Forecasting Elderly Well-Being through Decision Tree Modeling Techniques: Integrating Google Maps for Community Engagement in Bang Chakreng, Samut Songkhram Province, Thailand

Sudsawart, J.,[1] Korsanan, N.,[1] Pochanakul, K.[2] and Wattanaprapa, N.[3*]
[1]College of Allied Health Sciences, Suan Sunandha Rajabhat University, Thailand
E-mail: jirawat.su@ssru.ac.th, s61122236003@ssru.ac.th
[2]Faculty of Science and Technology, Phranakhon Rajabhat University, Thailand
E-mail: kullaphat.p@pnru.ac.th
[3]Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand
E-mail: natcha.wa@ssru.ac.th*
Corresponding Author

DOI: https://doi.org/10.52939/ijg.v20i10.3625

Abstract
This research is a quasi-experimental study focusing on the application of Decision Tree modeling techniques. The objectives are as follows: (1) to study the well-being data of the elderly in the Bang Chakreng community, Samut Songkhram Province; (2) to develop a predictive model and apply it to generate community health statistics using decision tree modeling; and (3) to incorporate Google Maps for managing data on community engagement. The sample group consists of 174 elderly individuals from the Bang Chakreng community, with the analysis focusing on four factors: overall health (Hel.), hygiene (Hyg.), environment (Env.), and economy (Eco.). The data was divided into training and testing sets using K-folds cross-validation and percentage split techniques. The results indicate that the optimal model yielded an accuracy rate of 91.40%. The predictive data can be utilized to plan and improve the well-being of the community moving forward.

Keywords: Bang Ja Kreng Community, Decision tree Model, Forecasting, Elderly Well-being, Google Map

1. Introduction

In this century, global aging is a phenomenon occurring worldwide due to declining birth rates in various countries, coupled with increased life expectancy. The United Nations has assessed the global population and noted that, for the first time in history, the majority of people are living into their 60s. Currently, 125 million people are aged 80 years or older, with 80% of the elderly population classified as middle-income. This aligns with the study by Christian Lindmeier, which states that global life expectancy has increased. By 2020, the number of people aged 60 years and older outnumbered children under 5 years old. Furthermore, by 2050, the global population aged 60 years and older is expected to reach approximately 2 billion, with 80% of these elderly individuals residing in low- to middle-income families [1].

The ASEAN region has a population of approximately 606 million people, accounting for one-tenth of the world's population. Based on the definition of an aging society using a percentage as an indicator, the population structure in 2013 shows that Singapore, Thailand, and Vietnam have already entered this category [2]. According to population projections for Thailand from 2010 to 2040 by the Foundation of Thai Gerontology Research and Development Institute, the birth rate of newborns has been continuously declining each year, while the elderly population has significantly increased. This is due to advancements in medical technology and the widespread availability of public healthcare services, leading to a growing number of elderly individuals. In 2013, the population of Thailand was 64.6 million, with 9.6 million classified as elderly. It is estimated that by 2030, the number of elderly people will rise to 17.6 million (26.3%), and by 2040, it will reach 20.5 million (32.1%) [3]. This shift in population structure has positioned Thailand as an aging society, where more than 10% of the population is aged 60 and above. This change has significant implications for the management of elderly healthcare by the government.

As the elderly population increases and the working-age population declines, the government faces rising healthcare costs while revenue trends may decline. Therefore, even though the elderly population is growing, maintaining good health among this group could help reduce future healthcare costs for the state.

Data mining is the process of working with large datasets to identify patterns and hidden relationships within the data for analysis and evaluation. It transforms simple data storage into a more structured form, such as a database, enabling efficient data retrieval and utilization [4]. Data mining techniques vary depending on the selection process, where data is relocated to the correct position based on comparison and swapping instructions in the merge algorithm. Ultimately, the pivot is moved to its correct position, and this algorithm operates iteratively [5]. For example, classification is a model used for categorizing data based on attributes to assign data to predetermined groups (classes). This process requires a portion of the data for the training set and another portion for the testing set, utilizing a decision tree algorithm.

