Received: 18 January 2023 | Revised: 21 February 2023 | Accepted: 8 March 2023 | Published online: 8 March 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2023, Vol. 1(1) 45–54
DOI: 10.47852/bonviewJDSIS3202669

BON VIEW PUBLISHING

ARM for Analyzing Factors Influencing Vaccinations During the COVID-19 Outbreak

Wipawan Buathong[1] and **Pita Jarupunphol**[1,*]

[1]*Department of Digital Technology, Phuket Rajabhat University, Thailand*

Abstract: This article investigates factors influencing Coronavirus 2019 (COVID-19) vaccinations and public concerns using association rule mining (ARM). The experiment was conducted in Phuket at the beginning of 2022 when many people were still not vaccinated. The online questionnaire comprises 26 questions with a sample of 403 participants living in Phuket, Thailand. Among the participants, 377 people were vaccinated. The experiment is based on two main ARM methods: Apriori and FP-Growth. The performance assessment criteria were 0.5 and 0.6 minimum support and 0.9 and 1.0 minimum confidence. The most appropriate values are based on 0.5 minimum support and 1.0 minimum confidence with lift > 1. The experimental results reveal factors influencing COVID-19 vaccinations in association rules. For example, most vaccinated people received Sinovac. They are female and knowledgeable about vaccines. They are concerned about severe illness or death and live in the Mueang district more than others. They are also concerned about injection site pain, nausea, vomiting, numbness, and hemiplegia. Enhancing herd immunity is another compelling reason for people to be vaccinated. In the meantime, these derived rules can be reversed to explain why many people are not vaccinated. For instance, the most available vaccine is Sinovac, and several vaccination points are in the Mueang district.

Keywords: Apriori, association rule mining (ARM), COVID-19, data mining, FP-Growth, vaccination

Foreword

Association rule mining is a well-recognized data mining technique for analyzing large datasets and identifying correlated factors with single or multiple attributes. It can handle nonlinear relationships and identify relationships between multiple variables, making it well-suited for analyzing complex data. Additionally, association rule mining can identify high-confidence rules that can be used to guide decision-making. In this case, the correlated factors can establish rules for explaining why some people agree, and some hesitate to be vaccinated during the COVID-19 outbreak between 2021 and 2022.

1. Introduction

The Coronavirus 2019 (COVID-19) outbreak commenced at the end of 2019 and spread worldwide at the beginning of January 2020 (Sohrabi et al., 2020). The virus affected social and economic health and caused fear among the population. According to the world's epidemic situation and infection of COVID-19 statistics of August 30, 2021, the global population has 217,939,497 infections and 4,524,239 deaths (WHO, n.d.). During that time, there were 1,190,066 infections and 11,399 deaths in Thailand. The first COVID-19 case was detected on January 22, 2020, when the number of infected cases slightly increased at the beginning and rapidly increased at an average rate of 25.1% per day after the mid-March of the same year (Tantrakarnapa & Bhopdhornangkul, 2020). The second round of outbreaks in Thailand started in late December 2020 and was perceived to be from the shrimp market in Samut Sakhon province, where migrant workers are densely inhabited. Samut Sakhon health officials visited the site and found many new asymptomatic patients among the workers. The second wave of the outbreak differs from the first wave, in which more infected people spread to many provinces. The COVID-19 pandemic in Thailand continued until the third wave of outbreaks, affecting several lives and a large population in distress.

In response to the situation, the COVID-19 management committee was established with the Prime Minister as the chairman, who has declared a state of emergency in Thailand and regulated several guidelines for disease control. For instance, infected patients or at risk of infection must be quickly isolated. People's movements must be monitored and restricted so that those exposed to the condition and the infection from the patient can be promptly checked. They are required to stay and work from home to maintain physical distance (physical distancing) and social distance (social distancing). They are also forced to wear masks and guided to wash their hands frequently. The emergency declaration has forced people to change their lifestyles. Family members must be separated because of the social distancing policy. COVID-19 has inevitably caused social and economic transformations affecting the way of life of many individuals. Therefore, various technologies must be used to manage such a

*__Corresponding author:__ Pita Jarupunphol, Department of Digital Technology, Phuket Rajabhat University, Thailand. Email: p.jarupunphol@pkru.ac.th

life. The way of life and community relationships must also be transformed to depend upon digital technology.

Besides several innovations proposed to facilitate the COVID-19 risk mitigation process (e.g., temperature detection, mask facial recognition, and vaccination tracking), digital technology can be utilized to deal with COVID-19 issues by presenting the information more appropriately and understandably. For example, Dong et al. (2020) developed an online interactive dashboard system to track COVID-19 in real time. The Center for Systems Science and Engineering at Johns Hopkins University hosted the system to visualize and track reported COVID-19 cases in real time.

Although various COVID-19 cases have been reported and sensationalized by different media channels every day, the public vaccination rate is still unsatisfactory, given that vaccination is an effective process of triggering immunity against the spread of coronavirus. Moreover, different countries launched incentive campaigns for people to get the vaccine. For example, tourist vouchers and car sweepstakes were offered to encourage people to vaccinate in Gunma Prefecture, Japan (Nagase, 2021). In Washington, D.C., United States, young people are encouraged to get vaccinated with souvenirs such as AirPods, GiftCards, and $1000 scholarships (Vakil, 2021). Nevertheless, the number of vaccinations that increased daily was still relatively small. For example, as of August 28, 2021, Thailand's immunization data in 65,228,120 people in Thailand, 34.2% of the population received the first dose, and 10.9% of the total population received the second dose.

