Prediction of Waiting Time in One-Stop Service

Yotravee Sanit-in and Kanda Runapongsa Saikaew

Abstract—**Customers of several popular services need to wait idly for a long time. These services usually have a limited capacity and can only serve a small number of customers at a time. It is impossible for customers to receive the service without waiting at all; thus, it will be advantageous for clients to know the approximate waiting time which they may choose to do other activities instead of standing in a service queue. This article proposes and evaluates approaches to predict the waiting time before a customer receives the service. Three approaches of waiting time prediction have been implemented and compared. These approaches include Queueing Theory, Average time, and Random Forest. The experimental results indicated that the supervised learning algorithm, Random Forest, achieved the highest accuracy at 85.76% of ear nose and throat clinic dataset and 81.7% of Khon Kaen University post office dataset. This article also investigated feature importance and found that the number of waiting queues was the most critical feature in waiting time prediction.**

Index Terms—**Machine learning, queueing theory, random forest, waiting time prediction.**

I. INTRODUCTION

Currently, people lifestyle and activities are always in a rush. Thus, their time is most valuable [1]-[3]. However, people often have to go to some over-crowded services, such as excellent health clinics or post offices. These services usually take customers a long time in waiting before receiving service. To make services more convenient and timely, service providers set up a one-stop service which combines the variety of services to only one place [4].

Nowadays, several one-stop services, such as post offices, banks or restaurants have high demand among many customers. Many times, customers have to wait for a long time before being served. Such a long waiting time is likely to affect customers' satisfaction. Waiting leads to negative relationship between customers and the service venues, and it affects the first impression that the customers have toward the services. It is clear that the first relationship influences the service [1].

Several recent research studies have attempted to tackle the problem of service waiting time. Rattanapan *et al.* [5] investigated on how to reduce waiting time by using Lean

Thinking and ECRS techniques (Eliminate, Combine, Rearrange, and Simplify) and the simulation model at a dental clinic. They concluded that the simulation model of the dental clinic could reduce waiting time for the services. Kumar *et al.* [6] designed an algorithm for waiting time prediction by using the average time of outpatient treatment, the implemented web-based and mobile application, and the algorithm for waiting time prediction. However, there was no experimental result reported. Rarh *et al.* [7] implemented the system to solve the problem of the waiting time of the restaurant table reservation by using time-series prediction. The system would inform waiting time to customers and help the provider to manage services more efficiently. However, there was no waiting time prediction accuracy reported. Carvalho *et al.* [8] developed a system for waiting time prediction of a store by using time-series prediction with Linear Regression technique and the features of the store service. The above discussed related work had approaches to solve the problem by using the simulation, the calculation of the past average time, or using time-series prediction. This proposed article differs from such related work by using classification. We chose classification instead of regression because it would take an excessively large amount of training data to predict waiting time in a regression manner.

One of the early work that also employed classification technique was the work proposed by Mourão *et al.* [9] which proposed the four techniques that included Queueing Theory, Deep Learning, Gradient Boost Machine, and Random Forest for predicting the waiting time of a bank, but the features that built the predictive model were specific to the bank services.

Although most of the studies [6]-[8] implemented the system for waiting time prediction, but they did not study some critical features affecting waiting time prediction, such as the number of waiting customers. On the other hand, the most recent work [9] applied some features for the waiting time classification. However, their work could only apply to only banks in Brazil, not any other services. Furthermore, their prediction result was whether the waiting time was above the certain threshold or not. Notifying the waiting time range would be more useful and preferable.

This article proposes an algorithm for waiting time prediction of any one-stop service by using general features that are not specific to any particular service. To the best of our knowledge, our article is the first that studies and reports the feature importance in the predictive model for waiting time of one-stop service.

The rest of the article is organized as follows: Section II describes the background of this study, Section III explains the methodology in preparing data and the waiting time prediction, and the experimental results are presented and analyzed in Section IV. Finally, the conclusions are drawn in Section V.

Yotravee Sanit-in and Kanda Runapongsa Saikaew are with the Faculty of Engineering, Khon Kaen University, Thailand (Corresponding author: Kanda Runapongsa Saikaew; e-mail: krunapon@kku.ac.th).

II. BACKGROUND

The study investigated three approaches for solving waiting time in one-stop service problem. These approaches included Queueing Theory, Average Time, and Random Forest. We chose to study these approaches because Queueing Theory is a traditional method of the prediction of waiting time, Average Time is a quick and easy calculation method, and Random Forest is a classification method that learns data to build the prediction model.

