Development of Business Intelligence System and Prediction with Data Mining of Lam Sam Kaeo Town Municipality, Thailand

Watchawee Wongart[1] and Somchai Lekcharoen[2]

[1,2] College of Digital Innovation Technology, Rangsit University, Pathum thani, Thailand
E-mail: watchawee.w63@rsu.ac.th, somchai@rsu.ac.th

Received: June 13, 2021 / Revised: July 12, 2021 / Accepted: July 27, 2021

Abstract—**The purpose of this research was to develop a business intelligence system and make predictions with data mining of Lam Sam Kaeo Town Municipality. The process of development used Microsoft Visual Studio (SSDT) with SQL Server Integration Services (SSIS) to create a data warehouse, then using SQL Server Analysis Services (SSAS) to create cube and using SQL Server Reporting Services (SSRS) to create reports then publish to web browsers for local administration officers to make decisions.**

The sample data used in this project covered local government taxpayers living in Lam Sam Kaeo Town Municipality in 2020: altogether about 200 taxpayers. The analysis was based on the new land and building tax act 2019. RapidMiner Studio was used to create the analysis model to determine the factors that cause tax delays in Lam Sam Kaeo Town Municipality. A comparison of three classification algorithms showed similar accuracy: J48 Decision Tree has accuracy = 96%, Naïve Bayes has accuracy = 92%, Neural Network has accuracy = 96.50%, ANOVA test found no significant difference at 0.05 level is not different so the researcher chooses the Decision Tree method for this research.

The results are that the most influential factors causing overdue tax payments are 'forget to pay', 'lack of advertising', and 'lack of e-payment method.

Index Terms—**Business Intelligence System, Data Warehouse, Decision Tree, Naïve Bayes, Neural Network**

I. Introduction

The Thai government has recently approved a new tax on land and buildings, as described in the Land and Buildings Tax Act 2019 [1], and the Land and Buildings Tax Reduction Act 2020 [2].

There are three parts to this new law: a land and buildings tax, a signboard tax, and a corporate building tax [1].

Lam Sam Kaeo Town Municipality is looking for a business intelligence system to handle the administration of this tax, and this paper describes the system that was developed for them. Currently, the data about the land size, its ownership, and use are stored in a spreadsheet, which has to be manually sorted and updated with each change to the land title deed holder, change of address of the landowner, newly-surveyed land that has no appraisal value, etc., This system is inefficient and error-prone, resulting in much manual checking and delay [3].

Having studied the problem, the following solution approach was adopted: firstly, use SQL Server technology to store data, then use the SQL Server Integration Services (SSIS) to create, extract, transform and load information into the data warehouse [4]. This will combine the relevant information and automatically make the various tax calculations for the people covered by the new land tax. Next, the results will be extracted as a dimensional model data structure and separated into a dimension table and fact table to import into SQL Server Management Studio to become a data warehouse. Finally, the SQL Server Analysis Services (SSAS) will be used to create a model for analyzing the data of the dimension table and fact table, allowing the user to view details of multidimensional data [5].

In addition, a cube was developed to support forecasting and decision-making. With this, SQL Server Reporting Services (SSRS) can be used to create a dashboard for summarizing information and to automatically generate customizable reports for executives which can be viewed with a Web Browser [6].

The first step in the process was to create a data gathering spreadsheet about village land, including land data, land address, landowner, landowner address, land utilization, date and time tax was paid, and the formula for calculating tax. All the data was sorted by village and was checked by government officers [7].

The second step was to use a dimensional model to structure the data, then to use the SQL Server Integration Services (SSIS) to create, extract, transform, and load information into the data warehouse. Dimension tables were created for land data, land address, landowner, landowner address, land utilization, and the date and time taxes were paid. A fact table was created for the formula to calculate tax [8].

It is important that the extract, transform and load (ETL) process be evaluated by the intended users until they are satisfied with the structure and knowledgeable enough to analyze the data and customize reports and create tables and charts for their management. A typical report might show which village pays the most tax, and how the land in a particular village is segmented by usage: residential, agricultural, commercial, or wasteland [9].

This information is essential to support local administration officers with operational insights and decision-making for overdue tax payments because the main income of local government is a local tax. And with this process, the information can be analyzed much more quickly and efficiently than the manual method [10].

At each stage, the government officers are required to verify that the information is accurate and can be used to make decisions.

Now the Lam Sam Kaeo Town Municipality didn't create a data warehouse. The researcher has developed an organized, accurate, and complete business intelligence system to help collect land and building taxes [5]-[7]. It can be used to check who paid taxes on time, who paid late, and who is overdue with their payment. Analyzing historical data by data mining also allows the user to predict the factors that cause tax delays [15]-[17]. This was done by importing the data into RapidMiner Studio to create three types of forecasts, namely Decision Tree [11], Naïve Bayes [12], and Neural Network [13].

II. Objective

1. To develop a business intelligence system to collect data of taxpayers with Visual Studio (SSDT) to build SQL Server Integration Services (SSIS) and then use SQL Server Analysis Services (SSAS) to create a cube. Additionally, use SQL Server Reporting Services (SSRS) to create reports [4]-[10].
2. To predict the factors of delayed tax payment in Lam Sam Kaeo Town Municipality using data mining techniques [17]-[25].

III. Related Work

Develop business intelligence with SQL Server, which consists of a business intelligence process that can be described as follows:

- The data warehouse must be designed carefully by dividing it into dimension tables and fact tables with primary keys and foreign keys. In this case, the design consists of dimension tables (dates, land, land address, owner address, and utilization) and one fact table (land tax) with relevant constraints defined by primary keys and foreign keys. Then the raw spreadsheet data can be extracted in Table I and II. [4], [5], [7].

TABLE I
Dimension Table (Dim Date)

Date_Key	Full_Date_Land_Tax	Year_Land_Tax	Month_Land_Tax	Day_Land_Tax	Quarter	EnglishMonthName	EnglishDayName	ThaiMonthName	ThaiDayName
1	2020-04-01	2020	4	1	3	April	Monday	มิถุนายน	วันจันทร์
2	2020-04-02	2020	4	2	3	April	Tuesday	มิถุนายน	วันอังคาร
3	2020-04-03	2020	4	3	3	April	Wednesday	มิถุนายน	วันพุธ
4	2020-04-04	2020	4	4	3	April	Thursday	มิถุนายน	วันพฤหัสบดี
5	2020-04-05	2020	4	5	3	April	Friday	มิถุนายน	วันศุกร์
6	2020-04-06	2020	4	6	3	April	Monday	มิถุนายน	วันจันทร์
7	2020-04-07	2020	4	7	3	April	Tuesday	มิถุนายน	วันอังคาร
8	2020-04-08	2020	4	8	3	April	Wednesday	มิถุนายน	วันพุธ
9	2020-04-09	2020	4	9	3	April	Thursday	มิถุนายน	วันพฤหัสบดี
10	2020-04-10	2020	4	10	3	April	Friday	มิถุนายน	วันศุกร์
11	2020-04-11	2020	4	11	3	April	Monday	มิถุนายน	วันจันทร์
12	2020-04-12	2020	4	12	3	April	Tuesday	มิถุนายน	วันอังคาร
13	2020-04-13	2020	4	13	3	April	Wednesday	มิถุนายน	วันพุธ
14	2020-04-14	2020	4	14	3	April	Thursday	มิถุนายน	วันพฤหัสบดี
15	2020-04-15	2020	4	15	3	April	Friday	มิถุนายน	วันศุกร์
16	2020-04-16	2020	4	16	3	April	Monday	มิถุนายน	วันจันทร์

