBCFWSM.

2.2. Factors description

2.2.1. Fake reporting

The fake reporting factor refers to news and messages that contain incorrect information or false report about the corona. The public is concerned about incorrect or misleading data regarding COVID-19 on social media such as Facebook, Google Search, YouTube, Twitter, and WhatsApp.

2.2.2. Fake remedies

The fake remedies factor denotes the wrongly spreading treatments of COVID-19 on any social media or any other platform.

2.2.3. Conspiracy

Conspiracy theories seek to explain the absolute causes of important social and political events and conditions by suggesting hidden conspiracies hatched by two or more influential actors. In the era of 2020, the Political and China-based corona conspiracy is trying to attempt confusion among the people.

2.2.4. Susceptibility

The susceptibility factor is being like to be influenced or harmed by COVID-19. Two types of Susceptibility are frequently explored in the corona mislead, like Socio-Economic susceptibility and epidemic susceptibility.

2.2.5. Vaccine rumors

Rumors and conspiracies can foster mistrust, contributing to vaccination hesitation, such as food and drug allergies, being carefully vaccinated, more side effects after getting vaccinated, etc.

2.3. Misinformation Factor-Based CF Recommendation Algorithm

2.3.1. Calculate the Similarity between Misinformation factors using cosine similarity

Cosine similarity is being used to assess responders' perceptions depending on the elements that may be represented as a vector of n-dimension, and the similarity of the components is determined by the angle of the respondents' perception vector. In general, the greater the similarity, the smaller the angle. The perceptions of the respondents are typically characterized as the perception vector $P_M\,\{P_{M_1}, P_{M_2}, \dots . P_{M_5}\}$. The similarity between the five misinformation factors is determined by comparing their perception vectors.

Cosine similarity is the traditional metric for calculating factor similarity. In cosine similarity, first the respondent's perception is gathered which represents the n-dimensional vector, and then the similarity among factors are calculated based on the angle of the factors perception vector. In general, the greater the similarity, the smaller the angle. This calculation used the primary data values for findings the similarity as well as considering the highest respondents count for calculating the similarity, such as 0.82, 0.57, 0.72, 0.55, and 0.70:

$$Sim_{A,B} = cos(A,B) = \frac{\sum_{i=1}^{n} A_i * B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} * \sqrt{\sum_{i=1}^{n} B_i^2}} \text{ As instead of } Sim_{M_1,M_2,M_3,M_4,M_5} = cos(\vec{P}_{M_1}, \vec{P}_{M_2}, \vec{P}_{M_3}, \vec{P}_{M_4}, \vec{P}_{M_5})$$

$$= \frac{\vec{P}_{M_1} \cdot \vec{P}_{M_2} \cdot \vec{P}_{M_3} \cdot \vec{P}_{M_4} \cdot \vec{P}_{M_5}}{\|\vec{P}_{M_1}\|_2 * \|\vec{P}_{M_2}\|_2 * \|\vec{P}_{M_3}\|_2 * \|\vec{P}_{M_4}\|_2 * \|\vec{P}_{M_5}\|_2}$$

$$= \frac{\sum_{i=A_{1\ to\ 5}}^{n} {}_{M_{1\ to\ 5}} M_1 * M_2 * M_3 * M_4 * M_5}{\sqrt{\sum_{I\in A_{R1,2,\ldots n}} P_{M_1}^2} \sqrt{\sum_{I\in A_{R1,2,\ldots n}} P_{M_2}^2} \sqrt{\sum_{I\in A_{R1,2,\ldots n}} P_{M_3}^2} \sqrt{\sum_{I\in A_{R1,2,\ldots n}} P_{M_4}^2} \sqrt{\sum_{I\in A_{R1,2,\ldots n}} P_{M_5}^2}} \quad (1)$$

Where $M_{1\ to\ 5}$ represents the misinformation factors and $A_{1\ to\ 5}$ denotes the affected factors by misinformation factors. Where $Sim_{M_1,M_2,M_3,M_4,M_5}$ represents the similarity between five misinformation factors, and $\vec{P}_{M_1}$ ... $\vec{P}_{M_5}$ represents the perception vectors of F_1 toF_5 respectively. $\|\vec{P}_{M_1}\|_2$ *to* $\|\vec{P}_{M_5}\|_2$ represents 2-norm of F_1toF_5, respectively, whereas F_1 *to* F_2 represents the perceptions of M_1 to M_5 on the affected factors A, respectively. A_1 to A_n represents the sets of affected factors by misinformation factors, respectively. The overall similarity between misinformation factors is calculated using equation (1).