A. Queueing Theory

Queueing theory is a mathematical study of service systems and waiting time of queues in a waiting line. Because the system has the limited resources, all customers cannot enter the services at the same time. Therefore, the waiting queues are created [10]. In queue discipline, the service orders include first come first served (FCFS), last come first served (LCFS), random selection for service (RSS), priority ordering (PRI), and general discipline (GD) [9].

According to Erdelić *et al*. [10], queue model has been explained using three factors, namely, *A/S/c* where *A* denotes the arrival distribution, *S* denotes the service distribution, and *c* denotes the number of services.

The case study in this article, the queue discipline is FCFS, and the queue system is *M/M/c*. Erdelić *et al*. [10] described that *M/M/c* is the multi-server system where *M* is an exponential distribution of arrival and service, respectively, and *c* is the number of multi-services.

For *M/M/c* queue system, the Utilization Factor (*U*) is a ratio of arrival, service rate, and service point number (2),

$$\rho = \frac{\lambda}{\mu} \tag{1}$$

$$U = \frac{\rho}{c} = \frac{\lambda}{c \cdot \mu} \tag{2}$$

where the mean arrival rate (λ) is the number of customers arriving at the service point in one minute, the mean service rate (μ) is the number of customers receiving the service in one minute, and *c* is the number of service point at the time [10].

Erdelić *et al*. [10] reported that if $U < 1$, the system is stationary. The measurement of the effectiveness of the queue system can be computed as follows:

The probability of no customers in the system (P_0) can be computed by using Equation (3).

$$P_0 = \left[\frac{c\rho^c}{c!(c-\rho)} + \sum_{n=0}^{c-1} \frac{\rho^n}{n!} \right]^{-1} \tag{3}$$

The average number of customers in queue for service (L_q) can be computed by using Equation (4).

$$L_q = \frac{c\rho^{c+1}}{c!(c-\rho)^2} \cdot P_0 + \rho \tag{4}$$

The average waiting time of customers in queue for service (W_q) can be computed by taking (L_q) divide by (λ) as illustrate in Equation (5).

$$W_q = \frac{L_q}{\lambda} \tag{5}$$

B. Average Time

The arithmetic mean is the sum of all interested values divided by the number of values [11]. In the case of time statistics, it is called average time.

For the average waiting time, Average Time (*AT*) can be calculated by waiting time and the number of customers,

$$AT = \frac{1}{n}\sum_{i=1}^{n} w_i = \frac{w_1 + w_2 + \cdots + w_n}{n} \tag{6}$$

where w_i denotes the waiting time of the *i*th customer, and *n* is the number of customers.

C. Random Forest

Supervised learning algorithms are machine learning methods that train data to create the model for prediction as regression or classification. In the classification, Random Forest is one of the supervised learning algorithms recommended for building the predictive model because of its high performance and speed [12].

Random Forest is a classification that includes decision trees and ensemble. To combine two algorithms as shown in Fig. 1, a decision tree is an algorithm that divides the data by attributing and building a tree to classify the category of an instance. An ensemble is a classification that divides the data for creating models and integrates the outcome prediction of each model as the class of an instance. In addition, Random Forest divides the data by randomly dividing the features for creating the models [9], [13].
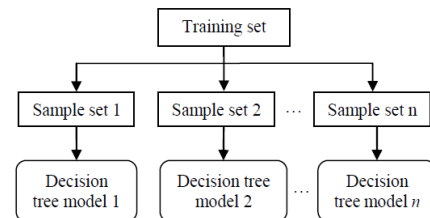

Fig. 1. The random forest model.

In Random Forest, there is the feature importance technique in which mean decrease accuracy is used to evaluate the predictive error and feature significance [14], [15].

III. METHODOLOGY

A. System Overview

The process of waiting time prediction includes data preparation, building a predictive model, and evaluation of predicted outcome as follows in Fig. 2.