Data mining techniques were employed to improve the quality of education in the Faculty of Engineering. Their study addressed the issue of poor academic performance among students, which stemmed from several factors, such as lack of engagement, inadequate exam preparation, and mismatches between the students' abilities and the courses they enrolled in. Their model was designed to predict grade trends in the following semester, offering students guidance on course selection and study habits [6]. Additionally, the use of Google Maps to assist in field visits and analyze the probability of community health status corresponds with the research by [7][8] and [9], which examined the effectiveness of healthcare services for the elderly in performing daily activities. The study applied Google Maps with symbols and color indicators to represent the level of the elderly's ability to perform daily activities, aiding in health communication and prioritizing home visits by healthcare staff and village health volunteers.

Well-being arises from actions, behaviors, and activities that individuals initiate and continuously engage in to maintain their health. This includes health promotion, disease prevention, medical treatment, and health rehabilitation. In terms of hygiene, it involves assessing various situations in the community; in terms of the environment, it relates to household drainage systems; and in terms of the economy, it involves income, work, and supplementary occupations. For the elderly to effectively care for themselves, they must possess knowledge and understanding of well-being before falling ill, enabling them to maintain their strength and learn how to prevent illness. Predicting future events and preparing for or mitigating the severity of natural disaster impacts is crucial [10] and [11]. When illness does occur, they should also know how to treat themselves until they recover. As outlined, the researcher emphasizes the importance of elderly well-being arising from self-care behaviors. The study aims to identify factors influencing the health behaviors of the elderly, as most health issues originate from personal behavior. By understanding these behaviors, potential health problems can be addressed. Information technology can now be utilized to analyze data from communities with good well-being, providing guidance and examples for other members of the community.

2. Method

2.1 Study Area

Bang Chakreng is a sub-district located in Samut Songkhram Province, Thailand (Figure 1), characterized by its rich cultural heritage and close-knit community. Nestled along the Tha Chin River, this area is known for its picturesque landscapes, traditional livelihoods, and vibrant local markets. The population is predominantly elderly, reflecting broader demographic trends in Thailand, where an increasing number of individuals are living into their senior years. This aging population presents unique challenges and opportunities for healthcare and community services, making Bang Chakreng an important focal point for studies aimed at enhancing elderly well-being. The integration of technology and data-driven approaches in this context can significantly improve healthcare delivery and community engagement, ultimately contributing to the overall quality of life for its residents.

2.2 Study Design

This research is a quasi-experimental study (Quasi-Experimental Design) on data mining using decision tree modeling techniques. The data was obtained from the Subdistrict Information System Development Project (Big Data) under the Rajabhat Strategy for Local Development of Suan Sunandha Rajabhat University for the fiscal year 2021. The dataset consisted of 13,920 records and 15 attributes. The Waikato Environment for Knowledge Analysis (Weka), a free software tool for machine learning and data mining developed by the University of Waikato, New Zealand, was used for analysis. The data analysis process followed the CRISP-DM (Cross Industry Standard Process for Data Mining) framework, which was developed in 1996 through the collaboration of three companies: Daimler Chrysler, SPSS, and NCR.