Furthermore, according to the National COVID-19 statistics compiled by the Ministry of Higher Education, Science, Research, and Innovation (MHESI, 2021), more than 60% of Thai people have not been vaccinated. Based on these statistics, some people participated in the COVID-19 vaccination, while some are very concerned about the vaccination and feel uncomfortable being vaccinated. According to COVID-19 vaccination information from the Phuket office of the Public Relations Department, as reported in the Phuket News (The Phuket News, 2021), about 338,980 people or 74% of Phuket's population have been vaccinated. The Center for Disease Control of the Coronavirus Disease 2019 (OCC) categorized Phuket into a low-risk group as there are a few thousand cases and less than 30 cumulative deaths in Phuket. On June 22, 2021, the government approved the campaign of opening a pilot area for tourists, namely "Phuket Sandbox," welcoming tourists from low- to medium-risk countries who had been vaccinated with the required number of doses certified by the World Health Organization and passed the COVID-19 testing process within 72 hours before departure to Phuket without quarantine. This campaign aims to stimulate the provincial economy from July 1, 2021. The province must prepare for the city's opening for tourists and public health measures. In the meantime, Phuket also needs to promote and encourage people to get enough vaccination to create herd community immunity to prevent infection and spread in Phuket. In this sense, understanding the factors underlying people's decisions toward COVID-19 vaccinations in Phuket is required. Several types of research have identified factors influencing public attitudes regarding COVID-19 vaccinations (Guillon & Kergall, 2021; Mir et al., 2021; Potdar et al., 2021; Toth-Manikowski et al., 2022). Despite many plausible techniques for analyzing such factors, this article aims to apply association rule mining (ARM), a well-known data mining technique, to understand the correlating factors driving and restricting people from getting vaccinated.

2. Literature Review

This section discusses topics related to this research, including data mining and its applications and factors influencing people's attitudes toward COVID-19 vaccinations.

2.1. Data mining applications

Data mining is recognized as an efficient technique for identifying and selecting a data sample or "data dimension" from a considerable amount of data (Bhatia, 2019). It is one of the most outstanding data analysis methods as it acquires new knowledge from big data (Buathong & Jarupunphol, 2021). There are several data mining applications categorized into specific perspectives. For instance, terrorist network mining is proposed as a workable solution for terrorist detection and analysis, which can be used to reveal trends whose action is intrusive (Sharma, 2018). The efficiency of this theoretical proposal could be improved by connecting fuzzy with a genetic algorithm with an intrusion detection system, leading to an efficient detection process. In machine learning modeling, dimensionality reduction has long been perceived as an effective data preparation process of data mining when selecting data dimensions from extensive data confronts several understandability issues. There are two main perspectives of dimensionality reduction: 1) feature selection (irrelevant or unnecessary data features will be reduced, and only key data features will be selected) and 2) feature extraction (the number of data items representing a dataset will be reduced or discarded and new data features will be created). Several research works have been on optimizations and applications of feature selection and feature extraction (Buathong & Jarupunphol, 2018; Manek et al., 2017; Miao & Niu, 2016; Tumthong et al., 2021).

ARM is another prominent field of data mining that can be used to identify associations of data features. For example, Soni et al. (2017) applied ARM to analyze e-commerce algorithms and identify the algorithm weaknesses. Based on the experimental results, the authors proposed potential characteristics for designing an efficient e-commerce database algorithm to support incremental and interactive ARM. In Rojanavasu (2019), there were experiments on two data mining techniques in education, including ARM and a decision tree. ARM was applied to identify the association of information to facilitate admission preparation. In the meantime, the decision tree was used to forecast student employment after graduation. The author claimed that the research findings offer vital insights into intake preparation and work forecasting

2.2. Association rule mining

ARM is an unsupervised nonlinear algorithm to analyze interitem associations in a transaction from systematic mining processes. It is a principal data mining method for finding relationships among features in big data. Frequently occurring relationships can be identified as patterns, which can later be used in correlation analysis or predict various phenomena. For example, recommendations on e-commerce websites are generated from frequently purchased items in online databases. The association rules are written in a set of causal items (premise) to a collection of consequent items (conclusion). According to Katragadda et al. (2021), the association rule may be denoted by $A \Rightarrow B$, which are sets of various attribute-value pairs (itemsets), where A is the antecedent or the left-hand side (LHS) and B is the consequent or the right-hand side (RHS). The rule represents the hypothesis that when variables in A occur in the dataset, the variables in B also occur. The most frequently used database for ARM is based on a transaction database, and the data type can be in nominal or ordinal format only. Apriori and FP-Growth are practical ARM algorithms (Agrawal & Srikant, 1994; Vijayarani & Sharmila, 2016). Although the Apriori and FP-Growth work similarly, the FP-Growth is more effective than the Apriori (Arora et al., 2013). The FP-Growth is introduced to address Apriori flaws by sorting the frequency values in descending order before delivering them into the tree to find the correlation rule according to the specified criteria (Witten et al., 2017). There are measurement criteria for ARM, including support, confidence, and lift. The rules must have support and confidence not less than the minimum specified value.

2.2.1. Support

The "support" is the percentage of the total number of itemsets occurring in the database. It represents the item popularity, which can be described in (1), where *P(A)* is the percentage of cases containing *A*.

$$support(A \Rightarrow B) = support(A \cup B) = P(A \wedge B) \tag{1}$$

2.2.2. Confidence

Confidence indicates the likelihood of item (*B*) being purchased when one item is purchased (*A*). In other words, it represents a causal relationship from *A* to *B*, which can be expressed as $A \Rightarrow B$. The confidence is the percentage of the total number of itemsets in the database per the number of itemsets appearing on the LHS, which can be explained in 2.

