ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Uncovering the Most and the Least Factors Affecting Elderly Health Using Association Mining

Nichnan Kittiphattanabawon[1]

ABSTRACT

Thailand began the transition to an aging society in 2005 and is now stepping towards a completely aging society in 2021. The elderly's health problems are difficult to avoid. This research aims to analyze the most and least influential factors on the elderly's health status. The association analysis approach is applied to discover the relationship between various factors that affect good health. Data from 17,804 Thai elderly people from the National Statistical Office (NSO) of Thailand were employed. One hundred ninety-nine features were taken into the association rules mining modeling. The FP-growth algorithm was chosen to find the largest and the smallest factors affecting the health of the elderly. The interesting relationships among those factors are also disclosed. As a result, the feature that promotes good health is performing daily activities independently. Such a feature occurred with a support value of 99.99 percent. Additionally, several features that are less likely to appear, lower than 0.1 percent of support value, in healthy seniors are hard work, working at risk, living in an urban society, living with unfamiliar caregivers, having few children, and lacking sufficient government medical care. The knowledge gained from the findings can be considered for preparing health care for the aging society.

Article information:

Keywords: Elderly, Health, Aging Society, Association Mining, Association Rules, Itemsets

Article history:
Received: December 7, 2021
Revised: January 14, 2022
Accepted: April 23, 2022
Published: June 4, 2022 (Online)

DOI: 10.37936/ecti-cit.2022162.246993

1. INTRODUCTION

According to the United Nations (UN) standards, an aging society is defined as a region where the share of the population aged over 60 exceeds 10% of the whole population. If the share exceeds 20%, it is an aged society, or if over 28%, it is a super-aged society. Thailand has entered the aging society since 2005 due to the percentage of older persons, defined as those 60 years or older, is higher than 10%. It is estimated by the Office of the National Economic and Social Development Board (NESDB) that Thailand will become an aged society in 2021 and will become a super-aged society in 2031, since the rate of an aging population will be higher than UN standards. Such a situation affects the country regarding economic growth and social stability. One of the main factors is the health problems of the elderly, which are difficult to avoid because it is inevitable with the increasing age. Therefore, it is necessary to make arrangements for dealing with the aging society, such as health care and medical care. Knowing the various factors affecting the health status of the elderly will make the preparation more effective.

This research presents an analysis of factors affecting the health status of the Thai elderly by using representatives of the country's population as a whole. The study focuses on discovering factors that affect the good health of the Thai elderly and finding the relationship between these factors. An association analysis technique has been applied. The association analysis originates from market basket analysis, whose primary purpose is to find interesting relationships between products that customers are likely to buy together. Association analysis is utilized in various fields. For example, in Malaysia's Port State Control, association rule mining is applied to detect abnormal actions for identifying influential relationships between the inspection ports, flag states, the number of deficiencies raised, detention results, and ship categories [1]. Another example studied the distribution characteristics and affecting factors of diabetic

[1]The author is with School of Informatics, Walailak University, Nakhon Si Thammarat, Thailand 80160, Phone(+66)75672206, E-mail: knichcha@mail.wu.ac.th

retinopathy in diabetes mellitus patients. The researchers employed association rules to discover eye-related diseases and those with diabetic retinopathy [2]. It was also found that association mining has been conducted to analyze the relationships of combined drugs that cause side effects in the medical field. Association rules showed drug-drug interaction, in which the activity of one drug may change if taken with another drug. The discovery of structural relationships among them enabled providing essential guidance when making a co-prescription [3]. Implementing association rules happened in the education field. There existed an analysis of the relationship between the use of books and student grades [4]. Another example is analyzing the relationship between exercise styles and daily activities that should be done together. The findings from this research provide information to alleviate obesity caused by a lack of exercise [5]. One more interesting example is the decision to remove the ventilator based on the relationship between the patient's history and their symptoms. The discovered association rules can support physicians' consideration of weaning patients from mechanical ventilation [6].

Most research related to the study of factors affecting the health of the elderly is conducted using statistical analysis. A structural equation model explored both direct and indirect relationships of factors affecting depressive symptoms among the elderly in China. The depressive symptoms can harm the elderly's physical and mental health. The study variables included sociodemographic characteristics, poor health status, unhealthy habits, and sleep duration. Correlations and descriptive statistics, such as mean and standard deviation were measured to disclose the significant predictor for depressive symptoms in the elderly [7]. Another study in China occurred in Shanghai. The factors that concerned the health level of the younger elderly were investigated to improve the health of community living for them. One-way analysis of variance and stepwise regression analysis were employed to determine the influencing factors. The standard assessment instrument, SF-36 score, was selected to indicate a level of health [8]. Elderly living alone in South Korea were considered to identify the latent profiles based on the combined effects of self-esteem, life satisfaction, and depression. Multinominal logistic regression, correlation analysis, and descriptive statistics were conducted to analyze the latent profiles. The study defined the profiles as extremely depressed, severely depressed, mildly depressed, low life satisfaction, and positive adaptation [9]. Similarly, the elderly in India were analyzed. Lifestyle behaviors and background characteristics were investigated to analyze physical health and mental health through a statistical path. The study aimed to measure the extent of the symptom of psychological stress, find out the relationships of lifestyle behaviors with mental health outcomes, and investigate the moderating effects of lifestyle behaviors on mental health outcomes. Exploratory factor analysis, confirmatory factor analysis, multiple regression, and moderation analysis were executed for those purposes [10]. A study of mental health and quality of life among the elderly in Thailand was examined. The impact of resiliency, social support, health behaviors, work activities, and demographics were considered. The analysis was performed through descriptive statistics and hierarchical multiple regression. A comprehensive wellness program was invented based on the study results to promote the mental health and quality of life of the elderly [11]. It can be seen that most of the research used descriptive analysis with predetermined factors based on the research hypothesis. Although the findings are beneficial in explaining several factors related to the problems, they may not discover hidden knowledge resulting from factors in which they may have never had an interested.