TABLE II
FACT TABLE (FACT_LAND_TAX) WITH LAND_NUM , ORDER_NUM, UTILIZATION_CODE, DATE_KEY AS FOREIGN K.EY

SSN	Parcel_Num	Survey	Utilization_Code	Value_of_land	Utilization_rate	After_Tax_Rate	Discount_90	Payment_10	House_Tax_2019	Tax Relief	Tax Relief_Increase_50	Real_Tax_Pay	Full_Date_Land_Tax
3 3107 00719 08 5	60373	38217	4	1,750,000.00	0.3	5,250.00	4,725.00	525.00	150.00	375	187.5	337.50	2020-08-16
3 1013 00445 03 7	56321	16428	4	3,024,000.00	0.3	9,072.00	8,164.80	907.20	151.00	756.2	378.1	529.10	2020-08-17
3 1005 01723 45 9	20165	6739	4	550,000.00	0.3	1,650.00	1,485.00	165.00	152.00	13	6.5	158.50	2020-08-18
3 1006 00549 85 6	32470	9576	4	1,750,000.00	0.3	5,250.00	4,725.00	525.00	153.00	372	186	339.00	2020-08-19
3 2406 00286 57 9	29550	8291	4	1,400,000.00	0.3	4,200.00	3,780.00	420.00	154.00	266	133	287.00	2020-08-20
3 1001 00787 37 5	11317	3573	4	1,800,000.00	0.3	5,400.00	4,860.00	540.00	155.00	385	192.5	347.50	2020-08-21
3 1001 00787 39 1	11318	3574	4	1,800,000.00	0.3	5,400.00	4,860.00	540.00	156.00	384	192	348.00	2020-08-22
3 1009 05776 05 8	29499	8240	4	1,470,000.00	0.3	4,410.00	3,969.00	441.00	157.00	284	142	299.00	2020-08-23
3 1020 02465 07 3	47093	14612	4	875,000.00	0.3	2,625.00	2,362.50	262.50	158.00	104.5	52.25	210.25	2020-08-24
3 1010 00641 47 7	11314	3570	4	1,800,000.00	0.3	5,400.00	4,860.00	540.00	159.00	381	190.5	349.50	2020-08-25
3 1704 00100 57 0	47288	14902	4	550,000.00	0.3	1,650.00	1,485.00	165.00	160.00	5	2.5	162.50	2020-08-26
3 1005 01307 08 1	29672	8413	4	1,400,000.00	0.3	4,200.00	3,780.00	420.00	161.00	259	129.5	290.50	2020-08-27
3 1202 00070 87 1	29553	8294	4	700,000.00	0.3	2,100.00	1,890.00	210.00	162.00	48	24	186.00	2020-08-28
3 1009 03818 42 0	9018	2515	4	992,000.00	0.3	2,976.00	2,678.40	297.60	163.00	134.6	67.3	230.30	2020-08-29
3 1005 03934 06 6	32469	9575	4	1,312,500.00	0.3	3,937.50	3,543.75	393.75	164.00	229.75	114.875	278.88	2020-08-30
3 1017 00739 63 1	19908	6493	4	1,512,000.00	0.3	4,536.00	4,082.40	453.60	165.00	288.6	144.3	309.30	2020-08-31
3 1306 00004 80 9	21771	7420	4	1,000,000.00	0.3	3,000.00	2,700.00	300.00	166.00	134	67	233.00	2020-09-01
3 1002 00194 57 1	20008	6582	4	750,000.00	0.3	2,250.00	2,025.00	225.00	167.00	58	29	196.00	2020-09-02
3 1019 00120 44 9	2214	497	4	23,192,000.00	0.3	69,576.00	62,618.40	6,957.60	168.00	6789.6	3394.8	3,562.80	2020-09-03
3 7699 00228 02 1	19982	6556	4	1,050,000.00	0.3	3,150.00	2,835.00	315.00	169.00	146	73	242.00	2020-09-04
3 5208 00008 68 5	20031	6605	4	750,000.00	0.3	2,250.00	2,025.00	225.00	170.00	55	27.5	197.50	2020-09-05

- The next process is one of extracting, transforming, and loading (ETL) the spreadsheet data into the data warehouse using Microsoft Visual Studio (SSDT) and SQL Server Integration Services (SSIS). The dimension table will be imported first. The primary key and data type will be set appropriately before importing. The fact table will be imported last. To assure data integrity, we create foreign keys to associate with the primary key in each dimension table, then create an Entity Relationship Diagram (ER diagram) to check the database relationship. Each dimension table must have a primary key to point to the fact table, and the data that has been through this process is considered correct. It is stored in a data warehouse in Microsoft SQL Server Management Studio in Fig. 1 [4], [6], [7].
- Using Microsoft Visual Studio (SSDT) to create SQL Server Analysis Services (SSAS), then create data source to retrieve data that we need from the data warehouse in Microsoft SQL Server Management Studio by selecting the database name and selecting all the required tables. This must include a dimension table and a fact table. Then create a data source view and select all the tables that we want to display in the ER diagram, then create a cube by assigning a table of quantitative numerical values that can be calculated as a fact table and the rest will be a dimension table by selecting the fields that need to be analyzed. When finished, test the deployed data to see if it passes. If not passed, the error must be found and corrected. If passed, it is considered to be set up correctly in Fig. 2 [5], [6].
- Display, after the deployment is successful, go to tab browser to select the output as a spreadsheet, which is a symbol of the program Microsoft Excel. The display can select data in the columns of the dimension table and fact table as needed in order to be able to create a dashboard to summarize the selected information. Then the size, color, font, various graph styles, tables, and dashboard designs can be customized for the executive report Fig. 3 [6].
- Report display, use Microsoft Visual Studio (SSDT) to create SQL Server Report Services (SSRS) as a more complex report format. You can import data from the data warehouse to select only the attributes that you want to display. When deployed, the data will be imported into a web browser, which can be browsed using a web browser Fig. 4 [6]-[10].

