

Project-specific Characteristics Reduction to Reduce Processing Time for Sub-classification Analysis of Smart Community Projects Using Data Mining Techniques

Tipaporn Supamid^{1,*}, Surasak Mungsing²

^{1,2}Information Technology, School of Information Technology, Sripatum University

Received: 17 September 2021

Revised: 14 December 2022

Accepted: 15 December 2022

Abstract

Smart City projects have a large number of project-specific characteristics to be used in the analysis for classifying smart-project types. The processing time depends on the quantity of project-specific characteristics used in the analysis. This research investigated the correlation of project-specific characteristics, using data mining techniques, and selected essential project-specific characteristics for the analysis in order to reduce the processing time. The results showed that the quantity of project-specific characteristics was reduced by 33.65%, the processing time was reduced by 28.53%, while accuracy decreased by 1.17%.

Keywords: Data Reduction, Project Characteristics, Smart City, Smart Community, Data Mining

* Corresponding Author; E-mail: tuk_ta12@hotmail.com

Introduction

Smart City projects have a large number of project-specific characteristics to be used in the analysis for classifying smart-project types. The type of smart project can be classified in to seven categories: Smart Mobility, Smart Economy, Smart Environment, Smart Governance, Smart Energy, Smart living, and Smart People by Supamid and Mungsing, (2020). In each project there are many project-specifics characteristics to be used for smart-category classification. The project-specific characteristics are correlated among smart project categories. Therefore, a project can be classified as more than one category (hybrid). The processing time for smart project classification depends on the quantity of project-specific characteristics used in the analysis. The analysis process may take hours depending on the amount of the project-specific characteristics used. The challenge to select the number of the project-specific characteristics to reduce the computation time with no significant decreasing in accuracy. Investigating a correlation of project-specific characteristics among smart categories can speed up the classification process. Correlation analysis is a statistical method used to evaluate the strength of relationship between two quantitative variables. A high correlation means that two or more variables have a strong relationship with each other, while a weak correlation means that the variables are hardly related. Investigation the correlation of project-specific characteristics helps reducing number of project-specific characteristics by selecting common and strong relationship between smart project categories to be used for smart-project category analysis. With the smaller number of project-specific characteristics the processing time would be faster.

Research Objective

This research aims to study the following information.

1. The study a conceptual framework by smart city and smart community.
2. The project-specific characteristics reduction.
3. The researcher created a tool to test the project.
4. The comparing classification methods for data mining is K-Nearest Neighbors (KNN), Neural Networks, and Support Vector Machine (SVM).

Research Framework

The conceptual framework for project-specific characteristics reduction in order to reduce processing time for sub-classification analysis of smart community projects figure 1.

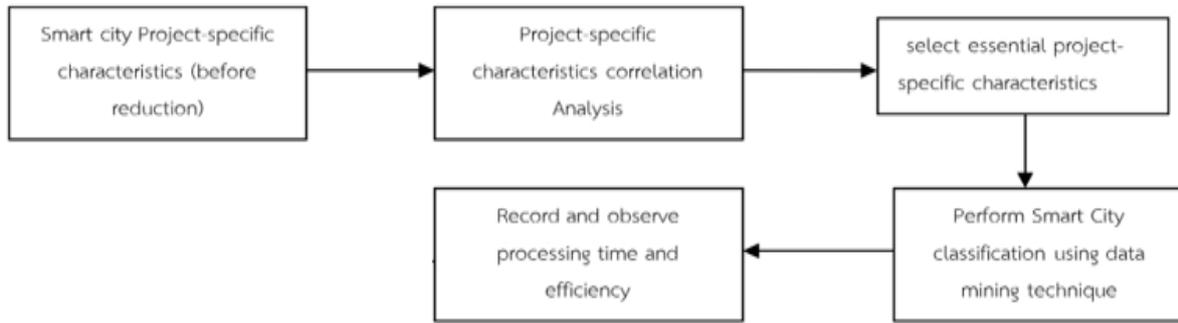


Figure 1 Conceptual framework for project-specific characteristics and smart community sub-classification analysis processing time

According to figure 1, Conceptual framework for project-specific characteristics and smart community sub-classification analysis processing time. There are procedures for studying and researching information in the work. and a system for analyzing and classifying features specific to the Smart Community project by Pichet (2017). The resulting in a more accurate understanding of the working system of the smart community smart project selection system Which has analyzed the needs of the decision-making system for smart community projects. After the development of the program is completed and can study the details of the steps under the topic research process.