Our research applies association rule mining techniques to the elderly data to reveal the relationship between the key factors affecting the health of the elderly. This technique can uncover entirely possible influencing factors. Several studies have conducted the association mining technique to find key factors in solving problems. They were searching hidden relationships among Twitter user-generated content [12], discovering relationships among news documents [13], highlighting meaningful rules between diagnosis types and diagnostic test requirements for emergency departments [14], and extracting dependencies among courses that help both students and advisors selecting the subject based on their performance in a previous study [15]. We utilized data from the NSO that is systematically sampled from the country's entire population. A association rule mining was applied to uncover the most and least relevant factors affecting the good health of the elderly to cope with the health problems of the aging society. The association technique does not need to determine fixed factors, but possible factors are selected. For this reason, the number of factors used in the process is greater than the number of factors determined by the research hypothesis. The finding of various and complete factors increase the efficiency in designing effective health promotion strategies for the aging society that occurs continuously in the country. Also, it helps find the knowledge that is useful for preparing the health of the Thai elderly.

The paper is structured as follows. A survey of the elderly persons in Thailand is detailed in Section 2. The conceptual framework in Section 3 has formed an overview of the research. Data preparation has been described step by step in Section 4. Section 5 is a presentation of the modeling methods of this research. The model evaluation is explained in Section 6. The findings of this research have been expressed in the experiments that appear in Section 7. A discussion of the experiments is presented in Section 8. Finally, conclusions are provided in the last section.

2. SURVEY OF THE OLDER PERSONS IN THAILAND

The NSO of Thailand performs as the center of statistics and information for the decision-making of the country. The elderly population is one group for which the NSO has conducted a survey every three years to gather statistical data. Our research utilized the elderly data from 2014. The NSO surveyed the data using Stratified Two-Stage Sampling, with the province as a stratum. In total, 83,880 households in every province across the country were investigated as a sample. Family members were interviewed to answer questions about the elderly in their homes. The questions were split into ten parts, with a total of 199 features, as detailed in Table 1. The first part is the household address where the elderly live. The second part is a question about the personal information of household members. For the third part, children's details of the elderly have been asked, such as the number of living children. The following part is about the work, income, and savings of the elderly. The fifth part consists of information on the house style where the elderly live, such as the floor or with whom they reside. The information in the sixth part is a question about how children support their elders and how often they visit them. The next part provides the health status of the elderly and caring for the elderly, including falls, and the ability to do activities by themselves. The question in the eighth part asks for information regarding the use of government services, which consists of participation in government-allocated activities and satisfaction with government services. Part 9 was obtained from asking elderly caregivers how much knowledge they have for elderly care. The last part is an interview of family heads of the elderly about the house style and the ownership of the property [?].

3. CONCEPTUAL FRAMEWORK

Before discussing the details of the research, the conceptual framework showing the overall operations must be established. This framework plays a role as a research framework derived from processing all of the concepts in the research and then placing in order. Figure 1 displays the conceptual framework of finding the most and least affected factors for healthy elderly and finding the relationships among them. The details within this framework will be expanded in the sections discussed below. The good health elderly dataset was already mentioned in Section 2. Details of data preparation, model creation, and model evaluation will be presented in Sections 4, 5, and 6, respectively.

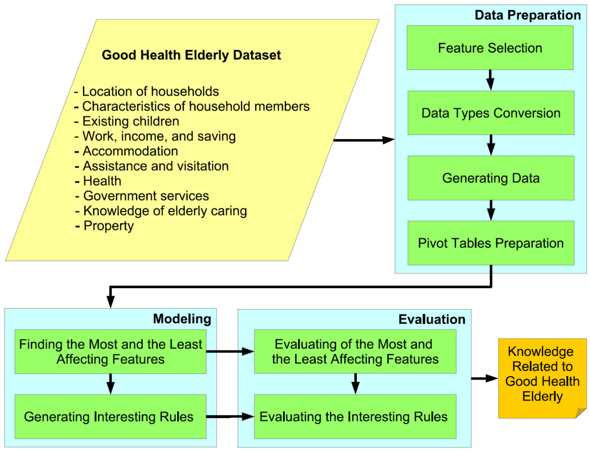

Fig.1: *Conceptual framework of finding the most and least affected factors for healthy elderly and finding their relationships.*

4. DATA PREPARATION

The NSO survey report [16] helps the process design of data preparation considerably. Referring to the NSO data of 83,880 households, we have only selected records of the elderly (aged 60 years and older), totaling 38,695 records. The data was categorized into 5 levels of health status: 1) very good, 2) good, 3) moderate, 4) poor, and 5) very bad. This research examined only the healthy elderly. That is, the health status of a good and very good levels, with a total of 17,804 examples. All these examples must be suitably prepared for the association model that will express the least and most affecting factors for the good health of the elderly. The crucial steps in preparing the data are described in the following subsections.

4.1 Feature Selection

The feature selection process involves selecting the necessary and appropriate factors for analysis without compromising the model performance. This research has selected the factors from Table 1, in which the chosen factors must have the opportunity to gain new knowledge relevant to healthy seniors. There are three main reasons why some factors were eliminated from the research. The first reason is that these factors cause the model to be too fit to the data. These features are the district, sub-district, village. The second is that they are identifiers that NSO uses to recognize records. Examples of the features, in this case, are enumeration area, sample household set, household number, month/year of interview. The last one removes factors are not relevant to the analysis, such as interviewee representatives, reasons for taking an interview on behalf of the elderly. As a result, we eliminated a total of 22 factors. That is, there are 177 remaining factors out of a total of 199 factors.

Table 1: The Number of Features for Each Part of Elderly Surveyed Data and Their Examples

No.	Topic	Number of Features	Examples of Features
1	Location of households	17	region, province, district, area, village, type of household
2	Characteristics of household members	13	gender, age, religion, education level, marital status, occupation
3	Existing children	5	number of living children, number of children living together
4	Work, income, and savings	24	working hours per day, average annual income, the value of savings or assets
5	Accommodation	21	the floor where the bedroom is located, people who live with them
6	Assistance and visitation	9	the amount received from children, the amount supported to children
7	Health	45	fall, ability to do activities (eat, wear clothes, take a shower, wash face/brush teeth, shave/comb, squat, lift heavy things, count change, pick your own medicine, etc.), happiness level
8	Government services	14	club membership, satisfaction in the government service system
9	Knowledge of elderly caring	17	relationship between elderly and caregivers, training of caregivers in caring for the elderly
10	Residential features and property ownership	31	housing type, consumer water source, ownership of appliances (TV, mobile phone, microwave, etc.)
	Total	**199**	