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{P(A \wedge B)}{P(A)} \tag{2}$$

2.2.3. Lift

Lift represents the likelihood of item (B) being purchased when one item is purchased (A) while the popularity of B is under control. The lift value considers how often the RHS is discovered together with confidence. The value expresses the probability that the RHS is selected as a segment after the LHS has been chosen based on the rules: 1) If lift < 1, there is no increase in the probability of the rule that the RHS data occur with the existing LHS data; 2) if lift = 1, the LHS and RHS datasets are independent or not correlated; and 3) if lift > 1, there is an increase of the probability of the rule that the LHS data occur with the existing LHS data. This rule, in other words, enforces that the LHS and RHS datasets are correlated or mutually reinforcing. The larger the lift value, the greater the probability that the RHS data occur along with the existing LHS data (Bhatia, 2019). Nowadays, ARM techniques have been utilized in various disciplines, such as telecommunication networks, risk and market management, inventory control, and medical diagnosis/drug tests.

$$lift(A \Rightarrow B) = \frac{\text{confidence}(A \Rightarrow B)}{P(B)} = \frac{P(A \wedge B)}{P(A)P(B)} \tag{3}$$

2.3. Factors influencing COVID-19 vaccination

There are several works on analyzing factors influencing COVID-19 vaccination. For example, Guillon and Kergall (2021) investigate factors associated with attitudes and intentions toward COVID-19 vaccination using ordered and multinomial logistic regressions in France. The experiment was conducted between 20th and 23rd November 2020 on a sample of 1146 participants representing the France population. The results indicated that risk perceptions toward COVID-19 vaccination are associated with intentions and attitudes. While 30.5% of participants agreed to be vaccinated in the first semester of 2021, 31.1% were unsure of their decisions. In addition, some participants perceived that COVID-19 vaccination is riskier to health than the virus. The authors suggested that enhancing vaccination uptake requires effective COVID-19 awareness campaigns.

Furthermore, Mir et al. (2021) analyzed factors influencing people's attitudes toward COVID-19 vaccination and their intentions to take up COVID-19 vaccinations. Variables associated with COVID-19 vaccinations, including perceived benefits, risk perceptions, social media exposure, social norms, and trust, were assessed using structural equation modeling. The authors discovered that the perceived benefits, social norms, and trust correlated significantly with people's attitudes toward COVID-19 vaccinations. Moreover, these three factors also strongly correlate with their intentions toward COVID-19 vaccinations. On the other hand, risk perceptions and social media exposure represented insignificant influences on people's attitudes and intentions toward COVID-19 vaccinations.

In Toth-Manikowski et al. (2022), the reasons behind the COVID-19 vaccination hesitancy of 1974 healthcare workers in the Chicagoland area were investigated using the Health Belief Model framework between March and May 2021. The results show that 85% of the workers expected COVID-19 vaccination, and they tended to receive the vaccines if there were convinced by their colleagues that the vaccination was essential. However, there are some influences related to political and health aspects restricting their decisions not receiving the vaccines. According to Potdar et al. (2021), India launched the major COVID-19 vaccination operation from 3006 sites in January 2021 to 21,872 sites in August 2021. However, the vaccination was not compulsory and unable to force people to comply with the operation. In this case, the authors experimented on gender disparities toward COVID-19 vaccination in Maharashtra States, India, using Pearson's Chi-square. The result illustrated that men were more aware of COVID-19 vaccination than women as the gender vaccination ratio is lower than India's gender ratio. Therefore, the authors suggested that an awareness campaign toward the importance of vaccination should be operated to increase the COVID-19 vaccination statistics.

While there are other methods for studying factors influencing COVID vaccination, ARM has its strengths and advantages that make it a valuable tool for this type of analysis. For example, Celik (2021) analyzed the patient's COVID-19 infection symptoms to provide rapid initiation treatment using ARM, based on SETM, AIS, and Apriori algorithms. The results showed that the Apriori could diagnose the COVID-19 infection symptoms of infected patients with 100% accuracy. Accordingly, Tandan et al. (2021) studied COVID-19 cases from Wolfram Data Repository and applied association mining techniques, including Apriori and FP-Growth, to identify common symptoms and shape patterns in the discovered rules. The most common COVID-19 symptoms of the patients include fever, cough, pneumonia, and sore throat. The rules for symptoms vary by gender and age. The condition of patients with chronic disease and death will be severe if the patients have diseases related to cardiovascular disease followed by pneumonia, fever, and cough, respectively.

3. Research Methodology

Before proceeding to the data mining process consisting of several steps to obtain knowledge, the researchers realize the importance of relevant legal aspects by collecting information by ethical principles and following the protection of personal information. Therefore, the research topic, data collection tools, and methods were submitted to the Human Research Ethics Committee of Phuket Rajabhat University and certified from July 15, 2021, to July 14, 2021. In this case, there was 1 year to carry out data collection under regulations, such as explaining the purpose to the respondents, data storage period, data confidentiality methods, and the right of the respondents to cancel the survey at any time if they feel uncomfortable. After collecting the data, we perform several steps to obtain the knowledge, including data selection, data preprocessing, data mining, data interpretation, and knowledge discovery. Figure 1 represents a process of searching for the body of knowledge.

3.1. Data selection

Data selection is an initial process involving inquiry design and data collection to identify the data sources for data mining. This process includes filtering the desired data from the database to create a data group for initial consideration. This research randomly collected a sample of 445 people living in Phuket via online questionnaires from July to August 2021. The questionnaire contains 30 questions ranging from demographics to participant concerns about COVID-19 vaccinations. At this stage, the data comprise 30 features and 445 instances.