Research Methodology

Correlation analysis Rule mining algorithm

1.1 Correlation analysis and Regression Analysis

Correlation analysis and regression analysis is the study of the relationship of variables. The main objective of regression analysis is estimating the value of a variable, which is called the dependent variable, commonly denoted by Y, based on knowledge of other variables called independent variables, commonly denoted by X. or in other words We use knowledge If only one variable X is used in estimating Y and the relation of Y and X is linear, we call it simple linear regression. The relationships It is the study of the degree or magnitude of the linear relationship between two variables corresponding to Kolozali, Bermudez-Edo, Farajidavar, Barnaghi, Gao, Intizar Ali, Mileo, Fischer, Iggena, Kuemper, Tonjes (2019). It is a symbolic aggregate estimation to reduce the size of a data stream and perform a quality analysis taking into account both single and multi-stream data streams. However, the measuring tool is called correlation of commonly shown in the following equation. $\text{corr}(X, Y) = \text{corr}(Y, X)$ lower letter. When r is measured as a number between -1 and 1. r is close to 1 from the two variables are highly related and have the same direction. X is large, Y is also large, and r is close to -1 from then the two variables are also very related, but in opposite directions, that is, if X is large, or X is large, Y is smaller or X is small. Then, X and Y are less

correlated and the correlation coefficient r is closest to 0. However, A key mathematical property of the Pearson correlation coefficient is that it is invariant under separate changes in location and scale in the two variables. That is, we may transform X to $a + bX$ and transform Y to $c + dY$, where a , b , c , and d are constants with $b, d > 0$, without changing the correlation coefficient. (This holds for both the population and sample Pearson correlation coefficients.) Note that more general linear transformations do change the correlation: see § Decorrelation of n random variables for an application of this.

1.2 Rule mining algorithm

Rule mining algorithm is an algorithm to search for correlation from big data to be used for the results of the analysis will be the answer to the problem. In which this type of analysis is used "Relationship Rules" (Association Rule) to find the correlation of information. There are several ways to have this relationship consistent with Datta, Mali and Roy (2017). It is an algorithmic data mining as a result of traditional mining, many rules are often created, including less or less important ones, which users expect is follow in the most important association rules.

Research Process

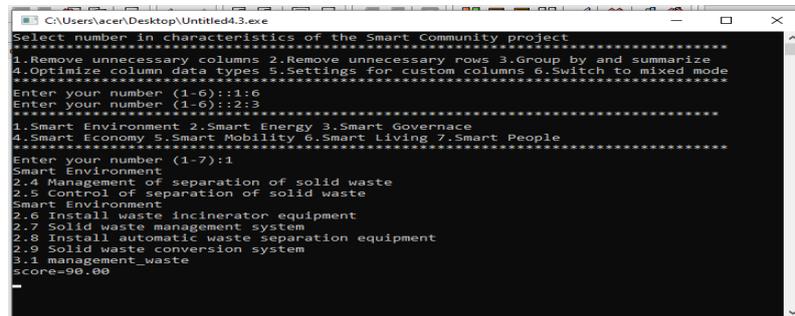
The research conducted in six steps as follows:

Step 1: Define all project-specific characteristics to be used for smart community classification. Study a research work by Supamid and Mungsing (2020), which defined project-specific characteristics for Smart Mobility, Smart Economy, Smart Environment, Smart Governance, Smart Energy, Smart living, and Smart People. These project-specific characteristics are to be used as the before-reduction input data. From the method of selecting the characteristics of the smart city project, the characteristics of various types such as Smart Energy are electric power sources to make use of various electrical devices in height and freedom. In general, more details can be studied in section 1. Smart city project specific characteristics.

Step 2: Perform correlation of those project-specific characteristics. The project-specific characteristics from in step 1 are used as input for correlation analysis, using WE-KA and MATLAB. Only strong relationship of characteristics will be considered. In general, more details can be studied in section 2. project specific characteristics correlation analysis.