4.2 Data Types Conversion

Data types are important in modeling. It is necessary to transform the type of data to be suitable for interpretation. This research creates a model of association rules by mining the relationships of factors related to the good health of the elderly. The model will work well if it is based on the proper type of data. We have transformed the numerical features containing so many values into categorical data types. This is known as discretization. Without doing that, the model can hardly identify interesting patterns from numerical data with large numbers of values. Such features include the age of the elderly, the age of the children, and the age of caregivers. We discretized these features by dividing data into bins of data ranges. The NSO has classified the elderly ages into three groups [16]. Therefore, the age of the elderly was divided into three bins, including the youngest elderly (60-69 years), the middle elderly (70-79 years), and the oldest elderly (greater than or equal to 80 years). The age of the children was transformed into 11 age ranges, including seven bins of working age (0–19 years, 20–24 years, 25–29 years, 30–34 years, 35–39 years, 40–44 years, and 45–49 years), the golden age (50–59 years), and three bins of elderly (the youngest elderly, the middle elderly, and the oldest elderly as mentioned above). The caregivers' age groups are divided into 12 bins: seven working-age bins, which are defined as the same as the age of the children mentioned above, the golden age, and four bins of elderly, which are the youngest elderly, the middle elderly, the early oldest elderly, and the late oldest-old. All of the above is summarized in Table 2.

4.3 Generating Data

Generating data is a step to construct new values for features. Here, four new features were generated: (1) the level of education, (2) the primary occupation in the last year, (3) business in the last year, and (4) the occupation in the last week. These features are too specific, with the levels of education of 1,000 possible values. The primary occupation last year and last week had 10,000 possible values The business last year has up to 100,000 possible values. Creating a model with these values can lead to overfitting since it does its best to account for every single point. A model that generalizes well to the actual data is needed, so we have to generate new features while not distorting the meaning. The level of education is a 3-digit code, where the first digit is the education level, the second digit is the type of education, and the third digit is a grade. We divided them by 10, resulting in a 2-digit code: cutting out the last digit, which is the grade. Synthesis of the primary occupation in the last year and last week were performed in the same way. They have a 4-digit code, with the first two digits showing the category, the third digit is the

Table 2: *Data Types Conversion from Numerical Values to Categorical Data Type*

Feature	Number of Bins	Bin Name
Age of elderly	3	youngest elderly, middle elderly, oldest elderly
Age of children	11	worker1-worker7, golden, youngest elderly, middle elderly, oldest elderly
Age of caregivers	12	worker1-worker7, golden, youngest elderly, middle- elderly, early oldest elderly, late oldest elderly

group and the fourth digit is the profession. These 4-digit codes were divided by 100 to become 2-digit codes. That means that only occupation categories and occupation groups will be used to build models. The business in the past year was a 5-digit code: the first two digits were a category, the third digit was a group, the fourth digit was a subgroup, and the fifth digit was an activity carried out. These codes were used as 2-digit codes. That is, we just used the business category. Consequently, we divided each of them by 1,000. The value of the generated features is shown with their meanings in Table 3.

4.4 Pivot Tables Preparation

This step prepares the data in the form of a pivot table, summarizing the data in a ready format for creating the association model. Here, the pivot table was structured as a 2-dimensional table, with each row (an example) representing one elderly person and each column (a feature) representing his or her characteristics. That means that the total number of rows equals the number of people, and the total number of columns equals all possible values of features. Two pivot tables were created for this research: a pivot table for the elderly with good health levels, and a pivot table for the elderly with very good health levels. Here, data was collected from 17,804 older people, consisting of 16,629 healthy elderly and 1,175 very healthy seniors. The healthy elderly category has 1,193 possible features, and very healthy seniors have 891 possible feature values. This led to two pivot tables, which were a pivot table for the elderly with a good level of health with 16,629 examples x 1,193 features, and a pivot table for the elderly with a very good level of health with 1,175 examples x 891 features. The pivot table for the very healthy elderly data with 1,175 examples and 891 features is illustrated in Table 4. The table headers represent features. In Table 4, $A4$ is the gender of the elderly, which has 2 possible values, male and female. Here, males are represented by 1, and females are represented by 2. Therefore, the male gender feature's value is also set to $A4\_1$, and the female gender feature is coded as $A4\_2$. $A5$ is a feature that refers to the age of the elderly. The possible values have been redefined as the youngest elderly (60-69 years), the middle elderly (70-79 years), and the oldest elderly (greater than or equal to 80 years) as specified in Section 4.2 $A5\_youngest$, $A5\_middle$, and $A5\_oldest$ refer to the elderly in early, middle, and late old age, respectively. $A7$ is a feature that represents education level. The value of $A7$ is a 2 digit code as described in Section 4.3 $A7\_61$ refers to bachelor degrees, while $A7\_66$ is a graduate diploma. There are two possible values for the body of the table, 1 and 0, where 1 means "yes" and 0 means "no". For example, the interpretation of the first elder, which appears in the first row, shows $A4\_1$, $A5\_youngest$, and $A7\_66$ with the value 1. It means that this person is a man aged between 60-69 years and graduated with a diploma degree.

5. MODELING

The construction of a model for examining Thai elderly health status for this research covers the analysis of the health status of the elderly with good and very good health. The data analytics tool used in modeling is RapidMiner Studio. The model is the association model, which explores the association rules relating to the health status of the elderly, leading to the factors that have the highest effect on the health of the elderly and the factors that affect the health of the elderly the least. Moreover, the relationships among the features will be discovered. There are two steps in creating an association model: (1) finding frequent itemsets, and (2) generating association rules. The details of both steps will be explained in the following subsections.

5.1 Finding Frequent Itemsets

This step is a process of finding features that appear together in a relation. Here, the feature is called *item* and the set of features that occur together is called *itemset*. An itemset may be composed of one or many features. The common n features are called *n-itemset*. That is, in the case of one common feature, it is called *1-itemset*, in the case of two common features, it is called *2-itemset*, and so on. The itemsets are written in the form of sets with members as items. The features shown in Table 4 which were described in Section 4.4 will be used as examples here. The first example is a *2-itemset*, such as $\{A4\_1, A5\_youngest\}$. Since there are two items in the set, this set is called a *2-itemset*. This example presents that $A4\_1$ coincides with $A5\_youngest$, which means that the elderly who are male are between 60 and 69 years old. Likewise, if it is a *3-itemset* such as $\{A4\_1, A5\_youngest, A7\_66\}$,

Table 3: *Feature Values Generated and Their Meanings*

Feature	Value Generated (before → after)	Meaning (before → after)
Education level	661 → 66	bachelor degree, graduate diploma, first year graduate diploma → bachelor degree, graduate diploma
Occupation type	5113 → 51	service staff, personal service staff, travel attendant, guide → service staff, personal service staff
Business type	03212 → 03	fisheries and aquaculture, aquaculture, sea aquaculture, sea shrimp farming → fisheries and aquaculture