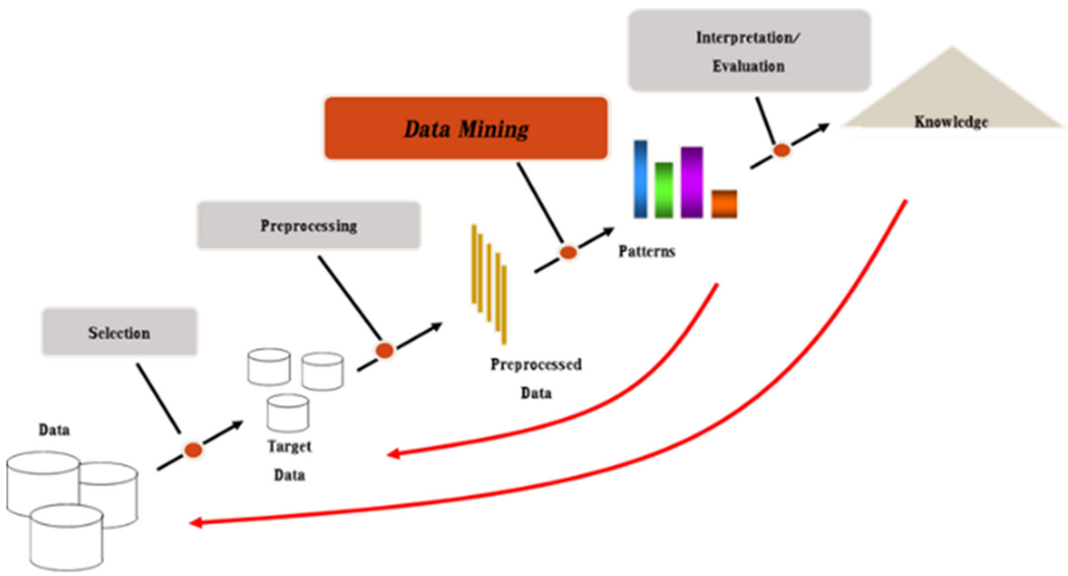
3.2. Data preprocessing

Data preprocessing is a crucial stage of data preparation containing essential data preparation steps, including data cleansing and data transformation. In data cleansing, incomplete or missing data are filtered out, and unimportant data features are eliminated to obtain the relevant features for further processing. This stage handles corrections and modifications of various data types and organizes them in the same standard format, suitable for application to algorithms and models for data mining. During the preprocessing phase, the data had to be translated from Thai to English. Some records were omitted as a result of a significant amount of missing or incomplete information within those records. Besides, some data values from the stored query were transformed into variables for ease of retrieval, analysis, processing, and interpretation. After that, data transformation converts the cleansed data from a query to a format that can be imported into WEKA and used with ARM techniques, such as transforming numerical (discrete/continuous) data to categorical (nominal/ordinal) data. The number of data after data preparation consists of 26 features, 403 instances (cleaned data) with baseline data: a total of 403 participants (164 males, 239 females), 377 vaccinated participants (148 males, 229 females), 26 unvaccinated participants, and 377 types of vaccine received (345 Sinovac, 32 AstraZeneca). Please note that the elimination of missing data did not impact the results and interpretation of the study, as the sample size of 377 vaccinated volunteers is considered representative of the population of Phuket, as determined through the application of Taro Yamane's formula.

3.3. Data mining

After the data have been transformed, data mining is conducted to search for valuable patterns using various techniques. Data mining is an essential step in knowledge discovery in databases (Kaur & Madan, 2015) applied in various fields to discover knowledge from massive data in databases. As this research aims to identify correlated features, two ARM algorithms, Apriori and FP-Growth, are mainly used in this data mining stage. The measurement criteria for the algorithmic efficiency are 1) minimum support at 0.5 and 0.6, 2) minimum confidence at 0.9 and 1.0, and 3) lift > 1.0. The threshold for support and confidence was determined based on values that were deemed high but not necessarily the highest. In all cases, a comprehensive correlation rule was established, with support values ranging from 0.1 to 1.0 and confidence values ranging from 0.1 to 1.0.

Figure 1
The search process for knowledge, adapted from Fayyad et al. (1996)



3.4. Data interpretation and evaluation

The results are analyzed and interpreted to determine whether the results are appropriate and consistent with the objective. For instance, minimum support at 0.6 indicates that the confidence of this research dataset is equal to 60% of the total data. In the meantime, minimum confidence at 1.0 means that the confidence in the required association rules is similar to that of association rules for 100% population. These measurement values can ensure that the association rules will be as accurate as possible. The knowledge is derived from this stage.

3.5. Knowledge discovery

This stage identifies and utilizes the body of discovered knowledge. Data attributes are discovered as knowledge such as most types of injected vaccines, gender, occupation, education level, place of residence who have been vaccinated, the reason for getting vaccinated, and concerns after vaccination. The discovered knowledge is further visualized using R to represent the rules associated with the vaccination. In this case, several insights can be obtained from the rules to answer why many people are still not vaccinated.

4. Result and Discussion

The experiment applies two ARM algorithms, Apriori and FP-Growth, with minimum support performance measurement criteria at 0.5 and 0.6 and minimum confidence at 0.9 and 1.0. After the data preparation, there are 26 features, 403 instances (164 males, 239 females), 377 vaccinated participants (148 males, 229 females), 26 unvaccinated participants, and two types of vaccines (345 Sinovac vaccinations, 32 AstraZeneca vaccinations). According to Table 1, the preprocessed data have been performed data wrangling to allow entry into the following data mining steps.

1. Filter out incomplete or missing data by eliminating 42 instances.
2. Eliminate six features that are insignificant and numeric.
3. Modify some of the data to be suitable for processing.
4. Correct the data, such as correcting the invalid data to an unknown value with the symbol "?"
5. Change data values from numerical (discrete/continuous) to categorical (nominal/ordinal).

After the above steps, preprocessed data are replaced with variables, as shown in Table 2. Finally, the examples of finished results are displayed in Table 3.

Only vaccinated participants and other relevant factors, including the reason for vaccination and concerns after vaccination, are considered to determine the association rules. Tables 4–7 show the association rules based on the confidence of the data (minimum support) and the confidence of the correlation rule (minimum confidence) according to the specified values.