Step 3: Select essential characteristics to be used for smart community sub-classification analysis. Therefore, when choosing a project, there are criteria to be considered as follows: (1) Knowledge, skills, experience, aptitude and interest in the selected project. (2) The project is appropriate and consistent with the community. (3) The project is suitable for local or community conditions. (4) The project is suitable for gender, age, capacity and the community budget (5) The duration of the project should not be too short and not more than 1 year. (6) It should be a project

that can research additional information from various sources. Can show the results of project analysis from any type of Smart Community examination at figure 2.



```
C:\Users\acer\Desktop\Untitled4.3.exe
Select number in characteristics of the Smart Community project
-----
1.Remove unnecessary columns 2.Remove unnecessary rows 3.Group by and summarize
4.Optimize column data types 5.Settings for custom columns 6.Switch to mixed mode
-----
Enter your number (1-6)::1:6
Enter your number (1-6)::2:3
-----
1.Smart Environment 2.Smart Energy 3.Smart Governance
4.Smart Economy 5.Smart Mobility 6.Smart Living 7.Smart People
-----
Enter your number (1-7):1
Smart Environment
2.4 Management of separation of solid waste
2.5 Control of separation of solid waste
Smart Environment
2.6 Install waste incinerator equipment
2.7 Solid waste management system
2.8 Install automatic waste separation equipment
2.9 Solid waste conversion system
3.1 management_waste
score=90.00
```

Figure 2 The results of the project analysis based on the criterion of Smart Community

According to figure 2 the results of the project analysis based on the criterion of Smart Community. Which the audit is finding information based on the likelihood, likeness or syntax of each Smart community database, and community suitability, etc. However, the selection method can be studied how to select it. project-specific characteristics. Which can study details in section 3. Select essential project specific characteristics.

Step 4: Perform classification using selected project-specific characteristics with data mining technique. It is a hybrid function estimation technique for creating a continuous working model with a value used to indicate the level of relationship. Researchers were able to analyze the data using tools that were used for data mining, such as Decision Tree, Naïve Bayes, and K-NN correspond to Supamid and Mungsing (2020), studied the genetic coefficient for learning algorithm in imbalanced data of software quality: A case study of the decision-making system for selecting the Smart project for initiating projects based on the concept of Smart City In general, more details can be studied in section 4. perform smart city classification using data mining techniques.

Step 5: Comparing processing time and accuracy before and after project-specific characteristics reduction. Therefore, the cleaning of time is important because the time from the action results to be used in the analysis of community smart project data in the attribute institution has to be used for cleaning. The timing of the data or the search for defects with the optimization of the interval data may occur or the area changes over time. In general, more details can be studied in section 5. Record and observe processing time and efficiency.

Project Analysis and Classification of Smart Community or Smart City Projects

1. Smart city project specific characteristics

The specific books (project characteristics) of the SmartCity project can be described as follows. (1) Smart Energy is a source of electricity to make use of various electrical equipment such as machine Heat, beauty, light etc. (2) Smart living is a city that has developed facilities based on the principle of civil architecture (Universal Design) to provide people with good health and quality of life and safe and happy in life. (3) Smart Economy is an economic system that allows individuals to make decisions on economic activities based on their abilities and opportunities based on markets and prices. (4) Smart Environment is a city that focuses on improving quality and increasing efficiency. management effectiveness and monitoring systematic environment and environment. (5) Smart People is a city that aims to develop knowledge, skills and environment. conducive to lifelong learning Reduce social and economic inequalities and open up for creativity, innovation and public participation. (6) Smart Mobility is a city that focuses on increasing convenience, efficiency and safety in travel and transportation. and environmentally friendly. (7) Smart Governance is a city that focuses on developing service systems to allow people to access government services conveniently and quickly, increasing channels for public participation. Including allowing the public to access information, resulting in transparency and verifiability.

2. project specific characteristics correlation analysis

Correlation analysis is the study of the relationship of 2 or more variables that they are related or not. How much they are related to each other the correlation analysis is divided into two types: (1) Simple Correlation Analysis. (2) Multiple Correlation Analysis.