Table 4: *Examples of a Pivot Table of the Elderly with a Very Good Level of Health*

No.	...	A4_1	A4_2	A5_youngest	A5_middle	A5_oldest	...	A7_61	...	A7_66	...
1	...	1	0	1	0	0	...	0	...	1	...
2	...	0	1	1	0	0	...	1	...	0	...
...	...	0	1	0	1	0	...	0	...	1	...
1175	...	0	1	0	0	1	...	0	...	1	...

it means that *A4_1*, *A5_youngest*, and *A7_66* occur together. It means that there are older men aged 60-69 years who have graduated with a diploma. If any of these itemsets occur together often, they will be called a *frequent itemset*. If $\{A4\_1, A5\_youngest\}$ is a frequent itemset, most male-elderly are between the ages of 60 and 69. This set will be called a *2-frequent itemset*. Similarly, the $\{A4\_1, A5\_youngest, A7\_66\}$ will be called called a *3-frequent itemset*. This itemset means that most male-elderly aged 60-69 and graduate with a diploma. To find frequent itemsets of this research, the algorithm FP-Growth was used. The FP-Growth algorithm has high efficiency in processing time because it uses a tree structure called FP-Tree to store data while processing [17], [18], [19]. For this research, frequent itemsets were applied to analyze common factors that reflect the good health of the elderly to discover knowledge that can be used to promote good health for the elderly. Frequent 1-itemsets, frequent 2-itemsets, and frequent 3-itemsets will be discovered in this research. The frequency of itemsets can be measured by a measure called *support*, which will be explained in Section 6. This measure gives an idea of how frequent an *itemset* is in all the transactions. In this research, we intend to find itemsets that occur both the most frequently and the least often to find the most common and least common factors that affect the health of the elderly.

5.2 Generating Association Rules

Generating association rules is the process of establishing relationships among the frequent itemsets obtained from Section 5.1 The relationships or rules, called association rules, reveal how items are related to each other. The association rule is in the form of $\{X\} \rightarrow \{Y\}$, where $\{X\}$ is called an *antecedent* which means the itemset(s) that occurs first, and $\{Y\}$ is called a *consequent* which means the itemset(s) that occurs later. As for $\{X\} \rightarrow \{Y\}$, it means that if there is an $\{X\}$ occurring, then $\{Y\}$ has a chance to appear. Both the antecedent and the consequent can be composed of more than one item. Association rules arising from a *2-frequent itemset* will be rules in the form of the 1-item antecedent and the 1-item consequent. For example, a rule generated from $\{A4\_1, A5\_youngest\}$ will have two rules, $\{A4\_1\} \rightarrow \{A5\_youngest\}$ and $\{A5\_youngest\} \rightarrow \{A4\_1\}$. Similarly, rules generated from a *3-frequent itemset* can have an antecedent with one item and a consequent with two items or an antecedent with two items and a consequent with one item. A *4-frequent itemset* can create rules with an antecedent of one item and a consequent of three items, those of an antecedent of two items and a consequent of two items, or those of an antecedent of three items and a consequent with one item. Rules will be formulated in this way for all *n-frequent itemsets*. Table 5 shows examples of association rules created from a *2-frequent itemset*, and a *3-frequent itemset*. If the relationships of the items in the rules are very closely related, the rules will be called strong association rules, which can be identified as interesting rules. If any rules exist for which the item is not very relevant, these rules can be called weak or uninteresting rules. For example, if the rule $\{A4\_1\} \rightarrow \{A5\_youngest\}$ is a strong association rule, it means that if older people are male, they tend to be between the ages of 60 and 69. On the other hand, when the rule is weak, the rule means that if seniors are male, they are rarely found between the ages of 60 and 69. Another example for rules generated by a *3-frequent itemset* is $\{A4\_1\} \rightarrow \{A5\_youngest, A7\_66\}$. If the rule is strong, it means that if seniors are male, they are often between the ages of 60-69 and graduated with a graduate diploma. In contrast to this rule, the strong rule $\{A5\_youngest, A7\_66\} \rightarrow \{A4\_1\}$ means that if an elderly is 60-69 years old and graduated with a graduate diploma, it is likely that they are men. However, if the rule $\{A4\_1\} \rightarrow \{A5\_youngest, A7\_66\}$ is weak, it means that if seniors are men, then it is unlikely that they will be aged between 60-69 years and graduate with a graduate diploma. On the other hand, the weak rule $\{A5\_youngest, A7\_66\} \rightarrow \{A4\_1\}$ means that if the elderly are aged between 60-69 years and have completed their graduate diploma, it is often not clear that they are men. However, the strength of the rule can be identified. A measure called *confidence* which will be discussed in Section 6, will be used.

6. EVALUATION

The evaluation process measures effectiveness of the association model created in the previous step. It determines how reliable the model is. The evaluation has two parts: the quality evaluation of frequent itemsets, and the association rules' quality evaluation.

6.1 Evaluation of Frequent Itemsets

Evaluating the effectiveness of frequent itemsets is a measurement of the percentage of frequent occurrences of features. The value applied for measuring the performance of frequent itemsets is *support*. Technically, support is the fraction of the total number of transactions in which the itemset occurs, as shown in Eq. 1.

$$support(\{X, Y\}) = \frac{\text{number of occurrences of both } X \text{ and } Y}{\text{total number of examples}} \times 100, \tag{1}$$

where X and Y are itemsets. The *support* has a maximum value of 100 percent. The percentage of itemset co-occurrences is specified with the support criteria to indicate the percentage of occurrences of itemsets. The required criteria can be specified as appropriate with the *support threshold*. If the support threshold is set to 90 percent, the frequent itemsets obtained from the algorithm have a support value of 90 percent or more. For example, if the $\{A4\_1, A5\_youngest\}$ has a support value of 95 percent, then this set will be a frequent itemset. The meaning of $\{A4\_1, A5\_youngest\}$ with support of 95 percent is that 95 percent of the elderly are male and aged between 60-69. However, if its support value is 80 percent, it will not be extracted as a frequent itemset. Any itemsets with a support value less than the support threshold will be called *infrequent itemsets*. In other words, these itemsets are those that rarely appear in the existing dataset. If the $\{A4\_1, A5\_youngest\}$ with a support value of 80 percent is infrequent, there are only 80 percent of the elderly who are male and aged between 60-69 years old.

