Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 13, No. 2, 2024

Association Rule Algorithm-Based Prediction Model for SCL-90 Psychological Measurement: A Comparative Analysis Before and During the COVID-19 Pandemic

Qiang Wang *, Daisy S. Yap

Graduate School, Angeles University Foundation, Angeles City, Philippines

* **Corresponding author**: Qiang Wang (Email: wang.quiang@auf.edu.ph)

Abstract: A prediction model for SCL-90 psychometric data was developed by association rule algorithm. The SCL-90 data reported by college students were used to predict possible psychological symptoms and psychological states by mining the association rules in the data. After the experiment, it is found that there are significant differences in the SCL-90 data in different time periods. As an example, the outbreak of COVID-19 had a significant impact on the mental health of college students in 2019 and 2020.The data in 2019 showed relatively few fixated mental states, while the data in 2020 showed a significant increase in fixated mental states. It can be inferred that this difference is related to the negative impact of the COVID-19 outbreak on mental health, with the COVID-19 outbreak increasing anxiety, which in turn led to the emergence of psychological symp-toms. Based on these findings, it is important to focus not only on physical health but also on changes in mental health in future mental health interventions and prevention efforts. The predictive model derived from this study can inform mental health management and help identify possible psychological symptoms and psychological states and take appropriate inter-ventions.

Keywords: Psychological Measurement; SCL90; COVID-19; Machine Learning; Apriori; Association Rule.

1. Introduction

The development of the mental health of Chinese college students has always been of great concern. With the influence of factors such as increasing social pressure, changes in family structure, and the popularization of the Internet, college students face many mental health problems, such as anxiety, depression, and interpersonal relationship problems. These problems not only affect the learning, life and development of individuals, but also pose a challenge to the healthy development of the education system.

SCL90 (Symptom Checklist-90) is a commonly used mental health assessment tool to assess the mental health status of individuals. It contains 90 items covering 9 dimensions, namely: anxiety, hostility, depression, phobia, paranoia, anxiety neurosis, somatization, obsessive-compulsive disorder, and psychotic symptoms. [1] The assessment of these items provides insights into the status of the assessed person on different mental health dimensions, which can guide mental health interventions and treatments. In this paper, we will predict the status of 9 dimensions in SCL90 through association rule algorithms in machine learning, so as to assist in guiding students to control college students' conditions.

Association rules are a technique in the field of data mining used to discover patterns of association in data. In the field of mental health, association rules can be used to analyze the correlations between the correlation index of SCL90 and other factors. For example, the degree of association between different mental health problems can be explored, or association rules between mental health problems and personal background factors (e.g., gender, age, academic stress, etc.) can be analyzed to help understand the causes and trends of mental health problems. The use of association rules to predict SCL90 correlation indices can help in the following areas: Risk identification and prevention: By analyzing the association rules between mental health problems, it can help to identify what other mental health problems may co-exist with individuals suffering from a certain mental health problem, so as to prevent or intervene in advance. Individualized Intervention: The analysis of association rules can reveal the potential links between different mental health problems, which can help to develop individualized mental health interventions to help college students improve their mental health problems. Policy making and resource allocation: By analyzing the association rules between SCL90 and individual background factors, it can provide a scientific basis for policy making and resource allocation, so as to improve the mental health level of college students and promote the development of education more effectively.[2] SCL90, as a tool for evaluating the mental health of college students, combined with the analysis of the association rules, can help to understand the characteristics and patterns of college students' mental health problems, so as to provide a better understanding of the mental health of college students. characteristics and patterns, so as to provide a scientific basis for mental health education and intervention, and to promote educational development and individual growth. [3]