IV. DEVELOPMENT OF BUSINESS INTELLIGENCE SYSTEM TECHNIQUE

A. Data Warehouse Design (Snow Flake Schema Dimension Model)

B. SQL Server Integration Services (SSIS)

Data imported into SQL Server Management Studio becomes a data warehouse with a snowflake schema dimension model. There is a processing list from spreadsheet to database in Fig. 1.

- Dim_Date import to be OLE DB Dim_date
- Dim_Land import to be OLE DB Dim_Land
- Dim_Land_Address import to be OLE DB Dim_Land_Address

- Dim_Owner import to be OLE DB Dim_Owner
- Dim_Owner_Address import to be OLE DB Dim_Owner_Address
- Dim_Utilizations import to be OLE DB Dim_Utilizations
- Fact_Land_Tax import to be OLE DB Fact_Land_Tax

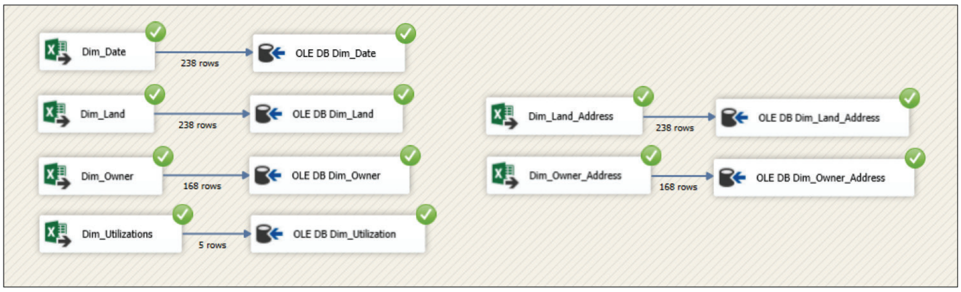


Fig. 1. Use SQL Server Integration Services (SSIS) to extract, transform, and load (ETL) process by using Visual Studio (SSDT) create SSIS using spreadsheet import into SQL Server Management Studio.

C. SQL Server Analysis Services (SSAS)

SSAS will bring Dimensional Table and Fact Table after verifying ETL process from SSIS into a data warehouse. Data were analyzed by using Cube. Tables were analyzed as follows: Dim_Date, Dim_Land, Dim_Land_Address, Dim_Owner, Dim_Owner_Address, Dim_Utilization, Fact_Land_Tax in Fig. 2.

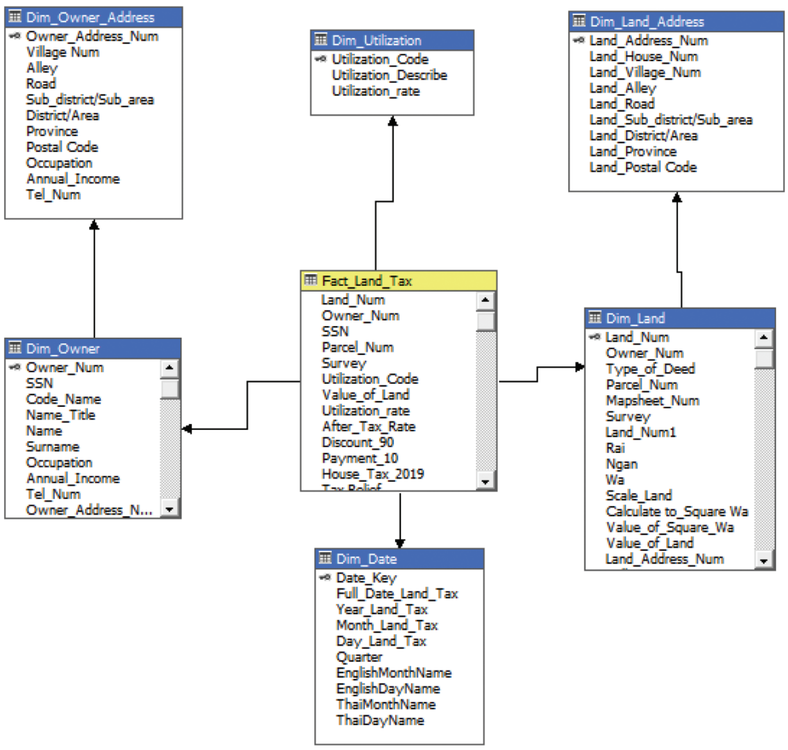


Fig. 2. Relation model Snow flake schema.

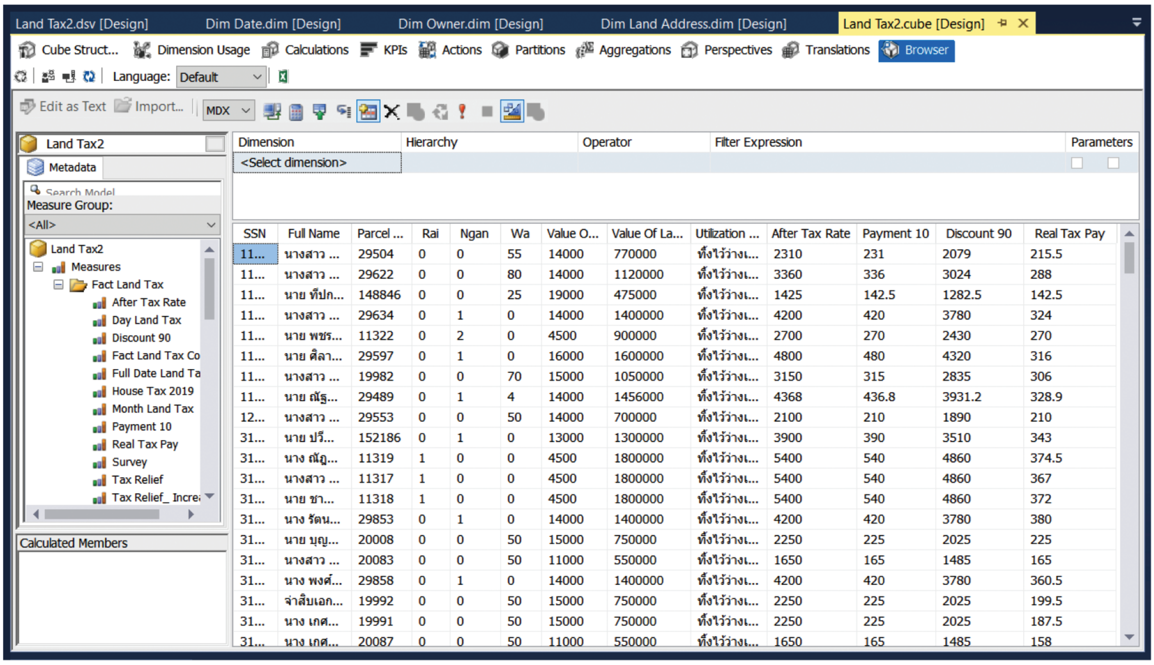
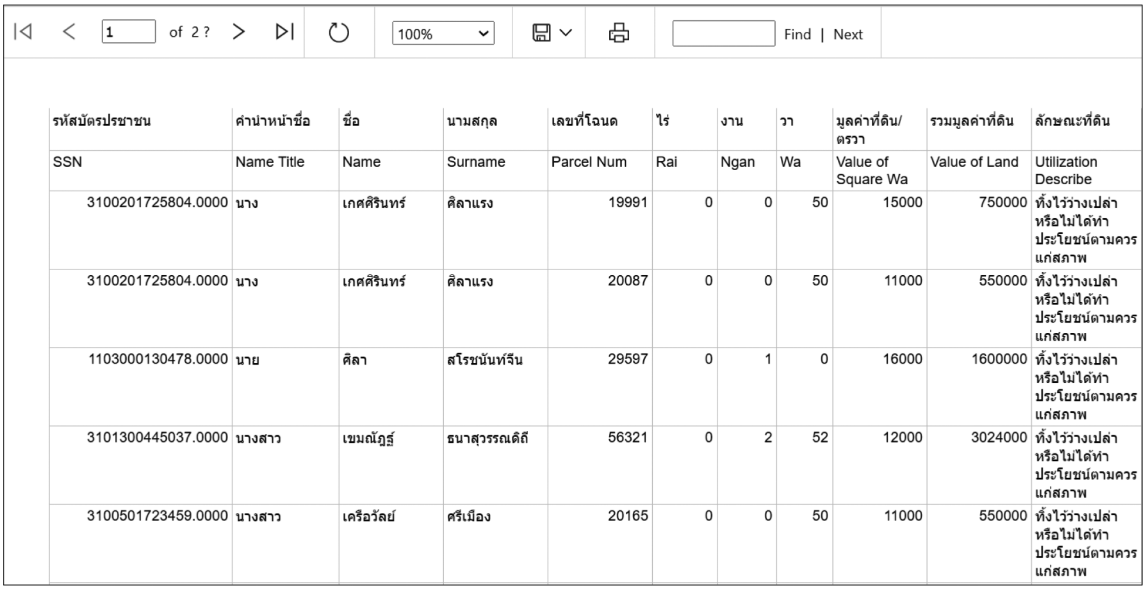