2.3.2. Find the Nearest Neighbors using the s_4 similarity formula

This analysis considered the (Augustine 2020) proposed similarity measures for finding the nearest neighbors between M_1 to M_5 to identify the most influencing misinformation factor. The $\boldsymbol{s}_4$ similarity measure is represented in equation (2).

$$s_4(A,B) = 1 - \frac{1}{4n}\sum_{i=1}^{n} [|\mu_A(x_i) - \mu_B(x_i)| + ||\mu_A(x_i) - V_A(x_i)| - |\mu_B(x_i) - V_B(x_i)|| + ||\mu_A(x_i) - \pi_A(x_i)| - |\mu_B(x_i) - \pi_B(x_i)||] \quad (2)$$

Where, $\pi_A(x_1) = \sqrt{\mu_A^2 + V_B^2}$ and $\pi_B(x_1) = \sqrt{\mu_B^2 + V_B^2}$

Augustine (2020) proposed that s_4 is the optimized formula for handling real-time similarities, and so s_4 is used to measure the influencing factor. From the primary data, there is the respondent's count, which is used to analyze; let us represent it as (P). From the secondary data, a sample of five respondents' perceptions is collected and analyzed based on the misinformation factor, let that it be represented as (S_{M_i}). Let us consider a set of misinformation factors M = $\{M_1, M_2, M_3, M_4, M_5\}$, where M_1= Fake reporting M_2= Fake remedies, M_3= Conspiracy,M_4= Susceptibility, M_5= Vaccine rumors. Now, equation (2) is modified as equation (3).

$$s_4(P, S_{M_i}) = 1 - \frac{1}{4n}\sum_{i=1} [|\mu_P(x_i) - \mu_{S_M}(x_i)| + ||\mu_P(x_i) - V_P(x_i)| - |\mu_{S_{M_i}}(x) - V_{S_M}(x_i)|| + ||\mu_P(x_i) - \pi_P(x_i)| - |\mu_{S_M}(x_i) - \pi_{S_M}(x_i)||] \quad (3)$$

In equation (3), first, find the π value for both primary and secondary data. For example:

$$M_1 = \mu_P(M_1) = 0.82, \mu_{S_F}(M_1) = 0.28, V_P(M_1) = 0.18, \text{ and } V_{S_F}(M_1) = 0.11$$

$$\pi_A(M_1) = \sqrt{\mu_P^2 + V_P^2}$$

$$= \sqrt{1 - 0.82^2 + 0.18^2}$$

$$= \sqrt{1 - 0.6724 + 0.0324}$$

$$= \sqrt{1 - 0.7048}$$

$$= \sqrt{0.2952}$$

$$\pi_P(M_1) = 0.5433$$

$$\pi_{S_F}(M_1) = \sqrt{\mu_{S_F}^2 + V_{S_F}^2}$$

$$= \sqrt{1 - 0.28^2 + 0.11^2}$$

$$= \sqrt{1 - 0.0784 + 0.0121}$$

$$= \sqrt{1 - 0.0905}$$

$$= \sqrt{0.9095}$$

$$S_F(M_1) = 0.9975$$

After finding the π values for both primary and secondary data and applying those values in equation (3). For example:

$$M_1 = \mu_P(M_1) = 0.82,\ \mu_{S_F}(M_1) = 0.28,\ V_P(M_1) = 0.18, \text{ and } V_{S_F}(M_1) = 0.11,$$
$$\pi_P(x_1) = 0.5433,\ S_F(x_1) = 0.9975$$

$$= s_4(P, S_M) = 1 - \frac{1}{4*5}\sum_{i=1}^{5} [\,|0.82 - 0.28| + ||0.82 - 0.18| - |0.28 - 0.11|| + ||0.82 - 0.5433| - |0.28 - 0.9975||\,]$$

$$= 0.54 + |0.64 - 0.17| + |0.2767 - 0.7175|$$

$$=0.54 + 0.47 + 0.4408$$

$$M_1=1.4508$$

After calculating the remaining factors, such as M_2=0.4618, M_3=0.4096, M_4=1.2962, and M_5=0.5127.