2. Related Work

2.1. Literature review

Throughout the timeframe spanning from 2020 to 2022, the global response to the COVID-19 pandemic necessitated the implementation of infection containment measures, notably isolation protocols and reduced physical activity. This period witnessed a notable surge in mental health disorders worldwide, particularly among university students, as highlighted by empirical research. The prevalence of adverse emotional states during the enforced containment measures underscores the impact of reduced physical activity and the emergence of unforeseen public health crises on this demographic. Empirical evidence underscores a significant correlation between the levels of physical activity, circadian rhythm variations, and the psychological well-being of university students. Moreover, the disparities in exercise habits among students contribute to varying degrees of academic pressure. Gender-specific discrepancies also manifest concerning the levels of academic and occupational stress experienced by male and female students. [4] The isolation measures instituted during the COVID-19 pandemic have engendered disruptions in psychological adaptation and resilience mechanisms. [5] While activating protective mechanisms such as engagement in leisure activities and social support, these measures concurrently attenuated the capacity for positive cognitive reappraisal. In response to the paradigm shift towards educational digitization, numerous institutions of higher learning have strategically integrated data mining technologies to develop comprehensive systems for managing student mental health. These systems incorporate a diverse array of algorithms tailored for psychological assessment, thereby facilitating counseling, educational interventions, and preventive measures targeting mental health concerns among university students. This concerted effort has introduced innovative strategies and methodologies aimed at addressing the multifaceted dimensions of student psychological well-being. Among the array of algorithms utilized, association rule mining emerges as a preeminent methodology for elucidating intricate relationships within student datasets. [6] By scrutinizing data pertaining to students' lifestyles and academic behaviors, association rule mining facilitates the discernment of latent connections between various attributes, thereby enabling educators to adeptly identify individual student needs and competencies, thus facilitating personalized pedagogical strategies. Likewise, the predictive modeling capabilities of association rules hold promise for forecasting students' psychological well-being, thereby extracting actionable insights to foster enhancements in mental health outcomes.

2.2. Methodology

2.2.1. Date Sources

The year 2019 is the year before the outbreak of the COVID-19 and the year 2020 is the year when the outbreak of the COVID-19 started. The purpose of using the data of 2019 and 2020 is to take advantage of the difference in the objective environment of the two and try to analyze whether the COVID-19 has an impact on certain factors of the college students' psychological state so that the correlation between the relevant psychological factors can be predicted. Now from a university to obtain 2019 and 2020 SCL90 data, 2019 and 2010 data volume are both 3277 entries, respectively, this data to obtain the consent of the university and data prediction, the data is the university's freshman enrollment when the mandatory program, the data is the subjective feelings of the students to present, the data structure of the following table 1 shown in the field:

The 10 psychological states are: degree of somatization, degree of obsessive-compulsive symptoms, degree of interpersonal relationships, degree of depression, degree of anxiety, degree of hostility, degree of terror, degree of stubbornness, degree of psychotic, and degree of other. Each of these 10 psychological states is divided into 5 ratings based on a Likert scale, which are shown in the table 2 below:

Table 1. Database Description

Item Name	Description
Name	The name of student
Gender	The Gender of student
Enrollment Year	Enrollment Year
Somatization	Degree of Somatization
Obsessive-compulsive symptom	Degree of Obsessive-Compulsive Symptoms
Interpersonal relation	Degree of Interpersonal Relationships
Depression	Degree of Depression
Anxiety	Degree of Anxiety
Hostile	Degree of Hostile
Terror	Degree of Terror
Stubborn	Degree of Stubborn
Psychotic	Degree of Psychotic
Other	Degree of Other

Table 2. Description of the degree of psychological state

Ratings	Description
None	Conscious absence of the symptom (problem)
Mild	Consciousness of the symptom, but it has no real effect on the person being tested, or the effect is slight
Middle	Self-perceived symptoms of the item, with some effect on the subject
Serious	Consciousness of the symptom is often present and has a considerable effect on the subject
Extremely Serious	Consciousness of the frequency and intensity of the symptom is so severe that it has a serious impact on the subject

The “impact” here includes the pain and worry caused by the symptoms, as well as the damage to psychosocial functioning caused by the symptoms. The specific definitions of “mild”, “middle” and “serious” are left to the self-evaluator's own experience and are not mandatory.
In order to facilitate the experiments conducted later, the five ratings in this SCL90 survey are therefore expressed as 1 to 5, where None is 1, mild is 2, middle is 3, serious is 4, and extremely serious is 5.