Fig. 3. Cube browser has the following data attributes: SSN, Full_Name, Parcel_Num, Rai, Ngan, Wa, Value of Square Wa, Utilization describe, House tax 2019, Discount 90, Payment 90, Real_Tax_Pay.

D. SQL Server Report Services (SSRS)

SSRS report issuing report is to bring data from the data warehouse to produce a more complex table report by selecting the attribute to be exported and able to deploy to display on a web browser.



รหัสบัตรประชาชน	คำนำหน้าชื่อ	ชื่อ	นามสกุล	เลขที่โฉนด	ไร่	งาน	วา	มูลค่าที่ดิน/ตรวา	รวมมูลค่าที่ดิน	ลักษณะที่ดิน
SSN	Name Title	Name	Surname	Parcel Num	Rai	Ngan	Wa	Value of Square Wa	Value of Land	Utilization Describe
3100201725804.0000	นาง	เกศศิรินทร์	ศิลาแรง	19991	0	0	50	15000	750000	ทิ้งไว้ว่างเปล่าหรือไม่ได้ทำประโยชน์ตามควรแก่สภาพ
3100201725804.0000	นาง	เกศศิรินทร์	ศิลาแรง	20087	0	0	50	11000	550000	ทิ้งไว้ว่างเปล่าหรือไม่ได้ทำประโยชน์ตามควรแก่สภาพ
1103000130478.0000	นาย	ศิลา	สโรชนันท์จีน	29597	0	1	0	16000	1600000	ทิ้งไว้ว่างเปล่าหรือไม่ได้ทำประโยชน์ตามควรแก่สภาพ
3101300445037.0000	นางสาว	เขมณัฏฐ์	ธนาสุวรรณดิถี	56321	0	2	52	12000	3024000	ทิ้งไว้ว่างเปล่าหรือไม่ได้ทำประโยชน์ตามควรแก่สภาพ
3100501723459.0000	นางสาว	เครือวัลย์	ศรีเมือง	20165	0	0	50	11000	550000	ทิ้งไว้ว่างเปล่าหรือไม่ได้ทำประโยชน์ตามควรแก่สภาพ

Fig. 4. Create a report on web browser.

V. Prediction Data with Data Mining Technique

Analyze the factors of delayed tax payment with RapidMiner Studio consisting of the following techniques:

1) Decision tree (J48*)* is a tool to help make strategic or operational decisions in a business, government agency, or other organization by using a decision tree. It resembles an upside-down tree. Where the very first note at the top will be the root of the tree called the Root Node, where each node will display attributes as a Branch, each branch will show the results in the test, and the Leaf Node will show the results of classification data in Fig. 5 [11].

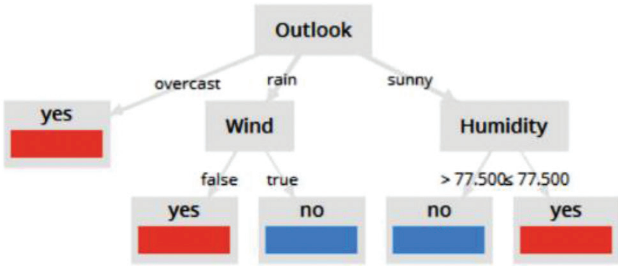


Fig. 5. Decision Tree

2) Naïve Bayes classification is a technique that uses probability theory according to Bayes' Theorem as an algorithm for classifying data. It works by learning the problems that arise to create new classification conditions. There is a principle for calculating the probability of predicting results. It is a technique for solving a classification problem that can predict the outcome. By analyzing the relationship between two or more variables to create a probabilistic condition for each relationship. Suitable for the case of a large number of samples and each attribute is independent of the other. The probability of the data in Eq. (1) [12].

$$P(h/D) = \frac{P(D|h)P(h)}{P(D)} \quad (1)$$

Naive Bay's equation

Where P(h/D) is the probability distribution of the hypothesis, h using the data D. According to the theory,

- P(h) is the previous probability of the hypothesis, h.
- P(D) is the previous probability of the sample data set, D.
- P(h|D) is the probability of h when known D.
- P(D|h) is the probability of D when known h.

3) A Neural network is a structure for processing information that is similar to the neurons in the human brain. It consists of a basic processor called neural processing and occurs in a sub-processor called a node, which simulates the behavior of a cell-signaling between nodes connected in three layers: the input layer, the hidden layer, and the output layer in Fig. 6 [13], [14].