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) |

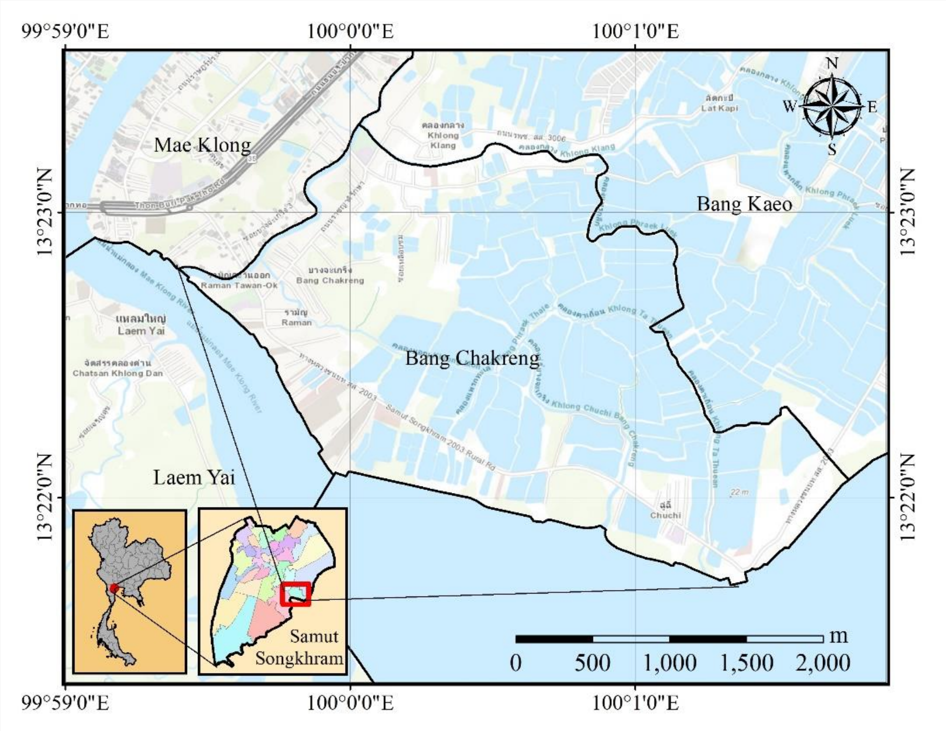

Figure 1: Bang Chakreng, Muaen Samut Songkhram district, Samut Songkhram province

Figure 2: The CRISP-DM process

The CRISP-DM framework comprises six phases, as shown in Figure 2. The data analysis process using the Cross Industry Standard Process for Data Mining (CRISP-DM) is as follows:

1. *Business Understanding:* This involves classifying the well-being data of the elderly, which includes overall health, hygiene, environmental, and economic information. The data is categorized for 174 elderly individuals from different villages, including Village 1 (Ban Raman West), Village 2 (Ban Bang Chakreng), Village 3 (Ban Khlong Klang), Village 4 (Ban Chuchi), and Village 5 (Ban Raman East).

ISSN: 1686-6576 (Printed) | ISSN 2673-0014 (Online) | © Geoinformatics International

2. Data Understanding: This stage focuses on comprehending the dataset under study. The dataset consists of 13,320 records and 15 attributes related to the well-being of elderly individuals in the Bang Chakreng community.

3. Data Preparation
This involves preparing the data before actual processing.

3.1 Data Cleaning: This step involves filtering out irrelevant data by removing records with missing values and removing records with errors or outliers

3.2 Dimensionality Reduction: Since the original dataset contains 15 attributes, some of which are irrelevant to the analysis, dimensionality reduction is applied using the Wrapper approach with Forward Selection. After this process, the number of attributes is reduced from 15 to 6 for further analysis, as shown in Table 1.

Table 1: Details of attributes used in the processing

Abbreviation	Description
Num	The respondent's sequence number
Hel.	Overall health information
Hyg.	Overall hygiene information
Env.	Overall environmental information
Eco.	Overall economic information
Results.	Outcome

4. Modeling: This stage involves identifying useful patterns from the available data using appropriate techniques and algorithms. In this study, the classification technique was chosen, with the Decision Tree algorithm selected for analysis.

