Original Article

Extracting Hidden Patterns of Iranian User Trust in Social Networks Regarding Coronavirus Disease 2019 Using Data Mining Techniques

Majid Jangi, Maryam Jahanbakhsh, Nahid Tavakoli, Hossein Bagherian, Asghar Ehteshami, Sakineh Saghaeian Nejad Isfahani, Mohammad Sattari

Health Information Technology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

Abstract

Aim: The coronavirus disease 2019 (COVID-19) pandemic caused the use of social networks in the field of information acquisition and transmission to increase, whereas the validity of the information available is questionable. Because people's trust in these networks is important, this study aimed to utilize three data mining techniques to identify the hidden rules for detecting the user trust level of social networks in the context of COVID-19. **Materials and Methods:** An electronic questionnaire containing 27 questions was provided to users. Out of the 12 questions selected, the final question asked about the level of user trust in social networks and was considered the target class. Based on the range in value, question 12 was divided into five classes. The relevance of the remaining 11 questions was then assessed using three decision tree-based data mining techniques. **Results:** The results showed that the random forest technique performed better than the other techniques. Most social network users have a moderate level of trust in information regarding COVID-19; in fact, the medium class is the most widely used target class with 60% utilization rate, which affects sensitivity and specificity. The values of these measures were much higher for this class than for the other classes. **Conclusion:** The educational content, both its type and the amount, regarding COVID-19 that is provided on social networks, impact on user trust. As the existence of inconsistent information has had a negative impact on user trust, a small percentage of users have high trust in these networks.

Keywords: Coronavirus disease 2019, data mining, social networks, trust

Introduction

Many countries were infected with Coronavirus disease 2019 (COVID-19), a new coronavirus disease, in late 2019 and early 2020 which caused various economic, political, social, and health challenges. The dissemination of information regarding this disease has directly or indirectly affected some of these challenges. Delivering the right information at the right time to the right person can reduce or slow down the progress of some challenges.[1] Information related to the emergence and epidemic of COVID-19 has created tremendous fluctuations in public opinion and has had an important impact on the new information environment.[2] A huge amount of information about the technology, medicine, social, political, and historical context of COVID-19 is published daily on virtual social networks with various typologies, including valid information, cheerful information, misleading information, quasi-information, and shocking information.[3]

Understanding the information dissemination process on social networks and its effects on public behavior can aid the design of effective strategies to implement public health interventions. Social networks strongly influence people's attitudes, beliefs, and behaviors toward any type of disease, including COVID-19.[4] Expanded information about COVID-19 can have a significant impact on people's behavior and the effectiveness of government measures for disease control.[5]

Address for correspondence: Dr. Mohammad Sattari, Health Information Technology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran.
E-mail: msattarimng.mui@gmail.com

Access this article online

Quick Response Code:

Website: https://journals.lww.com/IJEH

DOI: 10.4103/ijehe.ijehe_39_23

How to cite this article: Jangi M, Jahanbakhsh M, Tavakoli N, Bagherian H, Ehteshami A, Isfahani SS, *et al*. Extracting hidden patterns of Iranian user trust in social networks regarding coronavirus disease 2019 using data mining techniques. Int J Env Health Eng 2024;13:11.

Received: 02-07-2023, **Revised:** 29-02-2024, **Accepted:** 02-03-2024, **Published:** 12-07-2024

By providing direct access to an unprecedented amount of information, social networks reinforce the possibility of the spread of rumors and suspicious information. Moreover, users' preferences and attitudes and the increasing amount of content on social networks facilitate the dissemination of information.[5]

Determining the information-seeking process or avoidance of information and its effects on people is an important research challenge, especially when a news cycle changes the methods of information reporting, dissemination, and consumption. The term "infodemic" has been invented to describe the risks posed by anti-information phenomena during outbreak management.[2]

The use of new methods for data analysis, including data mining techniques in virtual social networks, has recently been considered, and thus, the volume of data in this space is expanding.[6] Data mining uses various statistical methods, machine learning, and artificial intelligence to discover the hidden relationships between data.[7] Classification and clustering are two common methods in data mining that are widely used in various sciences.[8] The decision tree is one of the most widely used classification techniques.[9]

Shu *et al*. used a linguistic-based data mining method for fake news detection. Characterization and detection are two phases of fake news detection.[10] Sutar used the k-means method to categorize information extracted from university students in social media.[11] Espíndola suggests a framework for the side effects of drugs. He used social media data to extract positive and negative responses about these side effects.[12]

It is important to trust or distrust social networks due to the large volume of rumors found on them and the high rate of use these networks have among users.[13] In most cases, these rumors can rile up the emotions, change the mood, and stick in your head. Several factors affect user trust in social networks. No paper assesses Iranian user trust regarding COVID-19. Three decision-tree techniques were used to extract hidden rules for detecting the trust level of virtual social networks regarding information on COVID-19. These hidden rules help make better decisions about pandemic news.