Table 1
Unpreprocessed data

1	เพศ	อาย	อาชีพ	ระดับก	ที่อยู่	ท่านมีค	ท่านได	ชนิดขอ
2	ชาย	ต่ำกว่า 25 ปี	อื่น ๆ โปรดระบุ	ปริญญาเอก	อำเภอถลาง	ไม่มี	ไม่ฉีด	
3	ชาย	41 - 60 ปี	Housekeeper	ปริญญาตรี	อำเภอเมือง	มี	ไม่ฉีด	
4	หญิง	ต่ำกว่า 25 ปี	รับราชการ	ต่ำกว่าระดับปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
5	หญิง	41 - 60 ปี	ลูกจ้างหน่วยงานต่าง ๆ	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
6	หญิง	25 - 40 ปี	รับราชการ	ปริญญาตรี	อำเภอถลาง	มี	ฉีด	ซิโนแวค (Sinovac)
7	หญิง	41 - 60 ปี	รับราชการ	ปริญญาโท	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
8	ชาย	41 - 60 ปี	ลูกจ้างหน่วยงานต่าง ๆ	ปริญญาโท	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
9	หญิง	25 - 40 ปี	พนักงานราชการ	ปริญญาตรี	อำเภอเมือง	ไม่มี	ฉีด	ซิโนแวค (Sinovac)
10	หญิง	25 - 40 ปี	ลูกจ้างหน่วยงานต่าง ๆ	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
11	หญิง	25 - 40 ปี	ลูกจ้างหน่วยงานต่าง ๆ	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
12	ชาย	25 - 40 ปี	รับราชการ	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
13	หญิง	25 - 40 ปี	รับราชการ	ปริญญาโท	อำเภอถลาง	มี	ฉีด	ซิโนแวค (Sinovac)
14	หญิง	41 - 60 ปี	รับราชการ	ปริญญาตรี	อำเภอเมือง	มี	ไม่ฉีด	
15	หญิง	25 - 40 ปี	ลูกจ้างหน่วยงานต่าง ๆ	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
16	หญิง	61 - 75 ปี	ข้าราชการบำนาญ	ปริญญาโท	อำเภอกะทู้	มี	ฉีด	แอสตร้าซิเนก้า (Astr
17	หญิง	ต่ำกว่า 25 ปี	นักเรียน นักศึกษา	ปริญญาตรี	อำเภอเมือง	มี	ไม่ฉีด	
18	หญิง	61 - 75 ปี	ตัวแทน	ปริญญาตรี	อำเภอเมือง	ไม่มี	ฉีด	แอสตร้าซิเนก้า (Astr
19	หญิง	41 - 60 ปี	บำนาญ	ปริญญาเอก	อำเภอเมือง	มี	ไม่ฉีด	
20	หญิง	61 - 75 ปี	ค้าขาย	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	แอสตร้าซิเนก้า (Astr
21	หญิง	ต่ำกว่า 25 ปี	นักเรียน	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
22	หญิง	41 - 60 ปี	ค้าขาย	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
23	หญิง	61 - 75 ปี	ทำงานเอกชน	ต่ำกว่าระดับปริญญาตรี	อำเภอถลาง	มี	ฉีด	แอสตร้าซิเนก้า (Astr
24	หญิง	ต่ำกว่า 25 ปี	นักศึกษา	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
25	หญิง	ต่ำกว่า 25 ปี	นักศึกษา	ปริญญาตรี	อำเภอเมือง	มี	ไม่ฉีด	
26	หญิง	ต่ำกว่า 25 ปี	นักศึกษา	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
27	หญิง	61 - 75 ปี	เกษียณ	ปริญญาโท	อำเภอเมือง	มี	ฉีด	แอสตร้าซิเนก้า (Astr
28	ชาย	ต่ำกว่า 25 ปี	นักศึกษา	ต่ำกว่าระดับปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
29	หญิง	ต่ำกว่า 25 ปี	นศพ	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)
30	ชาย	ต่ำกว่า 25 ปี	นักศึกษา	ปริญญาตรี	อำเภอกะทู้	มี	ฉีด	ซิโนแวค (Sinovac)
31	ชาย	41 - 60 ปี	ทำงานเอกชน	ปริญญาตรี	อำเภอเมือง	มี	ฉีด	ซิโนแวค (Sinovac)

Table 2
Variable representations to be applied to data mining algorithms and models

No.	Question	Answers and variables
1	Sex	Male, Female
2	Career	Government/State Enterprises: G.O., Private: PR, Trading/Business Owner: SA, General Employer: G.E., Farmers: F.A., Agency Employee: OF, Student/Student: S.T., Government Pensioners: P.E., Others, please specify: O.T.
3	Age	Under 25 years: A, 25–40 years: B, 41–60 years: C, 61–75 years: D, 75 years and over: E
4	Education level	Below Bachelor's Degree: C.E., Bachelor's Degree: BA, Master's Degree: M.S., Doctoral Degree: D.R.
5	Address	Mueang: MU, Kathu: KT, Thalang: TA
6	Knowledge of vaccines: KoV	Yes, No
7	Have you been vaccinated?	Yes, No
8	Type of vaccine: ToV	Sinovac: SV, AstraZeneca: AZ
9	Reasons for being vaccinated: RfV	Confident in vaccines: CV, Follow influencers: F.I., Confident in the government: C.G., Prevent severe illness or death: DE, Herd immunity: H.I., Facilitated conditions for vaccination: F.C., Public health standard: S.T., Phuket tourism operation: C.O., Travel in and out of Phuket: T.R., Others: O.T.
10	Concerns after vaccination: AfV	Allergic to vaccines: A.V., Mutation: M.U., Health: HE, No concerns: N.C., Others: O.T.
11	Reasons for not being vaccinated: RNV	Not confident in the vaccine: NCV, Not Confident in the government: NCG, Inadequate facilitated conditions for vaccination: IFC, Congenital disease: CD, Side effects: S.E., Others: OTH