5. *Model Evaluation*: This step assesses whether the results obtained from the model are appropriate. The evaluation of the classification model's performance involves four key metrics:

Accuracy (*AC*) represents the number of correctly predicted instances across all classes. It can be determined from Equation 1:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 1

Precision (*PC*) indicates the percentage of correctly predicted positive instances. The precision is defined in Equation 2:

$$PC = \frac{TP}{TP + FP}$$

Equation 2

Recall (*RC*) measures how many of the actual positive instances were correctly predicted by the model. Recall is determined from Equation 3:

$$RC = \frac{TP}{TP + FN}$$

Equation 3

F-measure is a metric used to evaluate the performance of a Machine Learning model which is determined from Equation 4:

$$F - measure = 2\frac{PC \times RC}{PC + RC}$$

Equation 4

Where:

TP is an outcome where the model correctly predicts the positive class
TN is an outcome where the model correctly predicts the negative class
FP is an outcome where the model incorrectly predicts the positive class.
FN is an outcome where the model incorrectly predicts the negative class

K-fold Cross-validation Testing: This method is used to evaluate the performance of a model by dividing the dataset into multiple parts, referred to as *K*-folds. In *K*-fold cross-validation, the data is split into *K* equal parts [12] and [13]. A subset of (*K-1*) parts is used as the training set, and the remaining part is used as the testing set. This process is repeated until each part has been used as a testing set, ensuring that all data subsets have a chance to be evaluated. Percentage Testing: This method divides the data into two parts. The first part, known as the training set, is used to build the model and contains no less than 60% of the data. The second part, called the testing set, is used to evaluate the model and contains no more than 40% of the data.

6. Deployment: This step involves applying the results obtained from the Decision Tree model in a practical context. The outcomes can be presented in the form of reports or dashboards, which are ready to be utilized for strategic planning, decision-making, and implementation in an efficient manner.

2.3 Ethical Consideration

This research proposal was reviewed and approved by the Suan Sunandha Rajabhat University Ethics Committee. The Ethics Committee issued certificate number COA.1-095/2021 and approved the implementation of this research proposal.

3. Results

This research collected data from a sample group of 174 individuals. The results are summarized as follows: The researcher analyzed and presented the demographic characteristics of the sample group using frequency and percentage values. This includes personal information such as gender, age range, education level, and occupation, among others.

3.1 General Information of the Respondents

According to Table 2, the majority of respondents were female, accounting for 104 individuals (59.77%), while 70 respondents (40.23%) were male. The largest age group was between 60 and 69 years, comprising 104 individuals (59.77%), followed by 40 individuals (22.99%) in the 70-79 age group, 20 individuals (11.94%) in the 80-89 age group, and 10 individuals (5.30%) aged between 90 and 99 years. In terms of education level, most respondents had completed lower primary education (62 individuals, 35.63%), followed by upper primary education (30 individuals, 17.24%), lower secondary education (20 individuals, 11.49%), upper secondary education (20 individuals, 11.49%), and diploma level (22 individuals, 12.65%). A small proportion had attained a bachelor's degree (10 individuals, 5.75%), while another 10 individuals (5.75%) had not received any formal education. Regarding occupation, the majority were engaged in farming/agriculture (74 individuals, 42.53%), followed by those who were unemployed (40 individuals, 22.99%). General contractors and service providers accounted for 30 individuals (17.24%), while laborers and those involved in trade or private business each made up 15 individuals (8.62%).

3.2 Predicting the Well-Being of the Elderly using Decision Tree Modeling

After testing the model by splitting the dataset to identify a Decision Tree model with high accuracy, the models were compared to select the one with the highest accuracy for predicting the well-being of elderly individuals in the Bang Chakreng community, Samut Songkhram Province. It was found that some of the selected factors may be correlated with the entire dataset, potentially affecting the model’s accuracy in prediction. Therefore, the data was analyzed using decision tree techniques through specialized software.