Materials and Methods

Article extracted from a research project "Investigating the role of social networks in shaping people's attitudes toward COVID-19" with code IR.MUI.RESEARCH.REC.1398.781.

The standard data mining process method released by Romero *et al*.[14] was applied in this study to gather and analyze data as described below.

Data gathering

Data were collected from an electronic questionnaire that investigated the effects of social networks on the attitude of people regarding COVID-19. The data were related to participants over 10 years of age in a 27-question framework. Data mining determined the accuracy of the hypothesis. Participants are Iranian users of social networks consisting of Telegram, WhatsApp, Instagram, and Iranian Social Networks. Cronbach's alpha equal to 0.86 confirmed the content and constructive validity and reliability of the questionnaire. The participants were randomly selected from 29 provinces across the country. The questionnaire link was distributed on various social networks. Everyone entered the questionnaire and completed it if they wished.

The questions were divided into subgroups of demographic and open questions and were answered using a Likert-type scale with the values of very high, high, medium, low, and very low. Overall, 1000 questionnaires were considered through March 24, 2020.

Data preprocessing

Data preprocessing has two parts. The first part is removing, where some unrelated questions are removed. Eleven questions were removed and 16 multiple-choice questions (Likert scale) remained. The second part is target selection and the last question is selected as target. The last question was selected as the data mining target,[1] and four questions not related to the target question were removed. Finally, 11 questions remained [Appendix Table 1].

Classification

The decision tree is a simple and robust method of multivariate analysis that presents classification results by simple and comprehensible graphs.[15] This approach is used in many areas, including pattern recognition, pattern classification, classification, decision support systems, and expert systems. Iterative Dichotomiser 3 (ID3) is one of the simplest decision-tree algorithms, which selects and classifies the first attribute with the highest information load.[16] Another algorithm is the random forest decision tree, which is a machine-learning method. In this algorithm, different trees of datasets are created with different properties, and the best decision from the trees determines the index to be associated with the class.[17] Chi-square automatic interaction detection (CHAID) is a decision tree-based method that divides the data space into different subsets in different iterations. It starts from the root node and creates two or more nodes in each segmentation. The algorithm uses the Chi-square test to decide on each subdivision for identifying child nodes.[18] A version of a neural network contains three layers called the input layer, the processing layer, and the output layer. The dependent variable (target) corresponds to the output layer, and the independent variables correspond to the input layer.[19]

RapidMiner, version 7(Altair Engineering, USA, Massachusetts), was the software tool used to analyze data classification. Decision tree algorithms were used to construct classifiers. The last question was selected for the class label. This question had five values (very high, high, medium, low, and very low), and each of these values was selected as one of the classes. The 11 remaining questions had five values considered class labels to split the decision tree.

Therefore, for each subject, five child nodes were created. Five techniques (CHAID, ID3, Random Forest, Neural Net, and AutoMLP [multilayer perceptron]) were used. There of the five (CHAID, ID3, Random Forest) were according to a decision-tree algorithm. These techniques predict a class label from the root of the tree. The random forest technique worked randomly and used several trees for decision-making. Two of the five (AutoMLP and NeuralNet) were according to neural network algorithms.

The data were split into training (learning model) and testing (evaluating model) set in the ratio of 70% to 30%. The training data and the last question (class label) were input into RapidMiner software. We use three techniques were performed on the training data, and the results were represented in both text and graphs. The steps of rapidminer for ID3 techniques are shown in Figure 1. Select attribute module choose attributes and set the role module to specify the set the target attribute. ID3 module does implemented based on specified attributes.

Hidden patterns extraction

Hidden patterns were extracted according to the trees created in previous steps. These patterns then determined the rules for producing the final class. The rules illustrated the impact of answering the questions on the target question.