6.2 Evaluation of Association Rules

A measure named *confidence* is applied to measure the interest of the rule. Confidence of $\{X\} \rightarrow \{Y\}$ defines the likeliness of occurrence of consequent (Y) in the example sets given that the example sets already have the antecedents (X), where X and Y are itemsets. Mathematically, confidence is the conditional probability of occurrence of a consequent given the antecedent is presented, as shown in Eq. 2.

$$confidence(\{X\} \rightarrow \{Y\}) = \frac{support(\{X\} \cup \{Y\})}{support(\{X\})} \times 100 \tag{2}$$

Like support, the confidence value has a maximum value of 100 percent. Similarly, *a confidence threshold* is defined to determine the interestingness of an association rule. How the confidence threshold is given depends on how much the relationship between items in the rule is needed. Rules with confidence values lower than confidence thresholds are designated uninteresting rules or weak rules. On the other hand, rules with a confidence equal to or higher than the confidence threshold will be interesting or strong rules. For example, for the 90% confidence threshold, if the rule $\{A4\_1\} \rightarrow \{A5\_youngest\}$ has the confidence value of 95 percent, it will be recognized as a strong rule. The rule implies that among older adults who are male, 95 percent are aged between 60-69 years old. On the other hand, the rule will be interpreted to mean that only 40 percent are aged 60-69 years among older men who are males when this rule's confidence value is 40 percent. It means that this rule is uninteresting or weak, since its confidence value is lower than the confidence threshold.

7. EXPERIMENTS

The model executed in the experiments was developed to analyze the relationships between features affecting the health status of the elderly. The main objective is to determine the factors that frequently occur with healthy seniors and establish the relationships between these factors. The experiment was separated into three parts to achieve the goal, as follows. The first experiment was to discover the factors that frequently occur with healthy seniors. This experiment was designed to uncover the factors that significantly impact how healthy the elderly are. The second experiment was performed to find of factors that rarely occur with healthy elderly. The experiment aimed to explore factors that have minimal effect on how healthy the elderly are. The final experiment was the finding of the relationships between various factors affecting the health of the elderly. The experiment was intended to reveal the relationships between various factors to discover the causes and effects on their health. Details and results of all three experiments will be explained next.

7.1 Discovered Frequent Itemsets

Table 5: *Examples of association rules generated from the frequent itemsets*

Itemsets Size	Frequent Itemsets	Association Rules
2-frequent itemsets	{A4_1, A5_youngest}	{A4_1} → {A5_youngest}, {A5_youngest} → {A4_1}
3-frequent itemsets	{A4_1, A5_youngest, A7_66}	{A4_1}→{A5_youngest, A7_66}, {A5_youngest}→{A4_1, A7_66}, {A7_66}→{A4_1, A5_youngest}, {A4_1, A5_youngest}→{A7_66}, {A4_1, A7_66}→{A5_youngest}, {A5_youngest, A7_66}→{A4_1}

The first experiment is designed to discover frequent itemsets related to healthy elderly. The discovered frequent itemsets will be used to find answers to the most affecting factors for the health of the elderly. The form of frequently occurring factors are both the single-occurring factor which is called *1-frequent Itemset*, and more than one factor occurring which will be studied over *2-frequent Itemset*, *3-frequent Itemset*, ..., *n-frequent Itemset*. The support threshold was set from 100 percent steadily decreasing to find the factors that affect the health of the elderly the most in this experiment. When the process ends, no itemsets were found at the support value equal to 100 percent. The highest support value that causes factors to be discovered is 99.99 percent. This kind of processing is consistent with a survey of the use of support in ways without setting a minimum value [20]. A minimum support value is not needed, but a k-value is needed, indicating the number of rules produced. This method is known as the top-k association rule method [21]. Here, determining the minimum support value is difficult since we do not know how many final association rules will be interesting enough for the health status of the elderly. Therefore, we tried to mine the top-k rules with the highest support that met the desired confidence, which will be specified in Section 7.3 As a result, Itemsets that have relationships with healthy seniors and very healthy seniors with a support value of 99.99 percent are exposed in Table 6. Considering the details in Table 6, the factors affecting the health status of the elderly, for elderly with good or very good health, almost 100 percent (almost everyone) is outstanding in doing daily activities by oneself. Six activities can be done by themselves, including (1) eating, (2) wearing clothes, (3) bathing, (4) washing the face / brushing teeth, (5) using the bathroom/toilet, and (6) shaving/combing. Besides, the discovered patterns revealed that the elderly are capable of doing one activity (1-frequent itemsets), two activities (2-frequent itemsets), three activities (3-frequent itemsets), four activities (4-frequent itemsets), and all five activities (5-frequent itemsets). In other words, the first six activities that most healthy seniors can do on their own are eating, putting on clothes, taking a shower, washing their face / brushing their teeth, using the bathroom/toilet, and shaving/combing.

7.2 Discovered Infrequent Itemsets

The second experiment is designed to find factors that happen together. However, unlike the first experiment, this uncovers factors that do not occur concurrently, while the first is to locate factors that are likely to appear together. The experimental results in this section found the least affecting factors of the health status of the elderly. The patterns of features that rarely occur together regularly tend to outnumber the frequent co-occurring features due to the large number of features considered. In this research, there are 1,193 features related to the elderly with good health conditions and 891 features associated with the elderly with very good health conditions. These features are very likely to happen together rarely. There are a vast number of patterns derived from these features. Consequently, only the least frequently occurring features will be presented here, and the pattern of more than one co-occurring feature will not be displayed. As a result, features that occur less than 0.1 percent of support with healthy seniors will be exposed. Details are expressed in Table 7. These findings revealed that the elderly with good health conditions were not found among those engaged in machinery and tour arrangements. It was also unearthed that healthy seniors do not live in condominiums, mansions, or small rooms. In other words, healthy seniors live in a family-friendly homes. Moreover, it can be seen that it is a group of the elderly with potential caregivers. That is, it was not discovered that there were elderly caregivers aged 20-29 years or younger caregivers. As for the findings, the elderly who are very healthy are as follows. It was not detected in medical community assistants, such as family planning assistants, hygiene, and dietary assistants. Furthermore, it was not recognized that a group of people worked too hard, more than 14 hours a day. More than that, it was also clear that very healthy seniors have an appropriate number of children. The results in the analysis show that there are no more than four children in the household and no more than six sons still alive. An interesting issue is that most elderly who are very healthy are often in a