From Fig. 2, the flowchart shows the order of the process of waiting time prediction. After datasets were cleaned, the datasets were randomly split into training and testing dataset. Random Forest used training dataset to build a predictive model for classification the waiting time class. Finally, the accuracy of three algorithms: Queueing Theory, Average

Time and Random Forest was tested by waiting time prediction of the testing dataset.
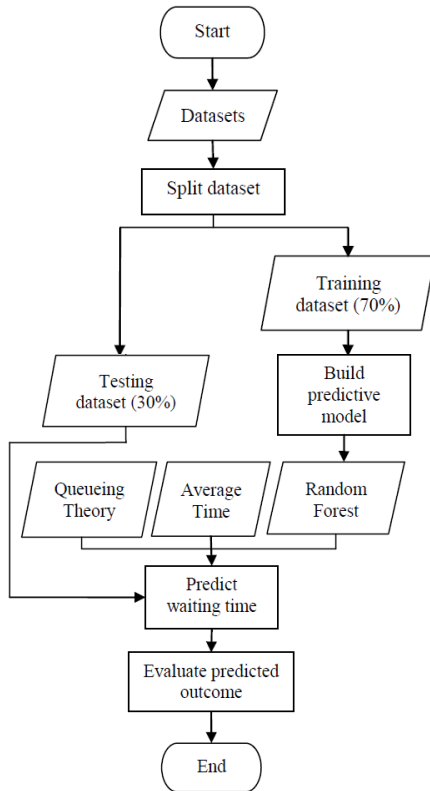

Fig. 2. The flowchart of waiting time prediction.

B. Data Preparation & Analysis

The first data of this study is the queue logs of Khon Kaen University post office from October 16 to November 17, 2017. Based on a mobile application, 3,480 records of data were collected.

The second data of this study is the queue logs of the ear nose and throat clinic, Srinagarind Hospital affiliated to the Faculty of Medicine of Khon Kaen University from June 29 to August 28, 2018. The data were collected 1,348 records by the queue management system of the clinic.

The data cleansing process included the elimination of attributes that were not used to build predictive models. The waiting time type classification from waiting duration in service was shown in Table I and II.

TABLE I: THE WAITING TIME CLASSES OF KHON KAEN UNIVERSITY POST OFFICE DATASET

Classes	Duration (Seconds)
Very short	0 - 60
Short	60 - 120
Medium	120 - 240
Long	240 - 480
Very long	> 480

Table I shows the waiting time classes that were classified from the waiting queue duration. The waiting time between 0 and 60 seconds was considered to be very short, the waiting time between 60 and 120 seconds considered short, and so on. We divided the waiting time class from the min (2 seconds), max (625 seconds), average (69 seconds), and standard deviation (96.3 seconds) values of the waiting queue duration

for the waiting time class of the Khon Kaen University post office dataset.

TABLE II: THE WAITING TIME CLASSES OF EAR NOSE AND THROAT CLINIC DATASET

Classes	Duration (Minutes)
Very short	0 - 20
Short	20 - 40
Medium	40 - 60
Long	60 - 80
Very long	> 80

TABLE III: THE STATISTICS OF WAITING TIME CLASSES OF KHON KAEN UNIVERSITY POST OFFICE DATASET

Classes	Number of Records	Percentage
Very short	2,280	65.52
Short	521	14.97
Medium	468	13.45
Long	174	5
Very long	37	1.06

TABLE IV: THE STATISTICS OF WAITING TIME CLASSES OF EAR NOSE AND THROAT CLINIC DATASET

Classes	Number of Records	Percentage
Very short	723	53.64
Short	379	28.12
Medium	123	9.12
Long	87	6.45
Very long	36	2.67

TABLE V: THE ATTRIBUTES OF DATASETS

Features	Description / Type
Number	The queue number / Nominal – e.g., 1, 101, or 2001.
Shift	The period of the day / Nominal: {morning, afternoon}
Queue Type	The type of queue which customers have been served / Nominal: {mail, parcel, mail and parcel} or {yellow, purple, light green, white, pink, dark green, blue, general, sky blue, elderly people, orange}
Service Point	The customer service point / Nominal: {1, 2, 3, 4} or {1, 2}
Created Queue Hours	The hours when the customer started waiting / Timestamp: {HH}
Created Queue Day of Week	The day of week when the customer entered service / Nominal: {Monday, Tuesday, Wednesday, Thursday, Friday}
Waiting Duration	The service waiting time / Numeric (Seconds or Minutes)
Number of Waiting Queue	The number of customers who were waiting for services / Numeric (Queues)
Arrival Rate	The customer arrival rate at the time / Numeric (customer per minute)
Service Rate	The customer service rate at the time / Numeric (customer per minute)
Waiting Time Class	The target attributes are classified from waiting time duration / Nominal: {very short, short, medium, long, very long}

Table II shows the classification of waiting time classes from the waiting queue duration. The waiting time between 0 and 20 minutes was considered to be very short, the waiting time between 20 and 40 minutes considered short, and so on. The waiting time classes were divided from the min (2

minutes), max (113 minutes), average (25 minutes), and standard deviation (20 minutes) values of the waiting duration of the ear nose and throat clinic dataset.