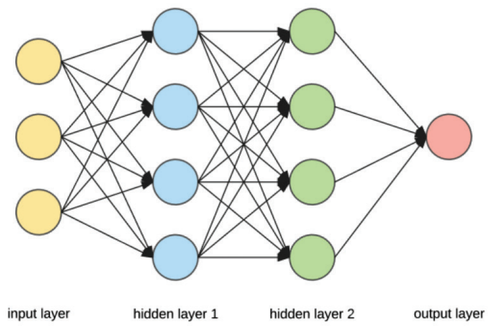


Fig. 6. Neural network

- The input layer is numeric data. If it is qualitative data, the data must be converted to a quantitative format acceptable to the neural network.
- The hidden layer is an intermediate step. They function like deep learning. Each layer can have as many neural nodes as possible, and the increase affects the performance.
- The output layer model is the actual result of the learning process of the neural network [13].

TABLE III
ORGANIZE THE DATA IN A TABULAR FORM FOR EASY CALCULATIONS.

Data No.	Process					
	1	2	3	...	k	
1	X_{11}	X_{12}	X_{13}	…	X_{1k}	
2	X_{21}	X_{22}	X_{23}	…	X_{2k}	
3	X_{31}	X_{32}	X_{33}	…	X_{3k}	
	.	.	.	.	.	
	.	.	.	.	.	
N	X_{n1}	X_{n2}	X_{n3}	…	X_{nk}	
Total	T_1	T_2	T_3	…	T_j	$T_.$
Mean	$\bar{X}_1$	$\bar{X}_1$	$\bar{X}_1$	…	$\bar{X}_1$	$\bar{X}_.$

Symbols used in the calculations

X_{ij} is the observation I in the procedure j when i = 1, 2, 3, …, n
j = 1, 2, 3, …, k

T_j is the sum of the observations in the procedure J.

$T_.$ is the sum of all observations.

$T_.$ is the number of data in the group J.

N is the total number of data.

4) The variance between the groups (Sum of Square Between the Groups (SSB)) is the variance between the mean of the data of each group. It is calculated in Eq. (2) [13], [14].

$$SSB = \sum_{i=1}^{k} (x_i - \overline{x})^2 \quad (2)$$

(4.1) Sum of Square Within the Groups (SSE) is the variance between data within the same group. It is calculated in Eq. (3) [12].

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2 \quad (3)$$

(4.2) The total variance (Sum of Square Total: SST) is the total variance of all data from all populations in Eq. (4) [13], [14].

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2 \ , SST = SSB + SSE \quad (4)$$

In this research, the researcher used Analysis of Variance (ANOVA) to compare the performance of the three forecasting models because ANOVA was able to compare two or more forecasting models. This makes it appropriate to compare the performance of the forecast model. The data to be analyzed for a variance must comply with the following prerequisites [13], [14].

- The samples were randomly drawn from the normal distribution.
- The Population must independent of each other .
- The number of population variances must have the same amount.

(4.3) Accuracy is a measure of accuracy to compare forecast models to find a suitable model. Determination of work efficiency in Eq. (5) [13].

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} x\ 100\% \quad (5)$$

TP is positive correct prediction value.
TN is negative correct prediction value.
FN is a positive error prediction value.
FP is negative faulty value.

(4.4) Precision is a measure of whether the forecast is true. The accuracy of the forecasting model can find the precision in Eq. (6) [13].

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

TP is a positive prediction value.
FP is a negative prediction error value.

(4.5) Recall is a measure of the forecast that is true. What is the ratio of all true values of the forecast model? Remembrance can be found in Eq. (7) [13].

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

TP is a positive prediction value.
FP is a positive prediction error value.

(4.6) Tool

In this research, the researcher used spreadsheet data to collect a survey for forecasting factors causing a delay in local tax payment. By extracting data from spreadsheet and using RapidMiner Studio as a tool to experiment with forecasting models each model could be compared [17]-[20].

VI. Results

The results of the research can be summarized into three parts:

Part 1 Comparison of Decision Tree, Naïve Bayes, and Neural Network Analysis.

Selecting three forecast models to find Accuracy, Precision and Recall in each forecast model, which will consist of the following operations. Import Data is to retrieve data in spreadsheet format, prepared for use in forecasting with various forecasting models. Select attributes is the selection of attributes to be used as a class in forecasting, and Cross Validation is to test the performance of each forecast model which uses RapidMiner Studio as a forecasting tool.

The process of creating a forecast model using a decision tree (J48) consists of operators as follows: first, forecasting with a Decision Tree forecast model, then the application of the created forecast model to make predictions. Finally, measure the performance of the forecasting model, for attributes such as Accuracy, Class Recall Accuracy, Class Precision accuracy, etc.

The forecasting process with the Decision Tree Forecasting Model is shown in the figure in Fig. 7.

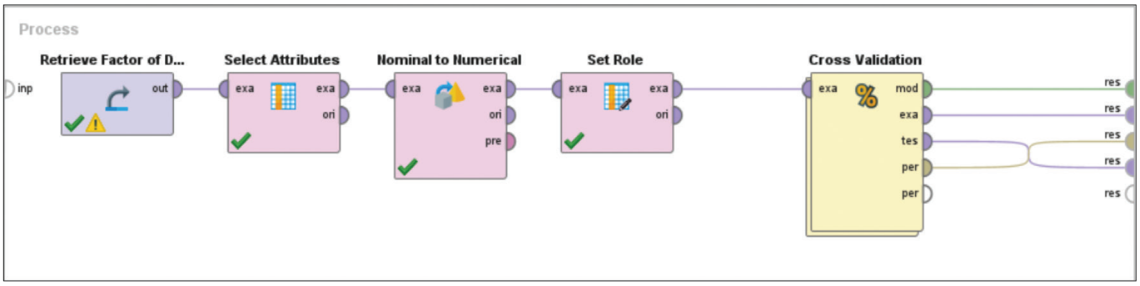


Fig. 7. The core process of forecasting with various forecasting models.

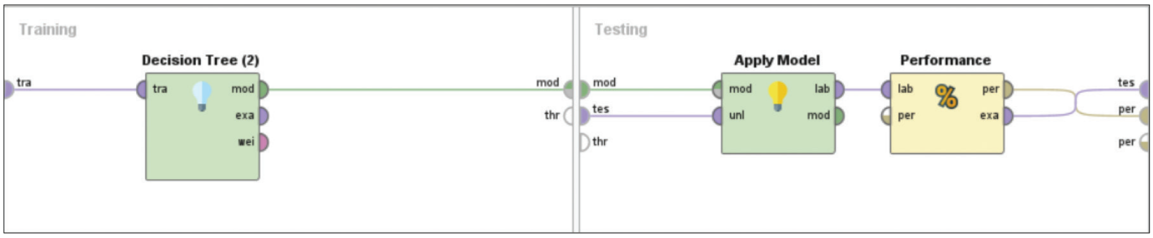


Fig. 8. Forecasting process with decision tree forecasting model

After forecasting with the Decision Tree Forecasting Model, the results are as shown in Table IV.