Results

Seven measures were used for evaluation. Accuracy was considered the ratio of true values compared to all values. Precision specified how many returned results were positive. Sensitivity specified how many positive results were returned, and specificity was how many negative results were returned. Spearman's correlation coefficient was used in this study.[20] The correlation coefficient was used to determine the direct or inverse relationship between two variables as well as the severity of the relationship between them. Other measures are *F*-score and G-mean.

$$GM = \sqrt{sensitivity \times specificity}$$

The statistical population of this study comprised 1000 respondents to the questionnaire, the demographic characteristics of whom are listed in Table 1.

Data were divided into two parts: training (70%; 700 respondents) and testing (30%; 300 respondents). The Appendix Table 1 lists the 12 selected questions. The first 11 questions were considered input, and the last question was taken as output or class. Table 2 presents the amount of trust that questionnaire respondents have in social network information regarding COVID-19. Most respondents (approximately 60%) reported a medium level of trust in social networks. Thus, the medium class is the most frequent in this study. Approximately 18%, 13%, and 59.8% of respondents expressed a low, high, and medium level of trust in social networks regarding COVID-19 information, respectively.

The CHAID, ID3, and random forest techniques, which are based on the decision-tree approach, were then used to extract hidden rules between classes and the responses to the other questions. Rapidminer software was used to apply these techniques. The random forest method (62.33%) was found to be the most accurate technique. Due to its random nature, the random forest algorithm was executed 10 times, and the average of runs was taken as the accuracy [Table 2]. The integration of 10 random trees can lead to the better performance of this method.

The results for each technique based on the precision measure are presented in Table 2. The random forest technique had the greatest precision in the low class, and the ID3 technique had the greatest precision in the very high class. Moreover, in the medium class, the CHAID technique performed better in terms of this measure [Table 2]. The minimum distances of 60% and 40% of the random forest technique in the low and high classes, respectively, were significantly different from the other techniques [Table 2]. CHAID got the highest *F*-score in the medium class and random forest also in this class reached the best performance based on f1-score. AutoMLP in high class got better performance than other methods [Table 2].

Table 3 represents the performance of classification sensitivity and specificity measures of the datasets using classification algorithms such as random forest, CHAID, ID3, Neural Net, and AutoMLP. In all three techniques, the best performance in terms of sensitivity and specificity was in the medium class, where the random forest technique performed better than the

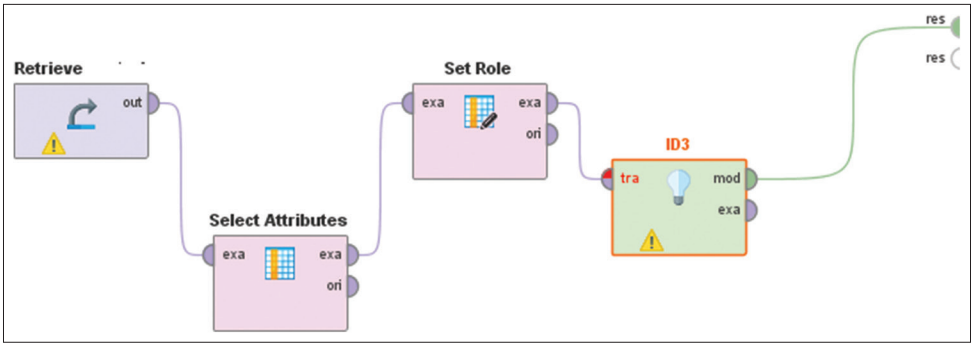

Figure 1: *Iterative Dichotomiser 3*-rapidminer

Table 1: Demographic information of respondents to the questionnaire

Age -on average	Gender		Job			Grade	
	Female, *n* (%)	Male, *n* (%)	Unemployed, *n* (%)	Employed by government, *n* (%)	Self-employment, *n* (%)	Bachelor degree or less, *n* (%)	Master's degree or PhD, *n* (%)
37	578 (57)	432 (43)	180 (33.3)	180 (33.3)	180 (33.3)	180 (50)	180 (50)