Table 2: Personal information of the respondents

Personal Information	Frequency (n=174)	Percentage (%)
Gender		
Male	70	40.23
Female	104	59.77
Age Group (years)		
60-69	104	59.77
70-79	40	22.99
80-89	20	11.94
90-99	10	5.30
Education Level		
Primary Education (Lower)	62	35.63
Primary Education (Upper)	30	17.24
Lower Secondary Education	20	11.49
Upper Secondary Education	20	11.49
Diploma (Vocational Certificate)	22	12.65
Bachelor's Degree	10	5.75
No Education	10	5.75
Occupation		
Unemployed	40	22.99
Laborer	15	8.62
Farming/Agriculture	74	42.53
Trade/Private Business	15	8.62
Employee/State Enterprise/Worker	0	0
General Contractor/Service Provider	30	17.24

Table 3: Experimental results

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
Cross Validation 5 Folds	84.50	87.70	84.50	84.80
Cross Validation 10 Folds	83.70	87.40	83.70	83.90
Cross Validation 50 Folds	85.60	88.40	85.60	85.90
Percentage Testing 60%	85.70	89.40	85.70	86.30
Percentage Testing 70%	84.60	89.70	84.60	85.40
Percentage Testing 80%	91.40	92.30	91.40	91.50

Each factor was removed one by one to observe its impact on accuracy. If a factor had no relationship with the dataset, removing it would result in higher accuracy. However, if removing a factor led to a decrease in accuracy, it indicated that the factor was related to the entire dataset and could not be excluded. The researcher, using data mining techniques with classification methods, found that the attributes indicating the well-being of the elderly in the Bang Chakreng community, Samut Songkhram Province, consisted of six attributes: Num., Hel., Hyg., Env., Eco., and Results. The accuracy of the experimental results for these attributes is shown in Table 3.

According to Table 3, the model development results compare the accuracy obtained from six designed methods, including three percentage testing methods (60:40, 70:30, and 80:20) and three cross-validation methods (5-fold, 10-fold, and 50-fold). The results show that the percentage testing methods of 60:40, 70:30, and 80:20 yielded the highest accuracy rates of 85.70%, 84.60%, and 91.40%, respectively. For the cross-validation methods, the highest accuracy rates were 84.50%, 83.70%, and 85.60% for 5-fold, 10-fold, and 50-fold cross-validation, respectively, with consistent results for the 10-fold and 50-fold datasets. When comparing the six testing methods, the model development results indicate that the percentage testing method with an 80:20 split provided the highest accuracy rate at 91.40%. Therefore, the model with the highest accuracy can be used to make precise predictions.

3.3 Application of Google Maps in Managing Community Participation Data

Currently, information technology plays a crucial role in managing and storing data in local areas, ensuring it aligns with the sample groups and enhances the accuracy of community data analysis and prediction. The research team recognized the importance of having a unified data collection approach, which would allow for the integration of such technology into planning and reporting processes by village. This approach provides a clear view of health trends within each village [14][15] and [16]. The prediction results were consistent with expectations, showing that information technology can be effectively used for home healthcare management, improving self-care among patients with support from family members. Field data collection during home visits can help prioritize services provided by healthcare personnel and teams. This enhances the efficiency of visits by students and healthcare professionals, resulting in better home healthcare services for both recipients and community healthcare teams. The use of Google Maps in the Bang Chakreng community (See Figure 3), located in Mueang Samut Songkhram District, Samut Songkhram, is crucial for visualizing and managing community participation data. The map can serve as a detailed geographical tool to:

1. *Identify Key Locations*: The map can be used to mark important locations within Bang Chakreng, such as the households of elderly residents, healthcare facilities, and areas where community participation is needed. By using Google Maps, local authorities and healthcare teams can gain a visual understanding of where to focus their efforts.
2. *Enhance Home Visit Planning*: As discussed in section 3.3, one of the primary applications of this technology is for planning home healthcare visits. By utilizing the map, healthcare teams can optimize their routes and prioritize visits based on the geographic distribution of the elderly population in Bang Chakreng. This ensures more efficient delivery of healthcare services, reducing time and resources.
3. *Monitor Community Engagement*: The map can also be used to monitor which areas of the community have higher or lower participation in healthcare programs. This allows healthcare officials to identify gaps in community involvement and target specific areas for further engagement.
4. *Real-time Data Integration*: Using Google Maps, the health status and participation data of the elderly population can be updated in real-time, enabling continuous monitoring of the community's health trends. This information is crucial for adapting healthcare strategies based on the needs of specific sub-areas within Bang Chakreng.