The data analysis in this section was considered a statistics of queue pattern as shown in Table III and IV.

The statistics of waiting time classes are shown in Table III. For Khon Kaen University post office dataset, the waiting time classes were classified from the waiting duration of the 3,480 records. The percentages of records in the five waiting time classes are as follows: very short 65.52%, short 14.97%, medium 13.45%, long 5%, and very long 1.06%.

Table IV shows the statistics of waiting time classes. the waiting duration of the ear nose and throat clinic dataset was classified to waiting time classes. The percentages of records in the five waiting time classes are as follows: very short 53.64%, short 28.12%, medium 9.12%, long 6.45%, and very long 2.67%.

The models were built from the features as shown in Table V. Only Average Time used the waiting time feature (Waiting Duration) for finding the arithmetic average.

C. Waiting Time Prediction

The datasets of this study were randomly split into two datasets: training (70%) and testing (30%) dataset. The Khon Kaen University post office dataset has training dataset: 2,436 records and testing dataset: 1,044 records. The ear nose and throat clinic dataset has training dataset: 943 records and testing dataset: 405 records.
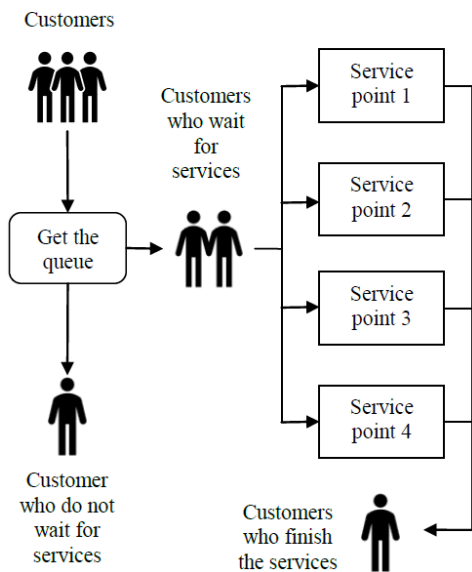

Fig. 3. The queue system diagram of Khon Kaen University post office.

For the waiting time prediction of Queueing Theory, the dataset is an *M/M/c* queueing system where *c* is the number of service points that can be run at the same time. The queueing system of Khon Kaen University post office is shown in Fig. 3. We considered variables following the equation (1)-(4) for predicting the waiting queue time. The waiting times were predicted by Equation (5) using the arrival rate and the average of customers in the queue from (4).

The queueing system of ear nose and throat clinic dataset is an *M/M/c* queueing system where *c* is two service points following Fig. 4. We considered variables following the equation (1)-(5) for predicting the waiting queue time.

Another method for predicting the waiting time is the calculation of the average waiting time in each day and hours. For the dataset of the study, the waiting time prediction followed the Equation (6), with the data divided into each day and time period to find the average of the waiting time.
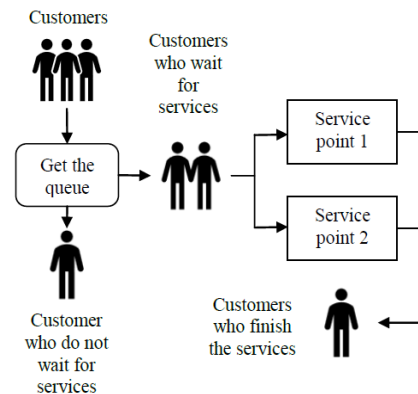

Fig. 4. The queue system diagram of ear nose and throat clinic.

The classification of Random Forest used R language to build a predictive model. The waiting duration was excluded from the training dataset for creating a model based on Random Forest because the classification used the waiting time as the target attribute. The waiting time duration was only used for calculating the average waiting time.

To find the most effective Random Forest model with a dataset, we experimented algorithm in the training model by configuring the number of trees from 100 to 500, by 50 to find out the most optimal number of trees. The values of other parameter settings were default values in the library provided in R language.

IV. EXPERIMENTAL RESULTS

The experimental results include the accuracy of three methods and feature importance of attributes affecting the waiting time prediction.

A. Accuracy

TABLE VII: THE ACCURACY OF KHON KAEN UNIVERSITY POST OFFICE DATASET

Methods	Accuracy (%)
Queueing Theory	65.23
Average Time	64.94
Random Forest	81.70

For the accuracy evaluation of Queueing Theory and Average Time, the accuracy is computed based on the predicted time and the waiting time class as in Table I and II. On the other hand, Random Forest accuracy is reported based on the output from a function in the randomForest library in R language. The accuracy of three methods are shown in Table VII and VIII.