Table 6: *Frequent Itemsets Whose support = 99.99 Percent (F68_3: eat oneself, F69_3: put on cloths oneself, F70_3: bathe oneself, F71_3: wash face/brush teeth by oneself, F72_3: use the toilet by oneself, F73_3: shave/comb by oneself)*

1-frequent itemsets	{F68_3}, {F69_3}, {F70_3}, {F71_3}, {F72_3}, {F73_3}
2-frequent itemsets	{F68_3, F69_3}, {F68_3, F70_3}, {F68_3, F71_3}, {F68_3, F72_3}, {F68_3, F73_3}, {F69_3, F70_3}, {F69_3, F71_3}, {F69_3, F72_3}, {F69_3, F73_3}, {F70_3, F71_3}, {F70_3, F72_3}, {F70_3, F73_3}, {F71_3, F72_3}, {F71_3, F73_3}, {F72_3, F73_3}
3-frequent itemsets	{F68_3, F69_3, F70_3}, {F68_3, F69_3, F71_3}, {F68_3, F69_3, F72_3}, {F68_3, F69_3, F73_3}, {F68_3, F70_3, F71_3}, {F68_3, F70_3, F72_3}, {F68_3, F70_3, F73_3}, {F68_3, F71_3, F72_3}, {F68_3, F71_3, F73_3}, {F68_3, F72_3, F73_3}, {F69_3, F70_3, F71_3}, {F69_3, F70_3, F72_3}, {F69_3, F70_3, F73_3}, {F69_3, F71_3, F72_3}, {F69_3, F71_3, F73_3}, {F69_3, F72_3, F73_3}, {F70_3, F71_3, F72_3}, {F70_3, F71_3, F73_3}, {F70_3, F72_3, F73_3}, {F71_3, F72_3, F73_3}
4-frequent itemsets	{F68_3, F69_3, F70_3 F71_3}, {F68_3, F69_3, F70_3 F72_3}, {F68_3, F69_3, F70_3 F73_3}, {F68_3, F69_3, F71_3 F72_3}, {F68_3, F69_3, F71_3 F73_3}, {F68_3, F69_3, F72_3 F73_3}, {F68_3, F70_3, F71_3 F72_3}, {F68_3, F70_3, F71_3 F73_3}, {F68_3, F70_3, F72_3 F73_3}, {F68_3, F71_3, F72_3 F73_3}, {F69_3, F70_3, F71_3 F72_3}, {F69_3, F70_3, F71_3 F73_3}, {F69_3, F70_3, F72_3 F73_3}, {F69_3, F71_3, F72_3 F73_3}, {F70_3, F71_3, F72_3 F73_3}
5-frequent itemsets	{F68_3, F69_3, F70_3 F71_3 F72_3}, {F68_3, F69_3, F70_3 F71_3 F73_3}, {F68_3, F69_3, F70_3 F72_3 F73_3}, {F68_3, F69_3, F71_3 F72_3 F73_3}, {F69_3, F70_3 F71_3 F72_3 F73_3}

Table 7: *Itemsets Whose support is Less Than 0.1 Percent*

No.	Discovered Features of Good Health Elderly	Discovered Features of Very Good Health Elderly
1	Elderly whose occupation is a machine operator and assemble machinery or have a business as a group of a travel agency and tour organization	Elderly who have a career as medical assistants in modern health (except nursing)
2	Elderly people living in condominiums, mansions or rooms in the house	Elderly who work more than 14 hours a day
3	Elderly with caregivers aged 20-29 years or with friends/neighbors/acquaintances as carers	Elderly with no more than 6 sons who are still alive or not more than 4 children in the same household
4	-	Elderly who do not receive public health benefits because they are not in their hometown, so they do not have rights
5	-	Elderly with caregivers aged 80 years or older, or caregivers are servants/employees, or caregivers in the same village or municipality, or have received unofficial care

domicile where they can use the government's medical privileges. Besides, they are a group of elderly with potential caregivers. There are no elderly caregivers older than the age of 80, and most caregivers are also relatives.

7.3 Discovered Association Rules

The last experiment intended to discover the relationships among the factors that frequently occur to learn which factors are causing the consequent factors. The analysis results of this experiment revealed how the relationships between the frequent itemsets are correlated. These relationships were expressed in the form of association rules ($X \rightarrow Y$) as mentioned in Section 5.2 The association rules were generated by the frequent itemsets discovered in Section 7.1 Those features consist of 6 features: self-eating, self-wearing clothes, self-bathing, self-washing/brushing, self-using the bathroom/toilets, and self-shaving/combing, which was discovered with a support value of 99.99 percent. The discovered rules indicated that these six factors are highly correlated - that is, the rules have 100 percent confidence. That is, the elderly who are healthy or very healthy, all have the same association rules. The association rules that have a confidence value of up to 100 percent can be interpreted to mean that if the features which are the antecedent (X) occur, then the features which are consequent (Y) will undoubtedly happen. The experimental results displayed that all six features can be turned to antecedent and consequent in all possible cases. In other words, for almost all elderly who are very healthy or have very good health, if they can do any one of the activities, they will unquestionably perform the rest. For example, for the 2-frequent itemsets $\{F68\_3, F69\_3\}$, the discovered rules are $\{F68\_3\} \rightarrow \{F69\_3\}$ and $\{F69\_3\} \rightarrow \{F68\_3\}$. Another example is that, for the 3-frequent itemsets $\{F68\_3, F69\_3, F70\_3\}$, the revealed rules cover all possible cases: $\{F68\_3\} \rightarrow \{F69\_3, F70\_3\}, \{F69\_3\} \rightarrow \{F68\_3, F70\_3\}$, $\{F70\_3\} \rightarrow \{F68\_3, F69\_3\}, \{F68\_3, F69\_3\} \rightarrow \{F70\_3\}, \{F68\_3, F70\_3\} \rightarrow \{F69\_3\}$, and $\{F69\_3, F70\_3\} \rightarrow \{F68\_3\}$. For the interpretation of the rules, some examples will be exhibited as follows. The $\{F68\_3\} \rightarrow \{F69\_3\}$ means that if the elderly can eat by themselves, they will surely be able to put on their clothes without needing any help. On the other hand, $\{F69\_3\} \rightarrow \{F68\_3\}$ means that if the elderly can put on their clothes alone, they will surely be able to eat by themselves. Other examples are $\{F68\_3\} \rightarrow \{F69\_3, F70\_3\}$ and $\{F69\_3, F70\_3\} \rightarrow \{F68\_3\}$. The former rule means that if the elderly can eat by themselves, they will definitely put on clothes and bathe by themselves. The latter rule expresses that if the elderly can put on clothes and take a bath themselves, they will assuredly eat by themselves.