Table 2: The precision, F1-score and accuracy of techniques

	CHAID		ID3		Neural Net		AutoMLP		Random forest	
Accuracy	5		59.3		62.6		6		6	
Class	Random forest		CHAID		ID3		Neural Net		AutoMLP	
	Precision (%)	F1-score (%)	Precision (%)	F1-score (%)	Precision (%)	F1-score (%)	Precision (%)	F1-score (%)	Precision (%)	F1-score (%)
Very low	55.56	35.71	40.00	35.29	38.89	35	31.85	35.29	47.06	42.1
Low	75	15	40.48	41.46	18.60	19.51	15.79	11.11	47.18	25
Medium	61.22	77.03	70.56	75.59	62.30	65.51	67.86	71.5	67.71	74.71
High	77.78	35.89	48.28	47.45	44	40	57.58	55.88	51.25	62.29
Very high	83.33	50	60.00	30	66.67	36.36	5	7.21	1	3.23

CHAID: Chi-square automatic interaction detection, ID3: Iterative dichotomiser 3

other two techniques in this class [Table 3]. The random forest technique had the highest sensitivity for the medium class. This indicates that this technique has returned approximately 98% of the positive results for this class [Table 3].

The rules have been arranged in Table 4 based on confidence scores. The 22 rules were extracted; however, the rules that support were more than 35 only considered. The best confidence rule was "If the reliability of tips and training gotten from social networks is medium, the amount social networks affect your awareness of COVID-19 signs and symptoms is much, and the spread of rumors about COVID-19 by these networks is high. Then, the reliability of information about COVID-19 on social networks was medium."

Rule 7 represents that the high reliability of social networks depends on the high introduction of preventive and high-risk behavior of COVID-19 if the contradictory content of social networks is medium.

Table 5 confirms the results shown in Table 4. Most of the questions listed in Table 5 have a higher correlation coefficient than other questions with the class (target) question. Furthermore, the negative correlation coefficient of question 9 is justified by rule 13.

Discussion

This paper used data mining techniques to assess the effects of social networks on public behavior regarding COVID-19. The results of the current paper show that a large number of users have a moderate level of trust in social networks regarding COVID-19. The best confidence represents that if the reliability of tips and training gotten from social networks is medium, the amount of effeteness of social networks on your awareness of COVID-19 signs and symptoms is much, and the spread of rumors about COVID-19 by these networks is high then the reliability of information about COVID-19 on social networks is medium. These rules show that the amount of effeteness of social networks on your awareness of COVID-19 signs and symptoms is much, the spread of rumors about COVID-19 by these networks is high, the reliability of tips and training gotten from social networks is medium, and the reliability of information about COVID-19 on social networks is medium.

Only 12% trust these networks much and accept them as a basis for getting information about COVID-19. These people have had a negative effect on other people's opinions. Given the rapid growth rate of rumors on social networks[13] and the impact these rumors have, overreliance on these networks can have devastating effects on people's lives. The source of information in this network is unclear, so not just any information can be trusted.

A new survey was conducted by the US Pew Research Center on social networking information about COVID-19. The results indicated that more than half of the people who participated in the survey did not trust the information received from social networks and believed that this information was fake.[21]

Studies have shown that mutual trust is one of the prerequisites for social network success.[22] Social influence and ease of use had significant and positive effects on trust in social networks. A positive and significant relationship was also found between the variables of trust in and continued use of social networks.[23] The findings of the current study represent that the more reliable the tips and training gotten from social networks are, the more reliable the information about COVID-19 on social networks will be. Rule 9 represents that If the reliability of tips and training gotten from social networks is very high, the amount you and your family take

Table 3: Sensitivity, specificity and G-mean of techniques

Class	Random forest (%)			CHAID (%)			ID3 (%)			Neural net (%)			Auto MLP (%)		
	Sensitivity	Specificity	G-mean	Sensitivity	Specificity	G-mean	Sensitivity	Specificity	G-mean	Sensitivity	Specificity	G-mean	Sensitivity	Specificity	G-mean
Very low	25	97.86	49.46	30	95.1	53.41	35	94.61	57.54	30	93	52.82	40	95	61.64
Low	7.5	99.46	27.31	42.5	86.77	36.84	20	82.63	40.64	7.5	95	26.69	17.5	94	40.55
Medium	98.17	24.54	49.04	77.44	58.71	67.42	69.51	39.42	52.325	81.1	46	61.07	79.27	46	60.35
High	22.95	97.86	47.37	45.9	83.6	61.94	36.07	81.60	54.19	62.30	83	71.9	62.30	80	70.59
Very high	33.33	99.45	57.535	20	98.94	44.47	26.67	98.69	51.21	5	98	22.135	9	100	30

MLP: Multilayer perceptron, CHAID: Chi-square automatic interaction detection, ID3: Iterative dichotomiser 3

care of personal health issues is high and decreases the amount you and your family commute to public places a high amount, the reliability of information about COVID-19 on social networks is very high.