Table VII shows the precisions of the three methods from the Khon Kaen University post office dataset. Average Time has the least accuracy at 64.94%, Queueing Theory has the better accuracy at 65.23%, and Random Forest is the most effective model with a precision of 81.70%, and the number of trees is 450 trees.

TABLE VIII: The Accuracy of Ear Nose and Throat clinic Dataset

Methods	Accuracy (%)
Queueing Theory	-
Average Time	68.89
Random Forest	85.76

The precisions of the three methods are shown in Table VIII. Queueing Theory was not applied to waiting time prediction of ear nose and throat clinic dataset, Average Time has the least accuracy at 68.89%, and Random Forest is the most effective model with a precision of 85.76%, and the number of trees is 250 trees.

Regarding the accuracy when using Queueing Theory, we consulted variables following the equation (1)-(5) for the waiting queue time prediction. The Utilization Factor (U) was found more than 1 from the rate of arrival rate (λ), service rate (μ), and c. Based on the results, the waiting time will tend to infinity and the queueing system is not stationary. Queueing Theory cannot be applied to the ear nose and throat clinic dataset because the queueing system and queue priority are complex and the service pattern is irregular.

From Table VII and VIII, Random Forest model achieves the highest accuracy because Random Forest is a supervised learning algorithm that differs from Queueing Theory and Average Time which are based on mathematical formulas.

B. Feature Importance

In Random Forest model, feature importance is investigated to find out the features that affect the outcome prediction the most.

The essential features that were used for prediction of waiting time in one-stop service are illustrated via the mean decrease accuracy graph. The mean decrease accuracy graph shows the value of the features that are probably the primary cause of accuracy in prediction.
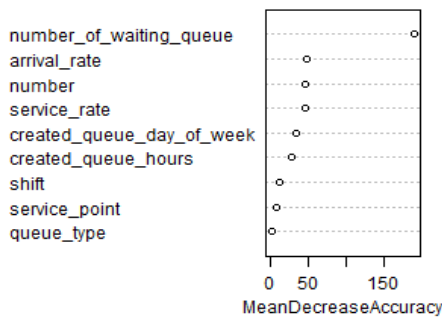

Fig. 5. The mean decrease accuracy of Khon Kaen University post office dataset.

From Fig. 5, for Khon Kaen University post office dataset, the number of waiting queues is the most effective feature in the classification of the waiting time. The next essential features are the arrival rate, the queue number of waiting time each day, the service rate, the created queue of the week, and the created queue in hours, respectively. On the other hand, the shift period of the day is less relevant to the queue system. The least significant features are the service point and queue type.

From Fig. 6, for ear nose and throat clinic dataset, the most effective feature in the waiting time classification is the number of waiting queues. The next significant features are

the service rate, the arrival rate, the created queue of the week, the created queue in hours, the queue number of waiting time each day, and queue type, respectively. On the other hand, the service point is less relevant to the queue system. The least significant feature is the shift period of the day.
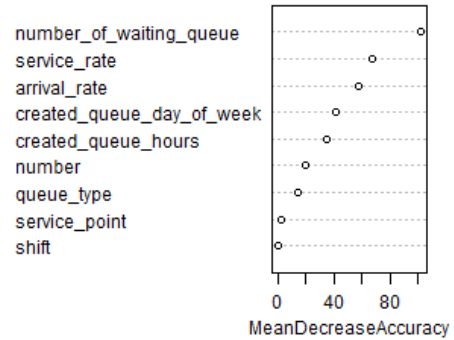

Fig. 6. The mean decrease accuracy of ear nose and throat clinic dataset.

From the feature important of both datasets, the number of waiting queues is the most important feature of classification and building the predictive model because the number of waiting queues is the feature that indicates the quantity of the service at the time. The features that show the difference of the service pattern, queueing system, and queue priority between the Khon Kaen University post office and the ear nose and throat clinic are the next importance features of each dataset. The arrival rate and queue number of the Khon Kaen University post office dataset indicate the service has the fixable pattern. The service and arrival rate of the ear nose and throat clinic dataset show that the waiting time of the queueing system depends on the service rate of service providers and the service pattern is not fixable.