Table 3
Examples of the preprocessed data

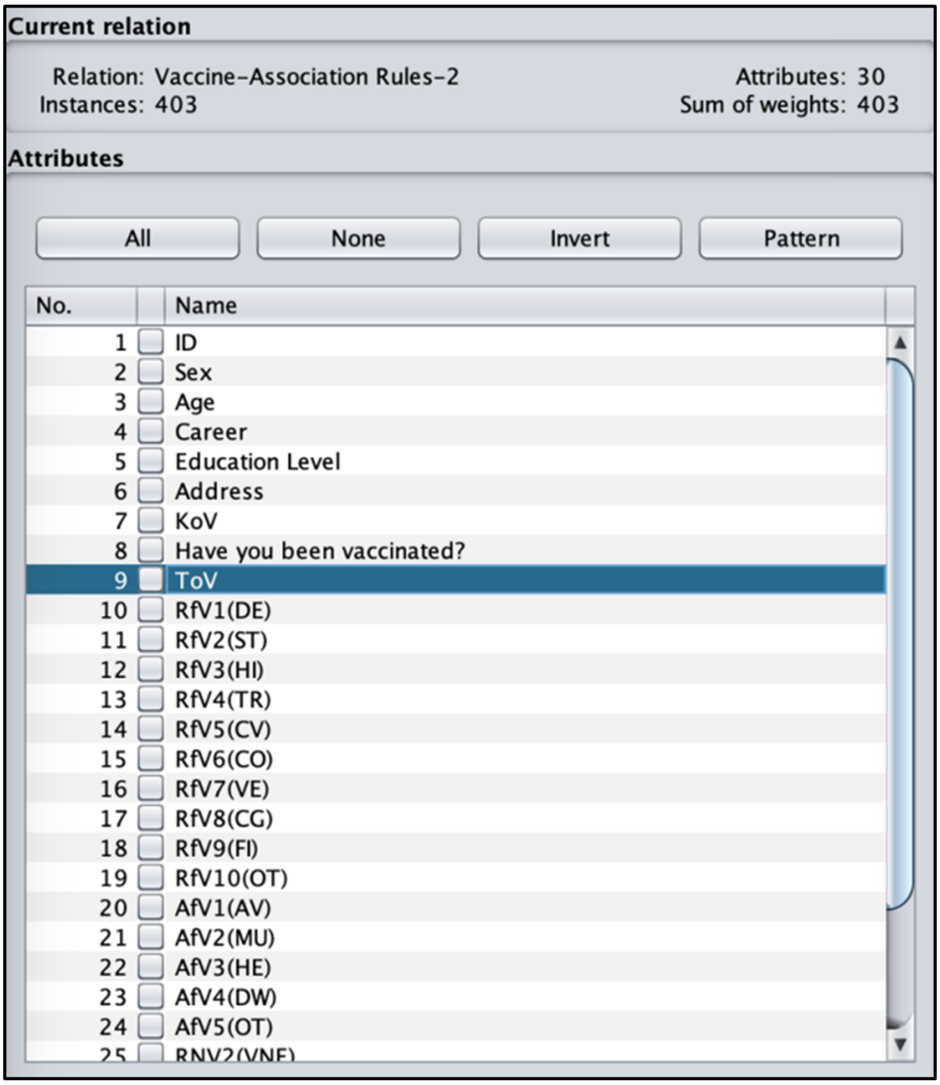


Table 4
The result of the association rules with minimum support = 0.6, minimum confidence = 1.0, and lift >1

No.	Association rules	Support	Confidence	Lift
1	ToV = SV → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
2	KoV = Yes and ToV = SV → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
3	RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
4	Address = MU and ToV = SV → Have you been vaccinated = Yes	>= 0.60	1.00	1.07

Table 5
The result of the association rules with minimum support = 0.5, minimum confidence = 1.0, and lift >1

No.	Association rules	Support	Confidence	Lift
1	ToV = S.V. → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
2	KoV = Yes and ToV = SV → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
3	RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
4	Address = MU and ToV = SV → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
5	ToV = SV and RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
6	ToV = SV and AfV1(AV) = Yes → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
7	Sex = Female and ToV = SV → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
8	KoV = Yes and RfV1(DE) → Have you been vaccinated = Yes	>= 0.50	1.00	1.07

Table 6
The result of the correlation rules with minimum support = 0.6, minimum confidence = 0.9, and lift >1

No.	Association rules	Support	Confidence	Lift
1	ToV = S.V. → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
2	KoV = Yes and ToV = SV → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
3	RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
4	Address = MU and ToV = SV → Have you been vaccinated = Yes	>= 0.60	1.00	1.07
5	ToV = SV and RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.60	0.94	1.07
6	KoV = Yes → Have you been vaccinated = Yes	>= 0.60	0.94	1.07
7	Address = M.U. → Have you been vaccinated = Yes	>= 0.60	0.93	1.07
8	KoV = Yes and Have you been vaccinated = Yes → ToV = SV	>= 0.60	0.93	1.07
9	Have you been vaccinated = Yes → ToV = S.V.	>= 0.60	0.92	1.07

Table 7
The result of the association rules with minimum support = 0.5, minimum confidence = 0.9, and lift >1

No.	Association rules	Support	Confidence	Lift
1	ToV = S.V. → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
2	KoV = Yes and ToV = SV → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
3	RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
4	Address = MU and ToV = SV → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
5	ToV = SV & RfV1(DE) = Yes → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
6	ToV = SV and AfV1(AV) = Yes → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
7	Sex = Female and ToV = SV → Have you been vaccinated = Yes	>= 0.50	1.00	1.07
8	KoV = Yes and Have you been vaccinated = Yes → ToV = SV	>= 0.50	1.00	1.07
9	AfV1(AV) = Yes → Have you been vaccinated = Yes	>= 0.50	0.99	1.06
10	Sex = Female → Have you been vaccinated = Yes	>= 0.50	0.95	1.01
11	Address = MU and Have you been vaccinated = Yes → ToV = SV	>= 0.50	0.94	1.10
12	Sex = Female and Have you been vaccinated = Yes → ToV = SV	>= 0.50	0.93	1.09
13	Have you been vaccinated = Yes and AfV1(AV) = Yes → ToV = SV	>= 0.50	0.92	1.07
14	KoV = Yes and Have you been vaccinated = Yes → ToV = SV	>= 0.50	0.92	1.07
15	Have you been vaccinated = Yes → ToV = S.V.	>= 0.50	0.92	1.07
16	RfV1(DE) = Yes → ToV = SV	>= 0.50	0.91	1.06
17	Have you been vaccinated = Yes and RfV1(DE) = Yes → ToV = SV	>= 0.50	0.91	1.06
18	RfV1(DE) = Yes → Have you been vaccinated = Yes and ToV = SV	>= 0.50	0.91	1.06
19	AfV1(AV) = Yes → ToV = SV	>= 0.50	0.91	1.06
20	AfV1(AV) = Yes → Have you been vaccinated = Yes and ToV = SV	>= 0.50	0.91	1.06

According to Table 4, four association rules were obtained from the experiment, which can be explained below.