8. DISCUSSION

The research results have shown that the most relevant factors for the health of the elderly are self-essential activities. The most common activities were eating food, putting on clothes, bathing, washing their face, brushing their teeth, using the bathroom, shaving, and combing their hair. Additionally, it has been discovered that if healthy seniors can carry out any of these activities by themselves, they also can perform the other activities themselves. This finding is in line with Phalasuek and Thanomchayathawatch [22] that suggested that the elderly who can carry out their daily activities will significantly reduce the burden of care for by family members or other people. It also helps the elderly have an excellent quality of life and happiness in self-reliance. Therefore, if there is sufficient health preparation for the elderly, the engaged organizations will reduce the burden of looking after the elderly.

As for the occupational issues, it was ascertained that the types of careers that cause the elderly to encounter health problems are occupations in the fields of machinery, tourism, and medical assistants, since these occupations might make the elderly feel insecure. Since the elderly themselves are as physically healthy as working-age people, safety is a matter that can have a significant impact on the elderly [23]. Besides, it is also following the Notification of the Ministry of Labour, Thailand, Subject: "Request for Cooperation and Support for the Elderly to Have a Job", announced on 8 March 2019 [24]. The vital point in the announcement is to request employment for the elderly in a way that is not dangerous to their health and which must be safe. Such work covers clerks, trades, and folk handicrafts. However, the duration of work should not exceed 7 hours per day. It can be seen that the content mentioned earlier corresponds to the knowledge obtained from our research that it is not found that overworked seniors are healthy people.

For the issue of caregivers who take care of the elderly, the family members should be encouraged to look after the elderly because the familiar caregivers have a better understanding of the elderly than others [22]. Our research has already revealed that it is rare that healthy seniors have unfamiliar caregivers in Thai society. This situation showed that Thailand is still living in a family-oriented way. Nowadays, even though family members are studying or working in a location far away from their family, they do not leave the elderly with someone unfamiliar. Family caring has a positive mental impact on the elderly, driving them to good physical health. Prasartkul and et al. [25] have analyzed the impact of situations where the elderly are not getting good care. They have summarized interestingly how the elderly feel lonely or without self-worth. They claim that their children do not have much time and that the elderly do not have any roles in the family. Therefore, it influences their mental state and continuously affects their health. They also point out the issues regarding the case of the elderly living in urban societies. They have said that the driving force in urban society is driven by socioeconomic, sickness, or public services. The elderly become vulnerable populations who do not receive justice in their human rights and freedoms in society. Fortunately, our research indicated that the elderly in Thailand rarely live in urban areas such as condominiums, mansions, or small rooms.

The last issue discussed here is the number of children of the elderly and their medical treatment rights from government welfare. This research has uncovered that there are not many healthy elderly who have a small number of children, or those not in the domicile who hold medical rights. This finding involves the hospitalization of the elderly during illness. Children are considered the prominent persons responsible for looking after the elderly when they are sick. Without a child to take care of them, the elderly may recover slowly from illness. It corresponds to the study of the situation in Thailand regarding the hospitalization of the elderly [26]. The study presented that the elderly who have carers during the recovery period will recover faster. They pointed out that the elderly with poor welfare take a higher average number of recovery days than those with civil servant rights. They also concluded that the elderly with insufficient medical rights have a small household size, so they lack sufficient care while recuperating or without constant caregivers. These significantly affects the health status of the elderly.

9. CONCLUSIONS

This research aimed to discover the factors that have the most effect and least effect on the health status of Thai elderly who have good health. The results can be taken into consideration to prepare for the health of the Thai elderly. The data used here is from 2014 from the NSO. The NSO surveyed data under the project named "The 2014 Survey of the Older Persons in Thailand". The project has provided data using statistical methods correctly according to academic principles. Data of 17,804 elderly who are in good health and very good condition were used for training.

It can be observed that most works used a data analysis method based on descriptive analysis. Our research uses the association mining technique as part of the machine learning field. We work with a wide variety of possible meaningful data, requiring optimized feature design for analysis. The results obtained are consistent with the results of other existing descriptive-analytical research studies. Nevertheless, our work can analyze a large amount of data faster and without pre-assumptions because the association mining technique can uncover all the probable factors. However, our research does not cover the elderly with poor health conditions. Further analysis should be performed to give clear and complete comparative results for all levels of health conditions. Remodeling should be treated by the refreshed Thailand elderly data (Thailand will survey data every three years) to acquire new knowledge fitting with the current aging society situation. Aside from this, more studies should be conducted in addition to health issues to discover new knowledge that is beneficial for application and reflects the need for implementation by the Government sector to cope with the aging society that is coming soon.

ACKNOWLEDGEMENT

This work was supported by Walailak University under the funding agreement no. WU61314. In addition to this, the author would like to thank the National Statistical Office of Thailand for their good support for the data they provided for this research.

References

[1] M. T. Osman, C. Yuli, T. Li and S. F. Senin, "Association Rule Mining for Identification of Port State Control Patterns in Malaysian Ports," *Maritime Policy & Management*, vol.48, pp. 1082–1095, 2020.

[2] X. Yao, X. Pei, Y. Yang, H. Zhang, M. Xia, R. Huang, Y. Wang and Z. Li, "Distribution of Diabetic Retinopathy in Diabetes Mellitus Patients and Its Association Rules with Other Eye Diseases," *Scientific Reports*, vol.11, no.1, pp.16993, 2021.

[3] F. Held, D. G. L. Couteur, F. M. Blyth, V. Hirani, V. Naganathan, L. M. Waite, M. J. Seibel, D. J. Handelsman, R. G. Cumming, H. G. Allore and D. Gnjidic, "Polypharmacy in Older Adults: Association Rule and Frequent-set Analysis to Evaluate Concomitant Medication Use," *Pharmacological Research*, vol. 116, pp. 39–44, 2017. Country in focus: Pharmacology in Australia.

[4] S. Verma and J. Doshi, *Correlation Between Text Book Usage and Academic Performance of Student in Higher Education Using R*, pp. 11–18. Springer, 04 2017.