2.3.3. Predict the highest perceptions of the misinformation factor

This section predicts which misinformation factor has the highest values that are more influencing and spreading the fake COVID-19 information among the people. For example, the M_i factor will be predicted using equation (3).

$$s_4(P,\ S_{M_1}) = 1 - \frac{1}{20}\sum_{i=1}^{5} 1.4508 + 0.4618 + 0.4618 + 1.2962+0.5127$$

$$1 - 0.05 * 4.1311$$

$$1 - 0.2065$$

$$s_4(P, S_{M_1}) = 0.7935$$

After predicting the remaining factors' values, such as $s_4(P, S_{M_2}) = 0.7499$, $s_4(P, S_{M_3}) = 0.7122$, $s_4(P, S_{M_4}) = 0.7676$, $s_4(P, S_{M_5}) = 0.7491$. From the proposed analysis, $s_4(P, S_{M_i}) = 0.7935$ has the highest value than other factors. So, the reporters have wrongly spread misinformation about COVID-19 among the people.

2.4. Affected Factor by Misinformation-Based CF Recommendation Algorithm

2.4.1. Calculate the Similarity between Misinformation factors using cosine similarity

Cosine similarity is being used to assess responders' perceptions depending on elements that may be represented as a vector of n-dimension, and the similarity of the components is determined by the angle of the respondents' perception vector. In general, the greater the similarity, the smaller the angle.

The perceptions of the respondents are typically characterized as the perception vector $P_A\{P_{A_1},\ P_{A_2}, \dots . P_{A_5}\}$. The similarity between the five misinformation factors is determined by comparing their perception vectors. Cosine similarity is the traditional metric for calculating factor similarity. In cosine similarity, first, the respondent's perception is gathered, which represents the n-dimensional vector, and then, the similarity among factors is calculated based on the angle of the factors perception vector. In general, the greater the similarity, the smaller the angle. This calculation used the secondary values for findings the similarity as well as considering the highest respondent's perception values for calculating the similarity. This calculation is similar to the Similarity between Misinformation factors using cosine similarity.

2.4.2. Find the Nearest Neighbors using the s_4 similarity formula

From the secondary data, a sample of five respondents' perceptions are collected and analyzed based on the affected factors by misinformation, let that it be represented as (S_{A_i}). Let us consider a set of misinformation factors A = $\{A_1, A_2, A_3, A_4, A_5\}$, where A_1= social media, A_2= Vitamin D prevents corona, A_3= Political Corona Conspiracy, A_4= Socio-Economic susceptibility, A_5= More side effects after getting vaccinated. Now, calculate the nearest neighbor to affected factors using eq (3).

In equation (3), first, find the π value for both primary and secondary data. For example:

$$A_1 = \mu_P(A_1) = 0.82,\ \mu_{S_F}(A_1) = 0.26,\ V_P(A_1) = 0.18, \text{ and } V_{S_F}(A_1) = 0.16$$

$$\pi_A(A_1) = \sqrt{\mu_P^2 + V_P^2}$$

$$= \sqrt{1 - 0.82^2 + 0.18^2}$$

$$= \sqrt{1 - 0.6724 + 0.0324}$$

$$= \sqrt{1 - 0.7048}$$

$$= \sqrt{0.2952}$$

$$\pi_P(A_1) = 0.5433$$

$$\pi_{S_F}(A_1) = \sqrt{\mu_{S_F}^2 + V_{S_F}^2}$$

$$= \sqrt{1 - 0.26^2 + 0.16^2}$$

$$= \sqrt{1 - 0.0676 + 0.0256}$$

$$= \sqrt{1 - 0.0932}$$

$$= \sqrt{0.9068}$$

$$S_F(A_1) = 0.9522$$

After finding the π values for both primary and secondary data and applying those values in eq (3). For example:

$$A_1 = \mu_P(A_1) = 0.82,\ \mu_{S_F}(A_1) = 0.26,\ V_P(A_1) = 0.18, \text{ and } V_{S_F}(A_1) = 0.16,$$
$$\pi_P(x_1) = 0.5433,\ S_F(x_1) = 0.9522$$

$$= s_4(P, S_{A}) = 1 - \frac{1}{4*5}\sum_{i=1}^{5}[|0.82 - 0.26| + ||0.26 - 0.18| - |0.26 - 0.16|| + ||0.82 - 0.5433| - |0.28 - 0.9522||]$$

$$= 0.56 + |0.08 - 0.1| + |0.2767 - 0.6922|$$

$$= 0.56 + 0.02 + 0.4155$$

$$A_1 = 0.9955$$

After calculating the remaining factors, such as A_2=0.7887, A_3=1.2336, A_4=1.7345, A_5=1.589