The findings also showed that the more people receive contradictory information from social networks, the lower their confidence will be in the networks. The qualitative research results of interviews with 115 chat users showed that participants tend to receive valuable information. Unpleasant information such as advertising and bad news weakens the motivation to continue using a network and affects the level of trust.[24] Rule 8 represents that if contradictory content regarding COVID-19 have you received from social networks is high, have social networks contributed to inform you about COVID-19 care centers is low and use of social networks has led to the spread of rumors about COVID-19 is high, and the reliability of information about COVID-19 on social networks is low. These results show the effectiveness of contradictory content regarding COVID-19 on the reliability of information about COVID-19.

Chang and Chuang found that trust, social interactions, and reputation had positive effects on the quality of knowledge-sharing behaviors in social networks.[25] Trust levels influence the willingness and behavior of individuals. Many studies represent that social trust is an important factor in the success of knowledge dissemination, and social networks have a direct effect on this attitude.[26] A study at the Azad University of Mashhad showed that there is a relationship between trust and the amount of usage of social networks.[27]

The decision tree is one of the best classification methods in data mining for modeling. Using a variety of algorithms, this method can create robust models with different datasets and perform accurate data classification. Among different machine learning methods, decision-tree algorithms provide the best results in classifying data.[28] The decision tree often performs much better than most other machine learning algorithms and makes stronger models for a small amount of data.[29]

Five data mining techniques, CHAID, ID3 and the random forest technique, neural network, and AutoMLP were used in the current study to predict user confidence based on 12 selected questions. The random forest technique was found to be more accurate than the other tested techniques. One reason for the higher accuracy of this algorithm is that it uses several trees to make a final decision.[30] Maroco *et al*. showed that the random forest technique was more accurate than other machine learning techniques.[31] The F1 score of all techniques in the medium class is higher than other classes. The precision of CHAID method in the medium class is more than other techniques.

Because the medium class covers 60% of the core classes, the sensitivity for all techniques in this class will be higher than the other measures. In fact, the rate of negative detection modes, namely, the absence of a class, was reduced. Thus, sensitivity measure was affected.[32] The specificity measures for the very

Table 4: Extracted rules

Rules	Conditions	Result	Confidence
Rule 1	If the reliability of tips and training gotten from social networks is medium, the amount social networks affect your awareness of COVID-19 signs and symptoms is much, and the spread of rumors about COVID-19 by these networks is high	The reliability of information about COVID-19 on social networks is medium	0.87
Rule 2	If social networks affect your awareness of COVID-19 signs and symptoms a medium amount and in decrease the amount you and your family commute to public places a medium amount	The reliability of information about COVID-19 on social networks is medium	0.81
Rule 3	Effective have social networks been in raising your awareness of COVID-19 signs and symptoms is high, contradictory content regarding COVID-19 have you received from social networks is medium	The reliability of information about COVID-19 on social networks is medium	0.75
Rule 4	If the amount social networks affect your awareness of COVID-19 signs and symptoms is high, decreases the amount you and your family commute to public places a high amount, and the amount you and your family take care of personal health issues is high	The reliability of information about COVID-19 on social networks is high	0.76
Rule 5	The effectiveness of social networks been in introducing to you preventive behaviors for COVID-19 is very high and the use of social networks has led to the spread of rumors about COVID-19 in medium	The reliability of information about COVID-19 on social networks is medium	0.56
Rule 6	If the reliability of tips and training gotten from social networks is high and the contradictory content of COVID-19 received from social networks is low	The reliability of information about COVID-19 on social networks is medium	0.56
Rule 7	If social networks been in introducing to you preventive behaviors for COVID-19 is very high, social networks been in introducing to you high-risk behaviors for COVID-19 is very high and the amount of contradictory content regarding COVID-19 have you received from social networks is medium	The reliability of information about COVID-19 on social networks is high	0.54
Rule 8	contradictory content regarding COVID-19 have you received from social networks is high, have social networks contributed to inform you about COVID-19 care centers is low and use of social networks has led to the spread of rumors about COVID-19 is high	The reliability of information about COVID-19 on social networks is low	0.52
Rule 9	If the reliability of tips and training gotten from social networks is very high, the amount you and your family take care of personal health issues is high and decreases the amount you and your family commute to public places a high amount	The reliability of information about COVID-19 on social networks is very high	0.5