1. People opt for vaccination as the vaccine type is Sinovac.
2. People choose to be vaccinated due to their understanding of vaccines and the vaccine type being Sinovac.
3. People get vaccinated out of concern for potentially severe illness or death from the virus.
4. People in the Mueang district choose to be vaccinated, with the vaccine type being Sinovac.

Table 5 shows eight association rules obtained from the experiments on 0.5 minimum support and 1.0 minimum confidence, further explained by rules 5 to 8 below.

5. People choose to get vaccinated due to worries about severe illness or death from the virus, and the vaccine type is Sinovac.
6. People opt for vaccination due to worries about potential side effects such as injection site pain, nausea, vomiting, numbness, and hemiplegia, and the vaccine type is Sinovac.
7. People get vaccinated as they are female, and the vaccine type is Sinovac.
8. People decide to get vaccinated as they are knowledgeable about vaccines and worried about severe illness or death from the virus.

Table 6 shows nine association rules were derived from using 0.6 minimum support and 0.9 minimum confidence. Again, there are repetitive rules in Tables 4–5, and a specific rule derived from rule no. 8 is "Participants are vaccinated because they are knowledgeable about vaccines and the vaccine type is Sinovac."

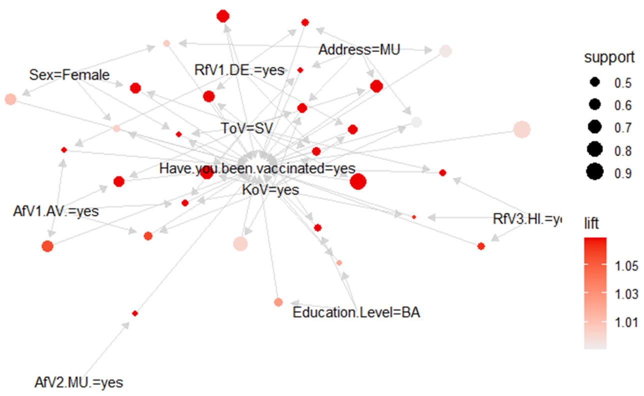
Table 7 shows the results using 0.5 minimum support and 0.9 minimum confidence. There are 20 association rules, and some do not directly associate with vaccinations. In addition, as mentioned previously, there are several rules, many of which are repetitive in Tables 4–6. The researchers performed data screening on 445 questionnaires and obtained 403 complete questionnaires with 26 attributes. There are correlations among the attributes after ARM techniques using Apriori and FP-Growth algorithms have been experimented with and measured using 0.5 and 0.6 minimum supports and 0.9 and 1.0 minimum confidences. Table 4 represents four correlation rules derived from the experimental results based on 0.6 minimum support and 1.0 minimum confidence.

Based on Tables 4–6, frequent factors accompanying vaccination appear in all association rules. Besides, Figure 5 shows that although there are differences in the number of rules, some are related and duplicate. However, we consider the rules of the vaccinated sample only (vaccination = yes) because other relevant factors were also considered, including reasons for vaccination and concerns after vaccination. Figure 2 summarizes outstanding factors influencing COVID-19 vaccinations derived from ARM experiments on different supports.

The experiments show that the researchers intended to reduce the minimum support and minimum confidence values by 0.1 to show the critical factors descendingly. Moreover, the presence of lift > 1 in all rules indicates that there is also a very high probability that the RHS data coincide with the existing LHS data. Based on the experimental results, the reasons people decided to vaccinate and their concerns can be summarized into seven association rules below.

1. People may choose to get vaccinated with Sinovac, as it is one of the available vaccine options.
2. People may get vaccinated by researching and gaining knowledge about the available vaccines.

Figure 2
Factors influencing COVID-19 vaccinations in Phuket using ARM



3. People may prioritize receiving Sinovac as they are concerned about the potential severe illness or death from the virus and have determined that Sinovac is the most suitable option for them.
4. After considering all available options, people may choose to receive Sinovac to contribute to the overall goal of herd immunity.
5. Despite concerns about potential side effects such as injection site pain, nausea, vomiting, numbness, and hemiplegia, some people still choose to get vaccinated.
6. People who live in Mueang district may choose to receive Sinovac as it is one of the available vaccine options in their area.
7. Women may choose to receive Sinovac, as it is one of the available vaccine options for everyone.

Based on these rules, they can be utilized to represent insights in the opposite direction, which can explain the reason of people not getting vaccinated. For example, the vaccination points were mostly available in Mueang district and in adequate in other districts. Perhaps, most people in Phuket were living in Mueang district during the COVID-19 outbreak. Furthermore, many people were not vaccinated because the most available vaccine type was Sinovac.

5. Conclusion

The researcher investigated the factors influencing people to be vaccinated in Phuket using ARM algorithms, including Apriori and FP-Growth. There are several processes until the knowledge can be discovered from data mining: data collection, data preprocessing, data cleansing, and data transformation. Minimum support, minimum confidence, and lift are values for measuring the efficiency of association rules. All these values are specified at different levels for the efficiency measurement criteria. In order of importance, the reasons why people received vaccinations were to 1) prevent severe illness or death and 2) enhance herd immunity, respectively, where the rules with the highest minimum support and minimum confidence are considered as the most weighed factor or reason causing people to be vaccinated and concerned after the vaccination. After the vaccination, the concerns of vaccinated participants include pain at the injection site, nausea, vomiting, numbness, and hemiplegia.

Interestingly, the results based on ARM are relatively consistent with other research works on public attitudes toward COVID-19 vaccinations despite different research approaches. The research results can be used as guidelines for the provincial public health or related agencies responsible for vaccination planning and public relations. Providing people with knowledge and understanding of vaccination, such as severe illness, fatality, and herd immunity, through different communication channels might encourage people to get vaccinated as required by the provincial administration. In addition, the indirect effect of getting the expected number of vaccinations will build the confidence of different tourism business operators in Phuket to welcome tourists, not just for the Phuket Sandbox Project but for the tourism business in the future as a world-class tourism destination.