V. CONCLUSION

The article proposes and evaluates the model for the waiting time prediction in one-stop services. Three approaches have been implemented and compared including Queueing Theory using *M/M/c* pattern, Average Time prediction, and Random Forest algorithm. Based on the experimental results, the method that achieves the highest accuracy is Random Forest algorithm: ear nose and throat clinic dataset which results in the accuracy of 85.76% and Khon Kaen University post office dataset which results in the accuracy of 81.70%. The most important feature of both datasets is the number of waiting queues.

In the future, the researchers are interested to use other datasets of one-stop service and study new features, such as the weather that may affect waiting time prediction.

REFERENCES

[1] J. F. Arroyo, "Primary study of the impact of waiting time management on customer satisfaction in the Thai hotel industry," in *Proc. Burapha Univ. Int. Conf.*, 2015, pp. 172–186.

[2]   R. L. Hensley and J. Sulek, "Customer satisfaction with waits in multi-stage services," *Manag. Serv. Qual. An Int. J.*, vol. 17, no. 2, pp. 152–173, 2007.

[3]   A. C. R. van Riel, J. Semeijn, D. Ribbink, and Y. Bomert-Peters, "Waiting for service at the checkout negative emotional responses, store image and overall satisfaction," *J. Serv. Manag.*, vol. 23, no. 2, pp. 144–169, 2012.

[4]   G. Martin. One stop shop - the meaning and origin of this phrase. [Online].   Available: https://www.phrases.org.uk/meanings/one-stop-shop.html

[5]   C. Rattanapan and P. Peerapattana, "The improvement of the service system for reducing queuing time by using lean thinking and simulation model: The case study of dental," *MBA KKU J.*, vol. 1, pp. 135–150, 2016.

[6]   M. A. Kumar, S. Mahesh, T. Kamalraj, and M. Azhagiri, "Patient treatment time prediction for out patient department OPD," *Int. J. Eng. Sci. Comput.*, vol. 7, no. 4, pp. 6238–6242, 2017.

[7]   F. Rarh, D. Pojee, and S. Zulphekari, "Restaurant Table reservation using time-series prediction," in *Proc. 2nd Int. Conf. Commun. Electron. Syst.*, pp. 153–155, 2017.

[8]   A. Carvalho and O. Belo, "Predicting waiting time in customer queuing systems," in *Proc. IEEE Int. Conf. Knowl. Eng. Appl.*, 2016, pp. 155–159.

[9]   R. N. Mourão, R. S. Carvalho, R. N. Carvalho, and G. N. Ramos, "Predicting waiting time overflow on bank teller queues," in *Proc. 2017 16th IEEE Int. Conf. Mach. Learn. Appl.*, 2017, pp. 842–847.

[10] T. Erdelić, T. Carić, and M. Ravlić, "Predicting waiting times at intersections," in *Proc. 59th Int. Symp. ELMAR-2017*, 2017, pp. 18–20.

[11] E. W. Weisstein. (2018). Arithmetic mean. *Math World*. [Online]. Available: http://mathworld.wolfram.com/ArithmeticMean.html

[12] C. Noyunsan, T. Katanyukul, and K. Saikaew, "Performance evaluation of supervised learning algorithms with various training data sizes and missing attributes," *Eng. Appl. Sci. Res.*, 2018.

[13] L. Breiman, "Randomforest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[14] H. Han, X. Guo, and H. Yu, "Variable selection using mean decrease accuracy and mean decrease gini based on random forest," in *Proc. 7th IEEE Int. Conf. Softw. Eng. Serv. Sci.*, pp. 219–224, 2016.

[15] N. Watcharenwong and K. Saikaew, "Spam detection for closed facebook groups," in *Proc. 2017 14th Int. Jt. Conf. Comput. Sci. Softw. Eng.*, no. 5, 2017.

Yotravee Sanit-in was born in Ubon Ratchathani, Thailand on October 15, 1994. He received the B.S. degree in computer engineering from Khon Kaen University, Thailand, in 2017. He had been a web application developer during his internship and cooperative education at Digio (Thailand) Co., Ltd., in 2016 and 2017, respectively. His current research interests include machine learning, big data, data mining and internet of things (IoT).

Kanda Runapongsa Saikaew received the B.S. degree in electrical and computer engineering from Carnegie Mellon University, Pennsylvania, USA, in 1997, and the M.S. and Ph.D. degrees in computer science and engineering from the University of Michigan at Ann Arbor, in 1999 and 2003, respectively. In 2003, she joined the Department of Computer Engineering, Khon Kaen University, as a lecturer, and became an associate professor in 2015. Her current research interests include machine learning and social network analysis