COVID-19: Coronavirus disease 2019

Table 5: Correlation coefficient of question to class (target) question

Row	Questions	*P*	Correlation coefficient
1	How effective have social networks been in raising your awareness of COVID-19 signs and symptoms?	<0.001	0.482
2	How effective have social networks been in introducing to you preventive behaviors for COVID-19?	<0.001	0.449
3	How effective have social networks been in introducing to you high-risk behaviors for COVID-19?	<0.001	0.443
4	How much have social networks led to you and your family taking care of personal health issues?	<0.001	0.445
5	How much have social networks led to decreasing the amount you and your family commute to public places?	<0.001	0.409
6	How much have social networks led to a decline in the amount you and your family go to parties?	<0.001	0.356
7	How much do you follow the tips and training you get from social networks?	<0.001	0.379
8	How much do you trust the tips and training you get from social networks?	<0.001	0.521
9	How much contradictory content regarding COVID-19 have you received from social networks?	<0.001	−0.240
10	How much have social networks contributed to inform you about COVID-19 care centers?	<0.001	0.332
11	How much use of social networks has led to the spread of rumors about COVID-19?	<0.001	−0.305

COVID-19: Coronavirus disease 2019

high class for all techniques than other classes. Maroco *et al*. also showed that the random forest and CHAID algorithms had similar specificity, but the random forest technique performed better based on the sensitivity algorithm.[31]

Conclusion

This paper shows that most users of social networks have a moderate level of trust in the information regarding COVID-19 they obtain from these networks. The random forest technique, used on the basis of the opinions of several trees, performed better than other techniques. These results suggest using these methods for other similar datasets. The educational content regarding COVID-19 that was provided by these networks can be effective in building trust in these networks. These contents can be provided by physician and medical professors. Conversely, the existence of inconsistent information has had a negative impact on user trust. Inconsistent information can spread as a rumor and effect user trust. Only a small percentage of users have a high trust in these networks.

Acknowledgments

We acknowledge the Health Information and Technology Research Center, Isfahan University of Medical Sciences,

Isfahan, Iran. This article was extracted from a research project, titled: "Investigating the role of social networks in shaping people's attitudes toward COVID-19," Research Grant #198243.

Financial support and sponsorship

Nil.

Ethics Code

IR.MUI.RESEARCH.REC.1398.781.

Conflicts of interest

Authors represent that they have no conflict of interest.

Authors' Contributions

Mohammad Sattari and Majid Jangi: manuscript writing; Maryam Jahanbakhsh, Nahid Tavakoli, Hossein Bagherian, Asghar Ehteshami, and Sakineh Saghaeian Nejad Isfahani: manuscript editing.