2.4.3. Predict the highest perceptions of the misinformation factor

This section predicts which affected factor has the highest values that are more caused by COVID-19 misinformation among the people. For example, the A_i factor will be predicted using eq (3).

$$s_4(P, S_{A_1}) = 1 - \frac{1}{20}\sum_{i=1}^{5} 0.9955 + 0.7887 + 1.2336 + 1.7345 + 1.589$$

$$1 - 0.05 * 6.3413$$

$$1 - 0.31706$$

$$s_4(P, S_{A_1}) = 0.6829$$

After predicting the remaining factors' values, such as $s_4(P, S_{A_2}) = 0.6906$, $s_4(P, S_{A_3}) = 0.6991$, $s_4(P, S_{A_4}) = 0.6977$, $s_4(P, S_{A_5}) = 0.7488$. From the proposed analysis, $s_4(P, S_{A_5}) = 0.7488$ has the highest value than other factors. Hence, the factor of getting more side effects after being vaccinated is the highly affected misinformation among the people.

3. Results and discussion

3.1 Similarity measure

Table 1 contains the s1 to s6 formulas and similarity calculated values between (x, y), (x, z), and (y, z).

Table 1 Analysis of the similarity measure.

Formulas	Similarity Measure		
	S (x, y)	S (x, z)	S (y, z)
s_1	0.8690	0.7813	0.9224
s_2	0.9224	0.7837	0.9479
s_3	0.8711	0.7211	0.9244
s_4	0.6745	0.6978	0.7235
s_5	0.8739	0.7817	0.9327
s_6	0.8773	0.7765	0.9410

Figure 3 represents the similarity measure using different similarity formulas. When comparing that, s4 calculates the accurate similarity of others as it provides the greatest similarity calculation between x and y, y and z, and x and z, respectively. Though S_1, s2, s3, s5, and s6 only satisfy the respective condition, it also calculates the approximate similarity more than s4.

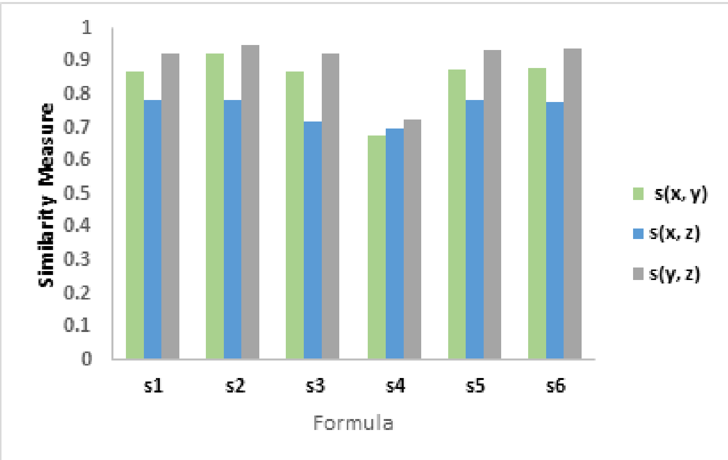

Figure 3 Compare the similarity.

3.2 Belief percentage measure

Table 2 contains the percentage values of both belief and non-belief respondents based on the vaccine rumors.

Table 2 Calculate the belief percentage among people.

Vaccine rumors	Percentage%	
	Belief	Non-belief
Scared of covid-19 vaccines	53	47
Chronic disease	68	32
Smoking	62	38
More side effects caused by a vaccine	93	7
Food/drug allergies	82	18

Figure 4 demonstrates that nearly 50 % of the respondents were initially concerned about the side effects of COVID-19 vaccination. Nearly 68% of the members' believed in chronic diseases based on the rumors before vaccination, as well as 62% of respondents believed that smokers are not to take the vaccine. 93% of respondents wrongly believed that there are more side effects caused by a vaccine. 83% of respondents believed that due to food and drug allergies, people are not taking a vaccine.