2.2.2. Association Rule Algorithms

Apriori algorithm is one of the most classical and commonly used algorithms in association rule mining, which is used to discover frequent association rules between items in a dataset. In association rule mining, items refer to elements in a dataset and rules refer to association relationships between items. [7] Apriori algorithm is in an important position in the field of association rule mining because it has the following advantages:
Scalability: the Apriori algorithm employs a layer-by-layer search strategy that gradually reduces the search space by generating sets of candidate terms and checking their frequencies. This strategy makes the algorithm able to handle large-scale datasets effectively and has good scalability.
Simple and easy to implement: the basic principle of Apriori algorithm is simple and easy to understand, easy to implement. It does not require complex mathematical theories and algorithms, so it is suitable for a wide range of application scenarios.
Widely applicable: Association rule mining is one of the important tasks in the field of data mining, and Apriori

algorithm is one of the main methods to realize association rule mining. It can be applied to market basket analysis, product recommendation, cross-selling analysis and other fields.

Strong interpretability: the association rules generated by Apriori algorithm have intuitive interpretability, which can help people understand the relationship between different items in the dataset, and thus provide reference for business decisions.[8] Apriori algorithm is widely used because it can effectively discover frequent item sets and association rules in a data set, which can help people understand the association relationship between data, discover potential patterns and trends, and then guide business decisions, marketing and product recommendation.

Next the results will be evaluated in terms of Support, Confidence, and Lift, and the equations for these three-evaluation metrics are shown below. [9]

Support:

Support measures how often an itemset appears in the dataset, i.e., the ratio of the number of occurrences of the itemset to the total number of transactions. For the itemset {X, Y}, its support can be calculated by the following formula:

$$\text{Support }(X \to Y) = \frac{\text{transactions containing } X \text{ and } Y}{\text{total transactions}} \quad (1)$$

Confidence

Confidence measures the trustworthiness of a rule, i.e., the probability of Y occurring if X occurs. For rule X ➔ Y, the confidence level can be calculated by the following formula:

$$Confidence(X \to Y) = \frac{\text{Support}(X \to Y)}{\text{Support}(X)} \quad (2)$$

Lift:

The degree of lift measures the degree of correlation between the terms in a rule, i.e., the ratio of the probability of Y occurring if X is known to the probability of Y occurring if no information is available. The formula for lift is as follows

$$\text{Lift}(X \to Y) = \frac{\text{Support }(X \to Y)}{\text{Support}(X) \times \text{Support}(Y)} \quad (3)$$

These metrics play an important role in association rule mining, supporting the analyst in identifying frequent item sets and valid rules, and helping to help us analyze the correlation of the 10 sub-metrics of the SCL90.

3. Results

To enhance the executable effect of the experimental process, the 10 mental states in SCL90 were represented by short codes as shown in the table 3 below:

Table 3. Simple codes for psychological states

Name	Simple Code
Somatization _ degree	S
Obsessive-compulsive symptom _ degree	O
Interpersonal relation _ degree	I
Depression _ degree	D
Anxiety _ degree	A
Hostile _ degree	H
Terror _ degree	T
Stubborn _ degree	U
Psychotic _ degree	P
Other _degree	E

In order to facilitate the correlation algorithm processing, the data in the dataset needs to be divided into rules, and the 10 psychological state items with no symptoms are assigned a value of 0, and the items with symptoms are assigned a value of 1, forming the results of the 2019 and 2020 data as shown in the table 4,5,6 and 7 below:

Table 4. Data set data values (2019)

ID	S	O	I	D	A	H	T	U	P	E
1	1	1	1	1	1	1	1	2	1	1
2	1	1	1	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	…

Table 5. Preprocessed data values (2019)

ID	S	O	I	D	A	H	T	U	P	E
1	0	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…

Table 6. Data set data values (2020)

ID	S	O	I	D	A	H	T	U	P	E
1	1	2	1	1	1	1	1	1	1	2
2	1	2	2	2	1	2	1	2	1	3
…	…	…	…	…	…	…	…	…	…	…

Table 7. Preprocessed data values (2020)

ID	S	O	I	D	A	H	T	U	P	E
1	0	1	0	0	0	0	0	1	0	1
2	0	1	1	1	0	1	0	1	0	1
…	…	…	…	…	…	…	…	…	…	…

Now for 2019 and 2020 data the Apriori algorithm in the association rule is calculated using Python programming language and the following results are obtained as shown in the table 8,9 and figure 1,2 below:

Table 8. Frequent Itemsets (2019)

	Support	Itemsets
0	0.46	O
1	0.32	I
2	0.21	D
3	0.20	U
4	0.29	(I, O)
5	0.20	(D, O)