Figure 3: Google maps in the Bang Chakreng community

In summary, the integration of Google Maps in Bang Chakreng, Mueang Samut Songkhram District, allows healthcare teams and researchers to manage and analyze community data geographically. This tool not only aids in improving healthcare access and planning but also ensures that healthcare interventions are targeted efficiently based on location-specific needs within the community.

4. Discussion

The findings from this research indicate that the decision tree model (Decision Trees) is an effective tool for predicting factors influencing the well-being of the elderly in Bang Chakreng, Samut Songkhram Province. The use of 50-fold cross-validation in the model development process resulted in the highest accuracy, with an accuracy rate of 85.60%, a precision of 88.40%, and an F-measure of 85.90%. This demonstrates the model's ability to handle large datasets and analyze multiple interrelated factors, making it more accurate than other models tested. The analysis also revealed that factors such as age, occupation, residence, and overall health status were critical in determining the well-being of the elderly. Removing any of these factors led to a decrease in the model's accuracy, emphasizing their interconnectivity and significance. These findings are consistent with previous research, which utilized data mining techniques to predict academic performance. Similarly, the decision tree model used in this research allowed for a comprehensive analysis of multiple factors, showing that advanced algorithms can improve predictions in various domains, including health and education.

Furthermore, the application of Google Maps in this study was beneficial for managing and analyzing community data. It facilitated the visualization of health trends within the community and supported more effective planning for healthcare services. This approach aligns with the study by Chakrapong Thisara, which highlighted the advantages of using Google Maps to assist in prioritizing home visits and communicating health information. The ability to map health data geographically is particularly valuable in communities like Bang Chakreng, where access to healthcare resources can be improved through strategic planning based on location-specific information. The integration of decision tree algorithms and geospatial technologies like Google Maps has proven to be an efficient strategy for enhancing healthcare services in the community. By using data mining techniques, healthcare teams can better understand the complex relationships between various health factors and tailor interventions accordingly. Moreover, the visual representation of health data through Google Maps provides actionable insights that can optimize resource allocation and service delivery in real time.

5. Conclusion

From the model development results of factors influencing the well-being of the elderly using decision tree data mining techniques, the Decision Tree algorithm was used to analyze and identify the factors affecting the well-being of the elderly in Bang Chakreng, Samut Songkhram Province. The data comprised 13,320 records collected from field visits in Bang Chakreng. The research findings can be summarized as follows: the analysis of all related factors showed that every factor was interrelated. When selecting factors for inclusion, cross-validation methods (5-fold, 10-fold, 50-fold) were used to determine the optimal data learning model. The model developed using 50-fold cross-validation produced the most accurate results. Subsequently, the data was randomly split for percentage testing with 60:40, 70:30, and 80:20 ratios. The analysis revealed that the factors of age, occupation, residence, and overall health information were significant; removing any of these factors reduced accuracy.

Therefore, it was concluded that all factors were interconnected and had a mutual influence. Upon analyzing and comparing all models using the Decision Tree technique, the model that provided the highest accuracy was the Cross Validation 50 Folds model, with an accuracy of 85.60%, a precision of 88.40%, a recall of 85.60%, and an overall F-measure of 85.90%. This indicates that the Cross Validation 50 Folds model using the Decision Tree technique was the most accurate compared to the other models. The model's use of 50 subsets of data for alternating training and testing resulted in greater analytical precision, producing higher accuracy and reliability.

Acknowledgment

We sincerely thank the reviewers for their insightful comments and valuable suggestions, which have significantly contributed to the improvement of this study. We also extend our gratitude to the College of Allied Health Sciences, Suan Sunandha Rajabhat University, Samut Songkhram Province, Thailand, for their generous support throughout the research process.