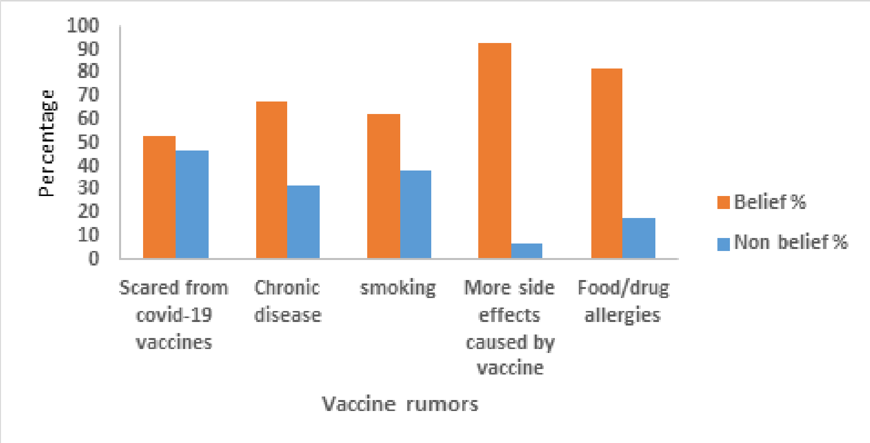

Figure 4 Compares the percentage of vaccine-based rumors between belief and non-belief.

3.3 Precision, Recall, and F-measure

Table 3 contains Precision, f-measure, and recall values for four different algorithms (Collaborative filtering, Model-based CF, Memory-based CF, and MBCFWS4).

Table 3 Precision, Recall, and F-measure comparison.

Parameters	CF (Alhijawi et al 2021)	Model-based CF (Zarzour et al 2020)	Memory-based CF (Ghazarian and Nematbakhsh 2015)	MBCFWS4
Precision	0.96	0.92	0.94	0.95
Recall	0.94	0.97	0.96	0.98
F-measure	0.93	0.94	0.98	0.99

The proposed approach performed better than the baseline works (CF, Model-based CF, and Memory-based CF), as seen in Figure 5. From the analysis, the precision, recall, and F-measure values are low for the existing works. Due to the issues of CF based techniques (Alhijawi et al 2021), it is unable to identify the latent relationship between the related factors. When there are no prior ratings from the people, CF has trouble making accurate suggestions. Then, the memory-based CF model (Ghazarian and Nematbakhsh 2015) suffers from data sparsity, which happens when the users score only a few factors out of a huge number of factors provided. Furthermore, when the number of users and factors rises, the processing complexity will also increase, resulting in limited scalability and difficulty in making suggestions. The model-based approaches need a significant amount of time and memory to create the model and may result in information loss when dimensionality reduction is used. Finally, the proposed approach's performance is found to be better in terms of Accuracy, F-score, and recall. As because the suggested approach lowers uncertainty and ambiguity, lowering these two factors will result in less confusion in decision-making and more efficiency in uncertain circumstances.

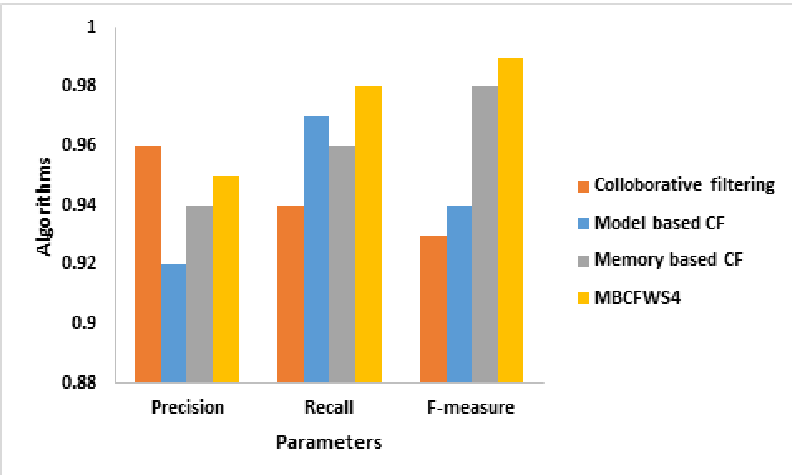

Figure 5 Compare the precision, recall, and F- measure.