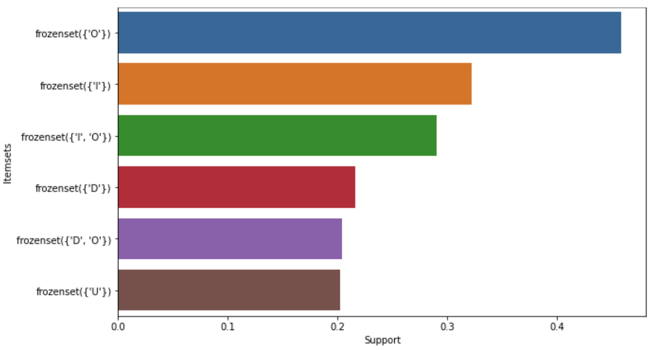

Figure 1. Top 10 Frequent Itemsets (2019)

From the 2 tables and figures above, it can be seen that the year before the outbreak of the COVID-19 in 2019, the results obtained by the apriori algorithm calculations can be seen that in the Frequent Itemsets Degree of Obsessive-Compulsive Symptoms, Degree of Interpersonal Relationships, Degree of Depression, and Degree of Stubborn are 0.46, 0.32, 0.21, and 0.20, while the frequency of symptoms in the Frequent

Itemsets is 0.46, 0.32, 0.21, and 0.20, and the frequency of symptoms in the Degree of Obsessive-Compulsive Symptoms and Degree of Interpersonal Relationships symptoms, and 0.29 and 0.20 for Degree of Obsessive-Compulsive Symptoms and Degree of Depression symptoms, respectively

Table 9. Frequent Itemsets (2020)

	Support	Itemsets
0	0.53	O
1	0.34	I
2	0.23	D
3	0.24	U
4	0.30	(I, O)
5	0.21	(D, O)
6	0.20	(U, O)
7	0.20	(D, I)

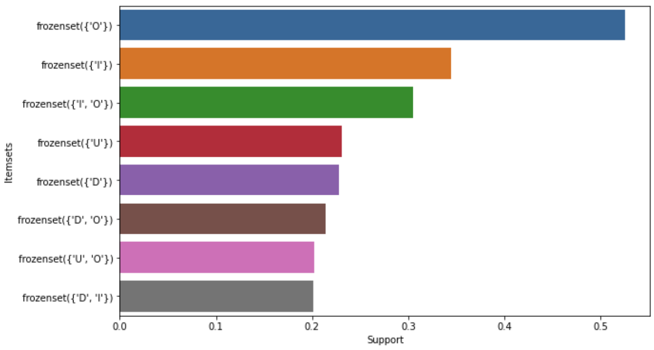

Figure 2. Top 10 Frequent Itemsets (2020)

Compared to the situation in 2019, the outbreak of the COVID-19 in 2020, the results of the calculations show that the frequency of frequency of 0.20 or more appeared frequently single symptoms and combinations of symptoms of a total of seven psychological states, of which the Degree of Obsessive-Compulsive Symptoms and the Degree of Interpersonal Relationships had the highest frequency of 0.52 and 0.34, and the combination of symptoms appeared in four of them, namely Degree of Obsessive-Compulsive Symptoms and Degree of Interpersonal Relationships, Degree of Obsessive-Compulsive Symptoms and Degree of Depression, Degree of Obsessive-Compulsive Symptoms and Degree of Stubborn, Degree of Interpersonal Relationships and Degree of Depression, with values of 0.30, 0.21, 0.20, and 0.20, respectively.

Comparison of the two clearly shows that the values taken for each of the items in 2020 are greater than those in 2019, from breadth to depth. Being attacked by the COVID-19 has a significant impact on the psychological state of college students.

In the table 9 and 10 above you can see that Degree of Obsessive-Compulsive Symptoms and Degree of Interpersonal Relationships are mutually exclusive in the 2019 Association Rule table and both have a support level of 0.20 or higher and a confidence level of 0.90 or higher, indicating that these 2 psychological states are more closely related. In 2020, there are no psychological states that are mutually exclusive in the association rule table, and there are 3 psychological states with Degree of Obsessive-Compulsive Symptoms as the latter term, which means that when Degree of Interpersonal Relationships, Degree of Depression, Degree of Depression, and Degree of Stubborn are closely related to Degree of Obsessive-Compulsive Symptoms, all with a support level above 0.20 and a confidence level above 80%, and it is important to note that Degree of Stubborn is a newly emerged association of psychological states in 2020. Indicator.