Recommendations

The finding revealed that people were vaccinated due to several reasons. For instance, they lived in Mueang district, and the most available type of vaccine was Sinovac. They were also concerned about severe illness or death from the virus. However, these reasons can be reversed to explain why some people were not vaccinated. They can be used as guidelines to get more people to participate in vaccinations, for example, vaccination points should be distributed throughout different Phuket districts. Furthermore, more effective vaccination types should be available for people to decide which one they would like to take. The findings derived from this research can be substantiated through comparison with report data obtained from multiple sources, such as a comprehensive report on the overall vaccination outcomes in Phuket.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases,* 487–499.

Arora, J. , Shelza & Rao, S. (2013). An efficient ARM technique for information retrieval in data mining. *International Journal of Engineering Research & Technology*, *10*, 1079–1084.

Bhatia, P. (2019). *Data mining and data warehousing: Principles and practical techniques*. UK: Cambridge University Press.

Buathong, W. & Jarupunphol, P. (2018). Further experiments on a combination of linear SVM weight and ReliefF for dimensionality reduction. In *Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality (2018),* 6–9.

Buathong, W. & Jarupunphol, P. (2021). Dengue fever prediction modelling using data mining techniques. *International Journal of Data Mining and Bioinformatics*, *25*(1–2), 103–127

Celik, A. (2021). Using Apriori data mining method in covid-19 diagnosis. *Journal of Engineering Technology and Applied Sciences*, *5*(3), 121–131.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534.

Fayyad, U., Piatetsky Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–54.

Guillon, M. & Kergall, P. (2021). Factors associated with COVID-19 vaccination intentions and attitudes in France. *Public Health*, *198*, 200–207.

Katragadda, S., Gottumukkala, R., Bhupatiraju, R. T., Kamal, A. M., Raghavan, V., Chu, H., Kolluru, R. & Ashkar, Z. (2021). Association mining based approach to analyse covid-19 response and case growth in the United States. *Scientific Reports*, *11*(1), 18635.

Kaur, J. & Madan, N. (2015). Association rule mining: A survey. *International Journal of Hybrid Information Technology*, *8*(7), 239–242.

Manek, S., Shenoy, D., Mohan, C., & Venugopal, R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web*, *20*(2), 135–154. https://doi.org/10.1007/s11280-015-0381-x

Ministry of Higher Education, Science, Research and Innovation. (2021). MHESI reveals more than 20 million Thai people have been Covid-19 vaccinated (in Thai). Retrieved from: https://www.mhesi.go.th/index.php/en/all-media/infographic/4187-98642.html.

Miao, J. & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science, 91*, 919–926.

Mir, H. H., Parveen, S., Mullick, N. H., & Nabi, S. (2021). Using structural equation modeling to predict Indian people's attitudes and intentions towards COVID-19 vaccination. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *15*(3), 1017–1022.

Nagase, Y. (2021). Gunma is offering travel vouchers and a car to entice residents to vaccinate. *Time Out*. Retrieved from: https://www.timeout.com/tokyo/news/gunma-is-offering-travel-vouchers-and-a-car-to-entice-residents-to-vaccinate-081121.

Potdar, M., Potdar, S. & Potdar, M. (2021). A study of gender disparities towards COVID-19 vaccination drive in Maharashtra State, India. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *15*(6), 102297.

Rojanavasu, P. (2019). Educational data analytics using association rule mining and classification. In *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering*, 142–145.

Sharma, U. (2018). Association rule mining and genetic algorithm (GA) for data mining based intrusion detection system: A review approach. In *2018 3rd International Conference on Contemporary Computing and Informatics*, 249–251.

Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C. & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus COVID-19. *International Journal of Surgery*, *76*, 71–76.

Soni, H. K., Sharma, S. & Jain, M. (2017). Plausible characteristics of association rule mining algorithms for e-commerce. In *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics*, 36–39.

Tandan, M., Acharya, Y., Pokharel, S. & Timilsina, M. (2021). Discovering symptom patterns of covid-19 patients using association rule mining. *Computers in Biology and Medicine*, *131*, 104249.

Tantrakarnapa, K. & Bhopdhornangkul, B. (2020). Challenging the spread of COVID-19 in Thailand. *One Health*, *11*, 100173.

The Phuket News (2021). Officials change definition of population of Phuket, claim island 74% vaccinated. Retrieved from: https://www.thephuketnews.com/officials-change-definition-of-population-of-phuket-claim-island-74-vaccinated-80369.php.

Toth-Manikowski, S. M., Swirsky, E. S., Gandhi, R. & Piscitello, G. (2022). COVID-19 vaccination hesitancy among healthcare workers, communication, and policy-making. *American Journal of Infection Control*, *50*(1), 20–25. https://doi.org/10.1016/j.ajic.2021.10.004

Tumthong, W., Meesud, P., and Jarupunphol, P. (2021). Automatic detection and recognition of Thai vehicle license plate from CCTV images. In *Proceedings of the 13th International Conference on Information Technology and Electrical Engineering*, 73–76.

Vakil, C. (2021). D.C. giving away AirPods to encourage youth vaccinations. *The Hill*. Retrieved from: https://thehill.com/policy/healthcare/566867-dc-giving-away-airpods-to-encourage-youth-vaccinations/.

Vijayarani, S. and Sharmila, S. (2016). Comparative analysis of association rule mining algorithms. In *2016 International Conference on Inventive Computation Technologies, 3*, 1–6.

WHO COVID-19 Dashboard. (n.d.). Geneva: World Health Organization. Retrieved from https://covid19.who.int.

Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques*. USA: Morgan Kaufmann Publishers.

How to Cite: Buathong, W. & Jarupunphol, P. (2023). ARM for Analyzing Factors Influencing Vaccinations During the COVID-19 Outbreak. *Journal of Data Science and Intelligent Systems* 1(1), 45–54, https://doi.org/10.47852/bonviewJDSIS3202669