3.4. Findings

One of the most important essential things for humans is the healthcare system. The COVID-19 pandemic is influenced by the availability of information, the majority of which is erroneous, causing a hazard to both physical and mental conditions. Despite the impact of sharing the COVID-19 pandemic misinformation, all people need to make the right decisions about COVID-19. This research found that the fake report factor has high impact and highly sharing of COVID-19 misinformation than all the others. Boards 2 and 3 depict the sample perceptions of the respondents on COVID-19 misinformation (M) and affected by misinformation (A) online. It is learned from the respondent's perception that most of the respondents' choice reporters are wrongly reporting on COVID-19. Public concern over COVID-19's incorrect or misleading content is placed on social media. Also, from the primary data analysis found in (Board 1), it is clear that counting based 82% of respondents gave their opinion that they are unsatisfied with social media.

4. Conclusions and recommendations

This study introduced a novel method for discovering fake health-related information, and it will aid in the detection of false information on COVID pandemic issues. The proposed work makes the factor-based analysis to conclude the more impact factor from the respondent's side. This analysis of the primary and secondary datasets helps to identify and collect the perceptions of the respondent's that are in the MBCFWS4 measure. In the MBCFWS4 process, first, the factors are categorized into two types, namely misinformation-based CF and affected by misinformation-based CF. The misinformation-based factor analysis found that the fake report factor has the most impact on the respondent's side and was affected by misinformation-based factor analysis found that social media has highly spread COVID-19 misinformation from the respondent's perception. Hence, it is recommended to avoid sharing COVID-19-based misinformation among the people. In state-of-the-art work, similarity measure of s4 is the optimized formula to measure the similarity between one or more variables and parameters. So, this analysis used s4 for calculating all the values based on factors during evaluation. From the efficiency analysis, the precision, f-score, and recall of the CF-based proposed work are higher than the baseline works. Also, 93% of respondents wrongly believed that more side effects were caused by the COVID vaccine, which is found from the belief percentage analysis. The study ends with a discussion of various limitations, which are associated with the data source, the schematization of misinformation, and the time required for data collection. More research is needed to better comprehend the contents, origins, effects, and further key components of COVID-19 misinformation in India, according to this analysis, which found a few knowledge gaps.

4.1. Recommendations

Some platform organizations capitalized on COVID-19 data centers and specialized knowledge panels, and they presented health authority information in several ways. They have also witnessed misinformation, prominent anti-vax organizations, and occasionally radical lockdown demonstrators, as well as concerted intimidation of journalists and specialists employed on COVID-19. The platforms most commonly utilized as a source of information or news on coronavirus vary by country, although in virtually all cases, Google Search, Facebook, Twitter, YouTube, and WhatsApp having high popularity. From the proposed analysis, it was found that the fake report factor has a high impact on COVID-19 misinformation because, from early 2020, social media had highly shared fake information about the COVID-19 pandemic.

List of Abbreviations

CF	Collaborative filtering
COVID	Coronavirus
MBCF	Memory-based collaborative filtering

Ethical considerations

Before taking part in the study, all individuals provided informed consent. The study was carried out under the Helsinki Declaration, and the protocol was approved by the Ethics Committee.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

This research did not receive any financial support.

References

Pak A, Adegboye OA, Adekunle AI, Rahman KM, McBryde ES, Eisen DP (2020) Economic consequences of the COVID-19 outbreak: The need for epidemic preparedness. Frontiers in Public Health 8. DOI: 10.3389/fpubh.2020.00241

Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman AL, Recchia G, Van der Bles AM, Van der Linden S (2020) Susceptibility to misinformation about COVID-19 around the world. Royal Society Open Science 7:201199. DOI: 10.1098/rsos.201199

Elhadad MK, Li KF, Gebali F (2020) Detecting misleading information on COVID-19. IEEE Access 8:165201. DOI: 10.1109/access.2020.3022867

Al-Zaman MS (2021) A thematic analysis of misinformation in India during the COVID-19 pandemic. International Information and Library Review 1. DOI: 10.1080/10572317.2021.1908063