3. Select essential project specific characteristics

Project features refer to a project that is a systematic organization of activities for the performance of organizational duties to achieve the goals effectively. Which a good project will make the efficiency of the operation and returns that the organization or agency will receive worthwhile This will bring about the development of that department. The summarized the good characteristics of the project as follows: (1) Able to meet the needs or solve problems of the organization or department. (2) Have clear objectives and goals. Able to operate and perform. (3) Project descriptions must be consistent and interrelated, that is, the project objectives must be consistent with principles and rationale. Operational methods must be consistent with the objectives, etc. (4) The details of the project can be easily understood. Convenient for implementing the project. (5) It is a project that can be put into action. Consistent with the main program of the organization and can be monitored and evaluated. (6) Projects must be based on factual information and it is information that has been analyzed carefully. (7) The project must have support in terms

of resources and management appropriately. (8) The project must have a period of time to implement. That is, must specify the start date and end of the project.

4. perform smart city classification using data mining techniques

It is a hybrid function estimation technique for creating a continuous working model with a value used to indicate the level of relationship. Moreover, it should know what kind of value relationships are structured in rule form and how to implement a set of instructions or rules of structure to create rules which can be structured in a rule form by using the equation and the relationship of an attribute display for the measure. The equation that ① from the equation of Correlation analysis is $I = i_1 \dots i_m$ and $D = t_1 \dots t_m$ ①. The equation that ② from the equation of Linear regression (Regression Analysis) is $Z_Y = B_1 Z_x$ ②. The equation that ③ from the Multiple Linear Regression equation is $g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$ ③. The equation that ④ from the equation of Correlation Analysis is $r = \frac{\sum Z=(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}}$.

5. Record and observe processing time and efficiency

Recording is a record of the processing to examine information that may make the information that is not inaccurate or inaccurate. Which method this result helps to observe and may be represented by symbols in long messages. Observe the processing time for the analysis of processing time, it is a method used to analyze data or observations that have changed in chronological order that occurred or changed in a variable over time. The nature of the change may or may not have a pattern. But if time shows up in a pattern that has changed over a certain period of time in the past in the past, it will be possible to predict in the future what kind of change should be in the future and able to forecast future data changes. Therefore, the analysis of time data is based on historical time changes. Efficiency to optimize the analysis results: (1) the amount of work is too large to exceed the capabilities of other methods, (2) the result is needed in a short time, (3) the nature of the work, has a repetitive workflow; When the processing is complicated.

6. Program (program smart community)

6.1 the screen design of the program

1. Login monitor screen the sign of the site used to login to use the system. The channel user has input user name and password to login show as figure 3 (a).

2. Main menu screen served as the center of use of the user to use functions of the website, you can add customers with button add project and the details of the customer to stay in the system show as figure 3 (b).

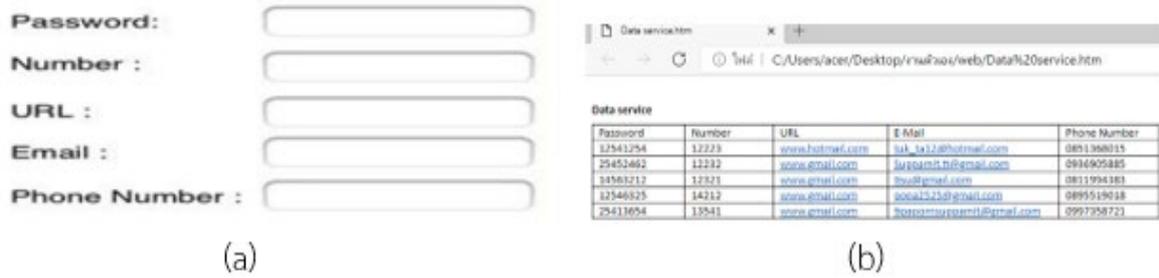


Figure 3. (a) login (b) main menu

3. Add Customer Information screen serves to add customer information, after pressing the ADD PROJECT button, there will be a space for the user to enter the customer's information to select a community project. Show as shown in figure 4 (a).

4. Screen showing the results of the considered projects acts to show a list of projects that have been considered with details. Show as shown in figure 4 (b).