TABLE IV
FORECAST RESULTS WITH DECISION TREE FORECASTING MODEL

Model	Accuracy	Class Recall (True Yes)	Class Recall (True No)	Class Precision (Pred. Yes)	Class Precision (Pred. No)
Decision Tree	96.00%	96.00%	96.00%	97.56%	93.51%

Decision Tree Forecasting Model Accuracy = 96.00% Accuracy of Tax Factor Forecasting True Variable Yes = 96.00% Accuracy of Tax Factor Forecast False Variable No = 96.00%

Forecasting modeling process using Naïve Bayes consists of the following operators: first is forecasting with the Naïve Bayes forecasting model, then the application of forecasting models. Finally, measure of the performance of the forecast model, such as the Accuracy, Class Recall accuracy, and the accuracy of the Class Precision class, etc.

Forecasting process with the Naïve Bayes forecasting model as shown in the figure in Fig. 9.

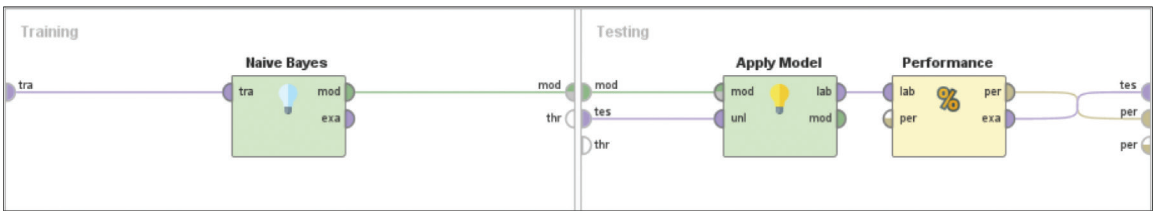


Fig. 9. Forecasting process with the Naïve Bayes forecasting model

After forecasting with the Naïve Bayes forecasting model, the results were as shown in Table V

TABLE V
FORECAST RESULTS WITH THE NAÏVE BAYES FORECASTING MODEL

Model	Accuracy	Class Recall (True Yes)	Class Recall (True No)	Class Precision (Pred. Yes)	Class Precision (Pred. No)
Naïve bayes	92.00%	100.00%	78.67%	88.65%	100.00%

Naïve Bayes Forecast Model Accuracy = 92.00 % Tax Factor Forecast Accuracy True Variable Yes = 100.00 % Tax Factor Forecast Accuracy False Variable No = 78.67%

The process of building a forecast model using a neural network consists of the following operators: forecast with a neural network forecast model, apply the forecasting model that has been created, then use measure the performance of the forecast model, in terms of Accuracy, Class Recall class accuracy, the accuracy of the Class Precision class, etc.

Forecasting Process with Neural Network Forecasting Model in Fig. 10.

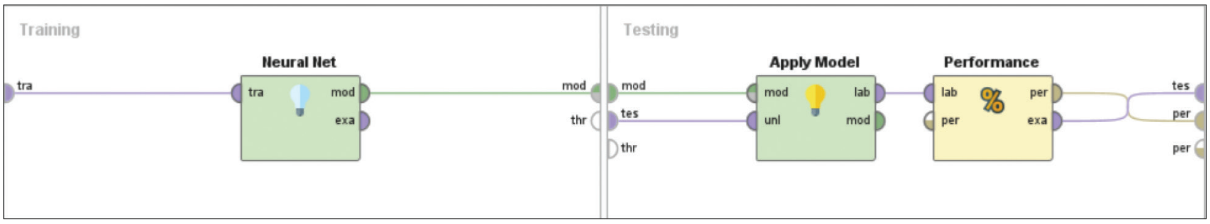


Fig. 10. The forecasting process with the Neural Network forecasting model.

After forecasting with the Neural Network forecasting model, the results are shown in Table VI.

TABLE VI
FORECAST RESULTS WITH NEURAL NETWORK FORECAST MODEL

Model	Accuracy	Class Recall (True Yes)	Class Recall (True No)	Class Precision (Pred. Yes)	Class Precision (Pred. No)
Neural Network	96.50%	96.00%	97.33%	98.36%	93.59%

Neural Network Forecasting Model Accuracy = 96.50% Accuracy of Tax Factor Forecast True Variable Yes = 96.00% Accuracy of Tax Factor Forecast False Variable No = 97.33%

Part 2 Compares the performance of Decision Tree, Naïve Bayes, and Neural Network forecasting models using ANOVA statistics in Fig. 11.

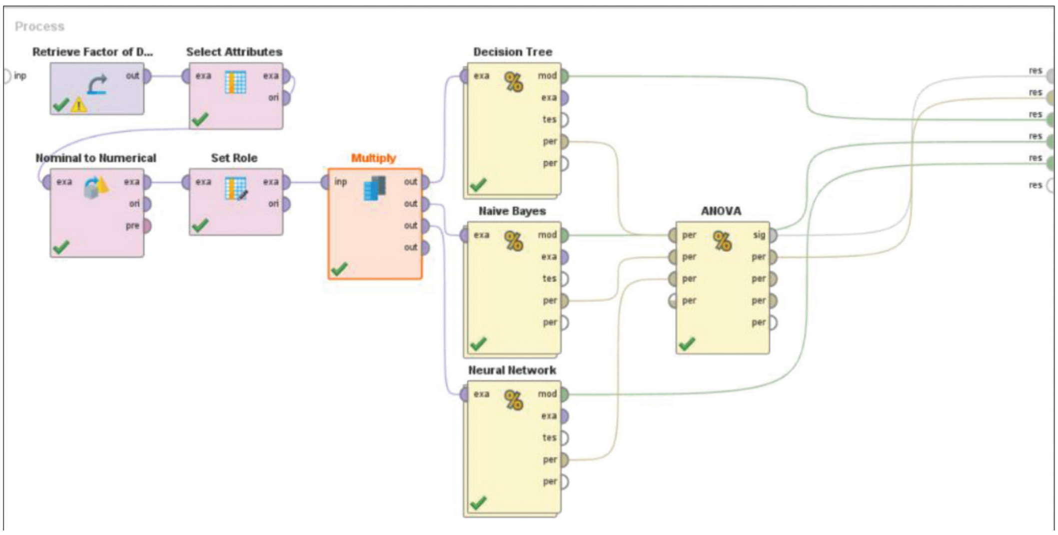


Fig. 11. Comparison of forecasting models with Rapid Miner Studio

Comparison of all 3 forecast models by using ANOVA statistics. Comparison of forecast models results as shown in Table VII.

TABLE VII
ANOVA FORECAST MODEL COMPARISON RESULTS

Source	Square Sums	DF	Mean Square	F	Prob
Between	0.007	2	0.004	1.817	0.182
Residuals	0.053	27	0.002		
Total	0.060	29			

The comparison of forecast models with ANOVA statistics revealed that the Decision Tree, Naïve Bayes and Neural Network forecast models had F = 1.817 and Prob = 0.182, which were not significantly different at the 0.05 level. The three forecasting methods were not significantly different. Therefore, any type of forecasting model can be used in research.