References

[1] Lindmeier, C., (2014). *WHO Communications Officer. WHO Statement on the Seventh Meeting of the IHR Emergency Committee Regarding MERS-CoV*. World Health Organization. [Online]. Available: https://www.who.int/news/item/01-10-2014-who-statement-on-the-seventh-meeting-of-the-ihr-emergency-committee-regarding-mers-cov. [Accessed: Jul. 5, 2024].

[2] Thaveesit, S., Sanpuwan, M. and Chuanwan, S., (2013). *Population and Society 2013: Population and Society in ASEAN, Challenges and Opportunities*. Institute for Population and Social Research, Mahidol University.

[3] Foundation of Thai Gerontology Research and Development Institute. (2013). *Situation of the Thai Elderly 2013*. [Online]. Available: https://thaitgri.org/?wpdmpro=situation-of-the-thai-elderly-2013. [Accessed: Jul. 5, 2024].

[4] Han, J. and Kamber, M., (2006). *Data Mining Concepts and Techniques (2nd ed.)*. Morgan Kaufmann. United States of America.

[5] Ketchaya, S. and Rattanatranurak, A., (2022). Analysis and Optimization of Dual Parallel Partition Sorting with OpenMP. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2022.01.003.

[6] Waimai, K., Songsiri, C. and Raktammanon, T., (2001). Using Data Mining Techniques to Improve the Quality of Education in the Faculty of Engineering. *NECTEC Journal*, Vol. 3, 134-142.

[7] Dulyawit, P. and Pijitra, J., (2020). Annual Rainfall Model by Using Machine Learning Techniques for Agricultural Adjustment. *Journal of Advances in Information Technology*, Vol. 11(3), 161-165. https://www.jait.us/uploadfile/2020/0714/20200714015750558.pdf.

[8] Kwangsawasdi, A., Nusawasdi, P., Khongmuak, W., Tipyakulruk, P. and Sanghanun, B., (2019). Stress Level Prediction System Using Decision Tree Diagrams. *Rattanakosin Journal of Science and Technology*, Vol. 1(2), 13–26. https://ph02.tci-thaijo.org/index.php/RJST/article/view/239865.

[9] Boonma, R. and Chirawichitchai, N., (2020). Classification of Diabetic Patients Using Data Mining Techniques and Feature Selection from Data Correlations. *PKRU SciTech Journal*, Vol. 3(2), 11–19.

[10] Pongsangworn, V., Thinsungnoen, T. and Thinsungnoen, M., (2018). Development of a Model of Factors Affecting Diabetes Using Decision Tree Techniques. *Journal of Science and Technology Mahasarakham University*, Vol. 1(1), 1-8. https://ph02.tci-thaijo.org/index.php/jstrmu/article/view/245737.

[11] Abirami, S. and Chitra, P., (2020). Energy-Efficient Edge-Based Real-Time Healthcare Support System. *Advances in Computers*, Vol. 117(1), 339–368. https://doi.org/10.1016/bs.adcom.2019.09.007.

[12] Gladchuen, R. and Saenraj, C., (2018). A Comparison of Algorithm Performance and Appropriate Feature Selection for Predicting the Academic Performance of Vocational Students. *Research Journal Rajamangala University of Technology Thanyaburi*, Vol. 17(1).

[13] Thisara, J., (2016). *Effectiveness of Health Care Services Concerning Activities of Daily Living Among the Elderly by Applying Google Map Program*. Master's Thesis, Thammasat University.

[14] Taniar, D., (2007). *Mining Graph Data Mining Technologies and Applications*. Monash University.

[15] La-orsirikul, K., Ratchaprapapornkul, P. and Kao-lean, S., (2024). Comparison of Efficiency for Imbalanced Data Classification via Simulation. *Information Technology Journal KMUTNB*. Vol. 20(1).

[16] Jongmuenwai, B., (2015). Comparison of Factors of Chronic Diseases in the Elderly Using J48 and Naivebayes Algorithms: A Case Study of Phoklang Public Health, Nakhon Ratchasima. *Computer Science and Information Technology Project Journal*, Vol. 1(2),43-51.