Ejegwa PA (2020) New similarity measures for Pythagorean fuzzy sets with applications. International Journal of Fuzzy Computation and Modelling 3:75. DOI: 10.1504/ijfcm.2020.106105

Peng X (2018) New similarity measure and distance measure for Pythagorean fuzzy set. Complex and Intelligent Systems 5:101. DOI: 10.1007/s40747-018-0084-x

Chen R, Hua Q, Chang Y, Wang B, Zhang L, Kong X (2018) A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks. IEEE Access 6:64301. DOI: 10.1109/access.2018.2877208

Hatmal MM, Al-Hatamleh MA, Olaimat AN, Hatmal M, Alhaj-Qasem DM, Olaimat T, Mohamud R (2021) Side effects and perceptions following COVID-19 vaccination in Jordan: A randomized, cross-sectional study implementing machine learning for predicting severity of side effects. Vaccines 9:556. DOI: 10.3390/vaccines9060556

Roy S, Ghosh P (2020) Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. PLOS ONE 15:e0241165. DOI: 10.1371/journal.pone.0241165

Wang J, Wang Z (2020) Strengths, weaknesses, opportunities and threats (SWOT) analysis of China's prevention and control strategy for the COVID-19 epidemic. International Journal of Environmental Research and Public Health 17:2235. DOI: 10.3390/ijerph17072235

Tariq Soomro Z, Waseem Ilyas SH, Yaqub U (2020) Sentiment, count and cases: Analysis of Twitter discussions during COVID-19 pandemic. 2020 7th International Conference on Behavioural and Social Computing (BESC). DOI: 10.1109/besc51023.2020.9348291

Junlan Z, Chengke Y (2020) Data-analysis-based discussion on COVID-19 pandemic shocks to the economy and policy responses: Cases in the United Kingdom. 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID). DOI: 10.1109/msieid52046.2020.00109

Ma X, Wang J, Qin H, Wang Y (2020) Public opinion analysis about the coronavirus COVID-19 based on micro-blog data. 2020 International Conference on Information Science and Education (ICISE-IE). DOI: 10.1109/icise51755.2020.00104

Vrindavanam J, Srinath R, Shankar HH, Nagesh G (2021) Machine learning based COVID-19 cough classification models – A comparative analysis. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). DOI: 10.1109/iccmc51019.2021.9418358

Alhijawi B, Al-Naymat G, Obeid N, Awajan A (2021) Novel predictive model to improve the accuracy of collaborative filtering recommender systems. Information Systems 96:101670. DOI: 10.1016/j.is.2020.101670

Ghazarian S, Nematbakhsh MA (2015) Enhancing memory-based collaborative filtering for group recommender systems. Expert Systems with Applications 42:3801. DOI: 10.1016/j.eswa.2014.11.042

Bhalse N, Thakur R (2021) Algorithm for movie recommendation system using collaborative filtering. Materials Today: Proceedings. DOI: 10.1016/j.matpr.2021.01.235

Nielsen RK, Schulz A, Fletcher R (2021) An ongoing infodemic: How people in eight countries access news and information about Coronavirus a year into the pandemic. Availale in: https://reutersinstitute.politics.ox.ac.uk/ongoing-infodemic-how-people-eight-countries-access-news-and-information-about-coronavirus-year Accessed on: November 4, 2021.

Ilie PC, Stefanescu S, Smith L (2020) The role of vitamin D in the prevention of coronavirus disease 2019 infection and mortality. Aging: Clinical and Experimental Research 32:1195.

Pummerer L, Böhm R, Lilleholt L, Winter K, Zettler I, Sassenberg K (2021) Dataset for: Conspiracy theories and their societal effects during the COVID-19 pandemic [Data set]. PsychArchives. Availale in: https://www.psycharchives.org/handle/20.500.12034/4095 Accessed on: November 4, 2021.

Augustine EP (2020) New similarity measures for Pythagorean fuzzy sets with applications. International Journal of Fuzzy Computation and Modelling 3:75. DOI: 10.1504/ijfcm.2020.10027701

Zarzour H, Jararweh Y, Al-Sharif ZA (2020) An effective model-based trust collaborative filtering for explainable recommendations. 2020 11th International Conference on Information and Communication Systems (ICICS). DOI: 10.1109/icics49469.2020.239555