[5] S. Sharma, "Concept of Association Rule of Data Mining Assists Mitigating the Increasing Obesity," *International Journal of Information Retrieval Researchs*, vol. 7, no. 2, pp. 1–18, apr 2017.

[6] Y. Gao, A. Xu, P.J.-H. Hu and T.-H. Cheng, "Incorporating Association Rule Networks in Feature Category-weighted Naive Bayes Model to Support Weaning Decision Making," *Decision Support Systems*, vol. 96, pp. 27-38, 2017.

[7] Y. Xie, M. Ma, W. Wu, Y. Zhang, Y. Zhang and X. Tan, "Factors Associated with Depressive Symptoms Among the Elderly in China: Structural Equation Model," *International Psychogeriatrics*, vol. 33, no. 2, pp. 157-167, 2021.

[8] Z. Zhu, D. Zhu, Y. Jiang, Y. Lin, Y. Yang and W. Luan, "Cross-sectional Study on the SF-36, the General Self-efficacy, the Social Support, and the Health Promoting Lifestyle of the Young Elderly in a Community in Shanghai, China," *Annals of palliative medicine*, vol. 10, no. 1, pp. 518-529, 2021.

[9] J. H. Park, S. Min, Y. Eoh and S. H. Park, "The Elderly Living in Single-person Households in South Korea: a Latent Profile Analysis of Self-esteem, Life Satisfaction, and Depression," *Quality of Life Research*, vol. 30, no. 4, pp. 1083-1092, 2021.

[10] P. Bhandari and B. Paswan, "Lifestyle Behaviours and Mental Health Outcomes of Elderly: Modification of Socio-economic and Physical Health Effects," *Ageing International*, vol. 46, no. 1, pp. 35–69, 2021.

[11] Paul Ratanasiripong, Nop Ratanasiripong, Monpanee Khamwong, Sarinya Jingmark, Ploen-

pit Thaniwattananon, Pennapa Pisaipan, Ladda Sanseeha, Nongnaphat Rungnoei, Wallapa Songprakun, Asawinee Tonkuriman, and Suchart Bunyapakorn. The Impact of Resiliency on Mental Health and Quality of Life Among Older Adults in Thailand. Journal of Health Research, 2021.

[12] L. Cagliero and A. Fiori, "Discovering Generalized Association Rules from Twitter," *Intelligent Data Analysis*, vol. 17, pp. 627-648, 2013.

[13] N. Kittiphattanabawon, T. Theeramunkong and E. Nantajeewarawat, "Region-based Association Measures for Ranking Mined News Relations," *Intelligent Data Analysis*, vol.18, pp. 217–241, 2014.

[14] G. Saryer and C. O. Taşar, "Highlighting the Rules Between Diagnosis Types and Laboratory Diagnostic Tests for Patients of an Emergency Department: Use of Association Rule Mining," *Health Informatics Journal*, vol. 26, no.2, pp. 1177–1193, 2020.

[15] R. Shatnawi, Q. Althebyan, B. Ghaleb and Mo. Al-Maolegi, "A Student Advising System Using Association Rule Mining," *International Journal of Web-Based Learning and Teaching Technologies*, vol.16, no. 3, pp. 65–78, 2021.

[16] National Statistical Office. The 2014 Survey of the Older Persons in Thailand. Technical report, Ministry of Information and Communication Technology, Bangkok: Text and Journal Publication, 2014.

[17] M. S. Mythili and A. R. Mohamed Shanavas, "Performance Evaluation of Apriori and FP-growth Algorithms," *International Journal of Computer Applications*, vol. 79, no. 10, pp.34–37, October 2013.

[18] K. Dharmaraajan and M. A. Dorairangaswamy, "Analysis of FP-growth and Apriori Algorithms on Pattern Discovery from Weblog Data," *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pp. 170– 174, 2016.

[19] R. Yusuf and Z. Lawal, "Performance Analysis of Apriori and FP-growth Algorithms (Association Rule Mining)," *International Journal of Computer Applications in Technology*, vol. 7, pp. 279–293, 2016.

[20] Erna Hikmawati and Kridanto Surendro. How to Determine Minimum Support in Association Rule. In Proceedings of the 2020 9th International Conference on Software and Computer Applications, ICSCA 2020, page 610, New York, NY, USA, 2020. Association for Computing Machinery.

[21] P. Fournier-Viger, C.-W. Wu and V. S. Tseng, "Mining Top-k Association Rules," in *Leila Kosseim and Diana Inkpen, editors, Advances in Artificial Intelligence*, pp. 61–73, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[22] R. Phalasuek and B. Thanomchayathawatch, "A Family Model for Older People Care," *The Southern College Network Journal of Nursing and Public Health*, vol.4, no. 3, pp.135–150, 2017.

[23] P. Chuanchom, T. Chantuk and P. Siriwong, "Job Styles for Older Workers," *VRU Research and Development Journal Humanities and Social Science*, vol. 13, no. 1, pp. 107–116, 2018.

[24] Ministry of Labour, Thailand. Notification of the Ministry of Labour: Request for Cooperation and Support for the Elderly to Have a Job, 2019. Announced on 8 March 2019.

[25] P. Prasartkul, S. Chuanwan and K. Thianlai. Elderly: Inner People to Be Marginalised. In Kulphawan Chansara and Krittaya Achavanitkul, editors, *Population and Society 2012: Marginalized Population and Fairness in Thai Society*, pp. 105–124. Publication in 8th Academic Conference: Population and Society, Nakhon Pathom: Population and Social Publishing, Institute for Population and Social Research, Mahidol University, 2012.

[26] K. Leawsuwan, T. Rattanachodpanich, K. Tisayaticom, W. Patcharanarumol, S. Limwattananon and C. Limwattananon, "Situation of Hospitalization and Having Carers During Recovery Period at Home Among Elderly People in thailand," *Journal of Health Systems Research*, vol. 11, no.2, pp. 248–256, 2017.

Nichnan Kittiphattanabawon received a bachelor's and master's degree in Computer Science from Rangsit University, Thailand, in 1993 and Prince of Songkla University in 1999, respectively, and a doctoral degree in Technology (International Program) from Sirindhorm International Institute of Technology, Thammasart University, Thailand in 2012. She is currently an Assistant Professor in Information Technology at the School of Informatics, Walailak University, Thailand. Her current research interests include text mining, data mining, document relation discovery, and association analysis.