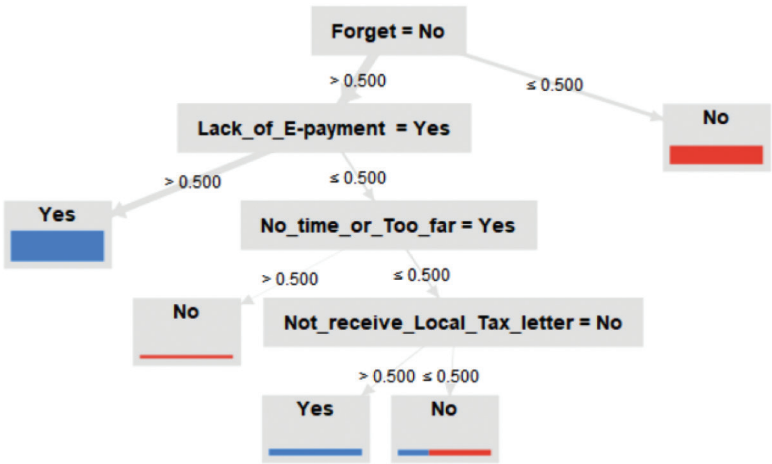


Fig. 12. Forecast results factors of delay payment with Decision Tree

From the forecasting of delayed local tax factors to Lam Sam Kaeo Town Municipality in 2020 with the Decision Tree forecasting model in Fig. 12.

1) IF Forget = "≤ 0.500" = "No"
#No means: Forget to pay tax".

2) IF Forget = "> 0.500" AND
Lack_of_E-payment = "> 0.500" = "Yes"
#Yes means: lack of E-payment method.

3) IF Forget = "> 0.500" AND
Lack_of_E-payment = "≤ 0.500" AND No_Time_or_Too_far = "> 0.500" = "No"
#No means: not worth the time and cost to travel to the local government organization.

4) IF Forget = "> 0.500" AND
Lack_of_E-payment = "≤ 0.500" AND
No_Time_or_Too_far = "≤ 0.500" AND
Not_receive_Local_Tax_Letter = "≤ 0.500" = "No"
#No means: not receive tax letter to inform tax from local government organization.

5) IF Forget = "> 0.500" AND
Lack_of_E-payment = "≤ 0.500" AND
No_Time_or_Too_far = "≤ 0.500" AND
Not_receive_Local_Tax_Letter = "≤ 0.500" = "Yes"
#Yes means: receive tax letter to inform tax from local government organization.

VII. Conclusion

The business intelligence system is a convenient way to collect data, calculate taxes, check tax payments, and make decisions about Land and Building taxes for Local Administration officers. The prototype developed in this project was given to five Local Administration officers to test. They found the program satisfied their needs and they could easily customize the reports to make them convenient.

A technique was used to predict the factors resulting in delayed tax payment: comparing the three forecasting models of Decision Tree, Naïve Bayes, and Neural Network, found no significant difference at 0.05 level (the ANOVA test found that F = 1.817 and Prob = 0.182.) The researcher chose the Decision Tree forecasting model to use in forecasting factors of late local tax payments in 2020.

The ANOVA test identifies whether there is a significant difference between groups. It allows you to decide whether you can reject the null hypothesis that there is no difference between the groups, and accept the alternate hypothesis that the groups are different.

It was found that the lack of electronic tax payment options (Lack of E-Payment) has the greatest impact, followed by forgetting to pay taxes (Forget) in second place. A solution approach is to add the payment status along with the payment due date and compare them. If the payment status is “NO” after the due date, that means the villager forgot. The delinquent villager can then be sent a reminder letter containing a QR code payment instruction to Lam Sam Kaeo Town Municipality bank account and ID Line to make payment easy and convenient.

Business intelligence (BI) can help companies make better decisions by showing present and historical data within their taxpayer context. Analysts can leverage BI to provide performance and competitor benchmarks (another local government organization) to make the organization run smoother and more efficiently. Analysts can also more easily identify the problems to increase tax. With this knowledge, the local administrators will be able to solve the problem in the following year.

VIII. Discussion

The research found the main factors that delayed local taxation to Lam Sam Kaeo Town Municipality in 2020. Three different forecasting models were compared using a sample of 200 local taxpayers in Lam Sam Kaeo Town municipality. The factors for delaying tax payment to Lam Sam Kaeo Town Municipality are discussed as follows:

From the forecast of factors causing the delay in local taxation to Lam Sam Kaeo Town Municipality it was found that

- Decision Tree analysis found that if the sample group forgot to pay the tax it would result in the most delay in tax payment.
- If the sample group has a problem with tax payment via electronic transaction system, i.e., Prompt Pay, it will result in the second most late tax payment.
- If the sample group has to pay the tax in person, and considers it not worth the time and cost to travel to the local government organization because the distance is far, this would result in delayed tax payment as well.
- If the sample does not receive a letter to inform them of the land and building tax, this will have the effect of delaying the payment of taxes.
- If the sample received a letter to inform them of the land and building tax, then there will be no effect on the late tax payment.

From the analysis, the factors affecting the late payment of local taxes to Lam Sam Kaeo Town Municipality were found. The two major reasons resulting in late payment of local taxes to the municipality are a) that people forgot and b) people found it difficult or inconvenient to make the payments.

Furthermore, according to the research comparing the efficiency and accuracy of the three forecast models, the Neural Network analysis model has an accuracy of 96.50%; the Decision Tree at 96.00%; and the Naïve Bayes model with an accuracy of 92.00% were not significantly different at level 0.05 therefore, any forecasting model can be used in forecasting. In this research, the researcher chose the Decision Tree forecasting Model to determine the factors affecting the late payment of local taxes to Lam Sam Kaeo Town Municipality because the decision tree model is a top-down tree-like structure that explains the decision-making rules for prediction.

At each step, a decision is made based on the attribute in question, and the result generates a branch of the tree. Using this divide and conquer approach, classification rules are generated. The rule is a path from the root node to an end node. An advantage of the Decision Tree is that the information gained can be used in feature selection as well [14]. That is why it is a widely popular and easy-to-understand algorithm.

XI. Suggestions

1) In the future this research should be a group analysis by village numbers. Adding a method for collecting data on surveys of the remaining uncompleted land areas to gain insights into the decisions of local administration officers.

2) Since it is the first year that the Land and Buildings Tax Act 2019 has been applied, there is still a lack of payment systems through the transaction

system. Electronic is Prompt Pay and the address of the taxpayer changes as well. As a result, the tax is delayed. Therefore, Lam Sam Kaeo Town Municipality can apply the research results to solve the problem in the coming years.