Table 9. Association Rules:(2019)

antecedents	consequent	antecedent support	consequent support	support	confidence	lift	leverage	conviction
I	O	0.32	0.46	0.29	0.90	1.97	0.14	5.54
O	I	0.46	0.32	0.29	0.63	1.97	0.14	1.85
D	O	0.21	0.46	0.20	0.94	2.06	0.10	9.59

Table 10. Association Rules:(2020)

antecedents	consequent	antecedent support	consequent support	support	confidence	lift	leverage	conviction
I	O	0.34	0.53	0.30	0.88	1.68	0.12	4.11
D	O	0.23	0.53	0.21	0.93	1.78	0.09	7.58
U	O	0.23	0.53	0.20	0.88	1.66	0.08	3.83
D	I	0.23	0.34	0.20	0.88	2.56	0.12	5.60

4. Conclusion

By analyzing the association rules of SCL-90 (Symptom Self-Rating Scale) data completed by college students in 2019 and 2020, we found that there were some differences between the two years. In 2019, the data completed by freshmen at the time of enrollment showed relatively few stubborn mental states, which could be related to the fact that the COVID-19 had not yet broken out at that time. However, the data from 2020 showed a significant increase in stubborn mental states, which could be interpreted as one of the manifestations of increased anxiety associated with the COVID-19 outbreak.

The COVID-19 outbreak has not only affected physical health, but also negatively impacted mental health, leading to a range of psychological symptoms, including anxiety. Therefore, in addition to paying attention to physical health, we also need to pay attention to the changes in the mental health of college students, especially in the face of the outbreak, and emphasize the intervention and preventive measures for mental health problems.

References

[1] H. Li and Y. Zhang, "Effects of Physical Activity and Circadian Rhythm Differences on the Mental Health of College Students in Schools Closed by COVID-19," IJERPH, vol. 20, no. 1, p. 95, Dec. 2022, doi: 10.3390/ijerph20010095.

[2] L. Yang, "Research on Strategies of Promoting Mental Health of Higher Vocational College Students Based on Data Mining," Wirel. Commun. Mob. Comput., vol. 2022, pp. 1–9, May 2022, doi: 10.1155/2022/3719882.

[3] K. Li, J. Huang, and J. Lin, “The Architecture of College Psychological Teaching Management System Based on Data Mining Technology,” Security Comm. Networks, vol. 2022, pp. 1–12, Aug. 2022, doi: 10.1155/2022/7968756.

[4] H. Kai, “EMPIRICAL ANALYSIS OF THE INFLUENCE OF PHYSICAL EXERCISE ON PSYCHOLOGICAL STRESS OF COLLEGE STUDENTS,” RACP, 2020, doi: 10.24205/03276716.2020.386.

[5] R. Mariani, A. Renzi, C. Di Monte, E. Petrovska, and M. Di Trani, “The Impact of the COVID-19 Pandemic on Primary Emotional Systems and Emotional Regulation,” IJERPH, vol. 18, no. 11, p. 5742, May 2021, doi: 10.3390/ijerph18115742.

[6] C. De, Psicológica, and C. Bi, “ASSOCIATION MINING BETWEEN IDEOLOGICAL AND POLITICAL TEACHING AND MENTAL HEALTH,” RACP, 2020, doi: 10.24205/03276716.2020.147.

[7] F. Gao, J. Yang, Y. Yang, and X. Yuan, “WOMDI-Apriori Data Mining Algorithm for Clustered Indicators Analysis of Specialty Groups in Higher Vocational Colleges,” INT J COMPUT COMMUN, Int. J. Comput. Commun. Control, vol. 18, no. 3, May 2023, doi: 10.15837/ijccc.2023.3.5045.

[8] M.-H. Tseng and H.-C. Wu, “Investigating Health Equity and Healthcare Needs among Immigrant Women Using the Association Rule Mining Method,” Healthcare, vol. 9, no. 2, p. 195, Feb. 2021, doi: 10.3390/healthcare9020195.

[9] U. M. Haque, E. Kabir, and R. Khanam, “Investigating school absenteeism and refusal among Australian children and adolescents using Apriori association rule mining,” Sci Rep, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-024-51230-4.