References

1. Ashrafi-Rizi H, Kazempour Z. Information typology in coronavirus (COVID-19) crisis; a Commentary. Arch Acad Emerg Med 2020;8:e19.
2. Yin FL, Lv JH, Zhang XJ, Xia XY, Wu JH. COVID-19 information propagation dynamics in the Chinese Sina-microblog. Math Biosci Eng 2020;17:2676-92.
3. Jahanbin K, Rahmanian V. Using twitter and web news mining to predict COVID-19 outbreak. Asian Pacific journal of tropical medicine: 2020; 13:378-80.
4. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, *et al.* Coronavirus disease 2019 (COVID-19): A perspective from China. Radiology 2020;296:E15-25.
5. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, *et al.* The COVID-19 Social Media Infodemic. arXiv Preprint arXiv: 200305004; 2020.
6. Barbier G, Liu H. Data mining in social media. In: Social Network Data Analytics. Germany: Springer; 2011. p. 327-52.
7. Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee YH, *et al.* Screening for prediabetes using machine learning models. Comput Math Methods Med 2014;2014:618976.
8. Dalal M, Harale N, editors. A Survey on Clustering in Data Mining. Proceedings of the International Conference and Workshop on Emerging Trends in Technology; 2011.
9. Jafari B, Azar A. Fuzzy decision tree; the new approach in strategy formulation. Manage Res 2013;6:25-39.
10. Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explor Newsl 2017;19:22-36.
11. Sutar SG. Intelligent data mining technique of social media for improving health care. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). India: IEEE; 2017. p. 1356-60.
12. Espíndola RP, Ebecken NF. On extending f-measure and g-mean metrics to multi-class problems. WIT Transactions on Information and Communication Technologies. 2005;35:25-34.
13. Doerr B, Fouz M, Friedrich T. Why rumors spread so quickly in social networks. Commun ACM 2012;55:70-5.
14. Romero C, Ventura S, García E. Data mining in course management systems: Moodle case study and tutorial. Comput Educ 2008;51:368-84.
15. Jawad SF, Towfik ZS. Application of decision tree as a data mining tool in a health care. J Kufa Math Comput 2012;1:102-9.
16. Bhatt H, Mehta S, D'mello LR. Use of ID3 decision tree algorithm for placement prediction. Int J Comput SciInf Technol 2015;6:4785-9.
17. García MN, Herráez JC, Barba MS, Hernández FS, editors. Random forest based ensemble classifiers for predicting healthcare-associated infections in intensive care units. In: Distributed Computing and Artificial Intelligence, 13th International Conference. Switzerland: Springer; 2016.
18. Hill DA, Delaney LM, Roncal S. A Chi-square automatic interaction detection (CHAID) analysis of factors determining trauma outcomes. J Trauma 1997;42:62-6.
19. Babinec T. Neural Networks and Statistical Models. In: Sawtooth Software Conference; 1997.
20. Zar JH. Spearman rank correlation. Encyclopedia Biostat 2005;15:7-17.
21. Hatmaker T. People Who Mostly get News from Social Networks have Some COVID-19 Misconceptions; 2020. Available from: https://techcrunch.com/2020/03/25/pew-coronavirus-covid-19-survey/. [Last accessed on 2020 Mar 26].
22. Zuo Y, Hu W-c, O'Keefe T, editors. Trust computing for social networking. In: 2009 Sixth International Conference on Information Technology: New Generations. USA: IEEE; 2009.
23. Abbas ES, Mohsen E. Factors affecting the continuance intention for using social networks; the role of mediators of trust (Case study: Telegram and Soroush messengers). Glob Media J Persian Edition 2019;14:29-49.
24. Primiero G, Raimondi F, Bottone M, Tagliabue J. Trust and distrust in contradictory information transmission. Appl Netw Sci 2017;2:12.
25. Chang HH, Chuang SS. Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. Inf Manage 2011;48:9-18.
26. Chow WS, Chan LS. Social network, social trust and shared goals in organizational knowledge sharing. Inf Manage 2008;45:458-65.
27. Golsoom K, Fatemeh DH, Hamideh S. Virtual social networks' type and extent of usage a case study of Mashhad Azad University students. Commun Res 2014;21:167-92.
28. Hammann F, Gutmann H, Baumann U, Helma C, Drewe J. Classification of cytochrome *P* (450) activities using machine learning methods. Mol Pharm 2009;6:1920-6.
29. Suenderhauf C, Hammann F, Maunz A, Helma C, Huwyler J. Combinatorial QSAR modeling of human intestinal absorption. Mol Pharm 2011;8:213-24.
30. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9:e1301.
31. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. BMC Res Notes 2011;4:299.
32. Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG Proc Health Care Life Sci Baltimore Maryland 2010;19:67.

Appendix Table 1: Selected questions

Row	Question	Answers				
		Very low	Low	Medium	High	Very high
1	How effective have social networks been in raising your awareness of COVID-19 signs and symptoms?	□	□	□	□	□
2	How effective have social networks been in introducing to you preventive behaviors for COVID-19?	□	□	□	□	□
3	How effective have social networks been in introducing to you high-risk behaviors for COVID-19?	□	□	□	□	□
4	How much have social networks led to you and your family taking care of personal health issues?	□	□	□	□	□
5	How much have social networks led to decreasing how much you and your family commute to public places?	□	□	□	□	□
6	How much have social networks led to reducing the amount you and your family goes to parties?	□	□	□	□	□
7	How much do you follow the tips and training you get from social networks?	□	□	□	□	□
8	How much do you trust the tips and training you get from social networks?	□	□	□	□	□
9	How much contradictory content regarding COVID-19 have you received from social networks?	□	□	□	□	□
10	How much have social networks contributed to informing you about COVID-19 care centers?	□	□	□	□	□
11	How much has the use of social networks led to the spread of rumors about COVID-19?	□	□	□	□	□
12	In your opinion, how reliable is the information about COVID-19 found on social networks?	□	□	□	□	□

COVID-19: Coronavirus disease 2019