3) The results of the analysis can be used as a guideline for solving the problem of late tax payments to Lam Sam Kaeo Town municipality. For example, the development of a business intelligence system to collect complete and current data about land ownership and taxation, and generate reports about who pays taxes on time, late or forgets to pay taxes. This will save a lot of administration time, as well as reduce the need to create and store a lot of paper forms. This will result in more tax collection covering the entire municipality and reduce delays in tax payments.

References

[1] Department of Local Administration, Ministry of Interior. (2021, June 10). *Land and Buildings Tax Act 2019*. [Online]. Available: http://www.dla.go.th/work/mixlaw2.pdf

[2] Department of Local Administration, Ministry of Interior. (2021, Jan. 10). *Royal Decree on reduces land and building taxes by 90%.* (2nd ed.) [Online]. Available: http://www.dla.go.th/upload/document/type2/2021/2/24981_1_1613034073130.pdf

[3] Department of Finance, Lam Sam Kaeo. (2021, May 10). *Data of Land and Building Taxpayers*. [Online]. Available lamsamkaeo.go.th/public/list/datal/index/Manu/1149

[4] Craig Stedman, Ed Burns. (2020, Sep.). *Business Intelligence (BI)*. [Online]. Available: https://searchbusiness.analyticstechtarget.com/definition/business-intelligence-BI

[5] O. Medovoi. (2019, Apr. 20). *Complete Guide to Business Intelligence and Analytics: Strategy, Steps, Processes, and Tools.* [Online]. Available: https://www.altexsoft.com/blog/business/complete-guide-to-business-intelligence-and-analytics-strategy-steps-processes-and-tools/

[6] P. Tiangsombun, (2018, Sep.). Business Intelligence Systems To Support Executive Forecasting And Decisions Making Case Study: Healthcare. *Veridian E-Journal, Science and Technology Silpakorn University.* [Online]. *5(4),* pp. 16-30. Available: https://ph01.tci-thaijo.org/index.php/VESTSU/article/view/148906

[7] S. Leelataweewud. (2020, May.) The Studying of Cloud Business Intelligence System Architecture for Asean Economic Community. *Suthiparitthat Journal DPU.* [Online]. *30(93)*, pp. 177-192. Available: https://so05.tci-thaijo.org/index.php/DPUSuthiparithatJournal/article/view/242764

[8] E. Sangiamkul and T. Panrungsri (2019, Jun.). Development of Decision Support System in Impact and Severity from Flood in Phuket. *Information Technology Journal.* [Online]. *15(1)*, pp. 60-70. Available: https://ph01.tci-thaijo.org/index.php/IT_Journal/article/view/199901

[9] P. Pipatjessadakul and O. Pinngern (2019, Dec.). Development of Business Intelligence System to Support Electrical Distribution. *Journal of project in Computer Science and Information Technology.* [Online]. *5(2)*, pp. 49-56. Available: https://ph02.tci-thaijo.org/index.php/project-journal/article/view/204706

[10] E. Naowanich and N. Jeerungsuwan (2017 Jun.). The Development of Business Intelligence System for Supporting the Administrations' Decision Making to Access Occupation Internationally for Students of Rajamangala University of Technology. *Academic Journal Bangkokthonburi University.* [Online]. *6(1)*, pp. 184-198. Available: https://so01.tci-thaijo.org/index.php/bkkthon/article/view/210782

[11] S. Hai-Jew. (2021, Jun. 10). *Using Decision Trees to Analyze Online Learning Data*. [Online]. Available: https://newprairiepress.org/cgi/viewcontent.cgi?article=1005&context=isitl

[12] S. Ray. (2021, Jun. 10). *6 Easy Steps to Learn Naive Bayes Algorithm with Codes in Python and R. India.* [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

[13] A. Dertat. (2021, Jun. 10). *Applied Deep Learning - Part 1: Artificial Neural Networks*. [Online]. Available: https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6

[14] H. Deng, G. Runger, and E. Tuv. "Artificial Neural Networks and Machine Learning," Presented at ICANN 2011. Conf. Artificial Neural Networks, Jun. 14-17, 2011.

[15] B. Hembree. (2006, Feb.). *One way ANOVA (Analysis of Variance)*. [Online]. Available: https://www.itl.nist.gov/div898/handbook/ppc/section2/ppc231.htm

[16] P. Siriattakul. (2012, Sep.). One-Way ANOVA: Social Science Research. *Journal of Interdisciplinary Research: Graduate Studies,* vol. 1, no. 1, pp. 1-11, Sep. 2012.

[17] N. Hongboonmee and P. Trepanichkul (2019, Jun.). Comparison of Data Classification Efficiency to Analyze Risk Factors that Affect the Occurrence of Hyperthyroidusing Data Mining Techniques. *Journal of Information Science and Technology (JIST).* [Online] *9(1)*, pp. 41-51. Available: https://ph02.tcithaijo.org/index.php/JIST/articleview/179188

[18] R. Caruana and A. Niculescu-Mizil. (2006, Jun.). *An Empirical Comparison of Supervised Learning Algorithms. Present at Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA.* [Online]. Available: https://www.researchgate.net/publication/215990477_An_Empirical_Comparison_of_Supervised_Learning_Algorithms

[19] A. Chutipascharoen (2018, Nov.). A Comparison of the Efficiency of Algorithms and Feature Selection Methods for Predicting the Success of Personal Overseas Money Transfer. *KKU Research Journal of Humanities and Social Sciences*. [Online]. *6(3),* pp. 105-113. Available: https://so04.tci-thaijo.org/index.php/gskkuhs/article/view/156370

[20] N. Hongboonmee and P. Trepanichkul. (2019, Jun.). Comparison of Data Classification Efficiency to Analyze Risk Factors that Affect the Occurrence of Hyperthyroid using Data Mining Techniques. *Journal of Information Science and Technology*. [Online]. *9(1)*. pp. 41-51. Available: https://ph02.tci-thaijo.org/

[21] R. Caruana and A. Niculescu-Mizil. (2006, Jun.). *An Empirical Comparison of Supervised Learning Algorithms. Present at Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA*. [Online]. Available: https://www.researchgate.net/publication/215990477_An_Empirical_Comparison_of_Supervised_Learning_Algorithms

Watchawee Wongart is a student at Master of Science (Information Technology), College of Digital Innovation Technology. He graduates B. in Business Computer Information System form Rangsit University in 2013 and graduates M. in Business Logistic Management in 2016.

Somchai Lekcharoen was born in Bangkok, Thailand in 1965. He received the M.Sc. degree in Information Technology from King Mongkut's University of Technology in 1998, Thonburi Campus, Thailand. He received his Ph.D. in Computer Science at Assumption University of Thailand in 2008. He is an assistant professor of faculty of digital innovation technology, Rangsit University, Thailand. He has published more than 90 scientific papers. He is scientific research and for leadership in the area of science in Telecommunication engineering and his research interest is today focused on congestion control, network of queue and data mining.