Figure 4. (a) The screen shows adding information to the customer who wants to locate the distribution point. (b) The screen shows the approved project

Findings

This research is a data analysis to reduce the amount of characteristic data in the Smart Community project for initiating projects based on the concept of Smart City. Besides, the researcher used the specific data for comparison of classification methods with data mining in the form of algorithms, data reduction analysis, characterization of smart community projects, results of time analysis of project-specific data and results of data reduction methods. In this case, it is relating with the specific information of the Smart Community project, and the results of the reduction analysis and the characteristics of the Smart Community project. As the result, the researcher performed an analysis to show the results according to Figure 5-10.

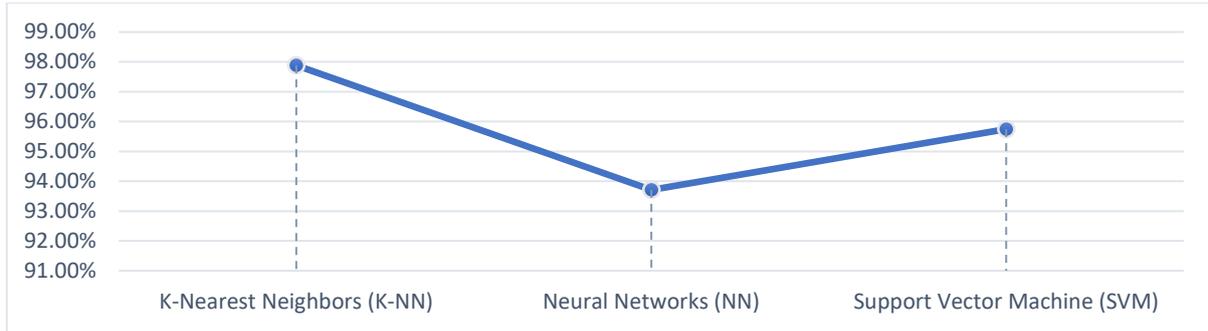


Figure 5. Comparison of Classification Methods

According to figure 5. it represented to the results of a comparison of classification methods. The techniques used in the comparison are as follows: K-Nearest Neighbors (KNN), Neural Networks and Support Vector Machine (SVM). Which class will represent a new condition or case? It was possible that the K-Nearest Neighbors (KNN) technique had a maximum accuracy of 97.88%.

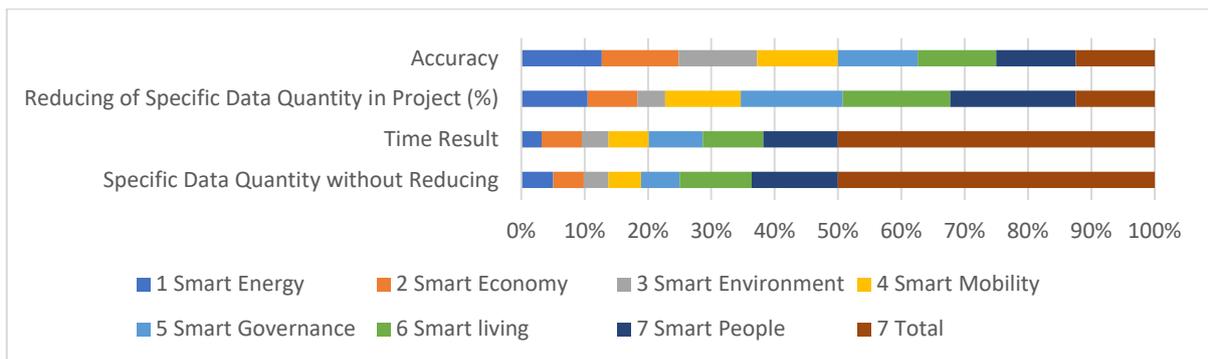


Figure 6. Reduction Analysis Results for the Smart Community Project Characteristics with the Percentage of Data Reduction in Each Feature

According to Figure 6, it represented to the results of the reduction analysis for the characteristics of the Smart Community Project with the percentage of data reduction in each feature. In this case, according to the result of data analysis it is based on the six project-specific data reduction segmentation methods to be compared: delete unnecessary columns and rows, group by and summarize, adjust column data types. optimize and set up for custom columns. As the result, it showed that researchers could use a variety of methods to obtain the most accuracy in reducing the amount of project-specific data.

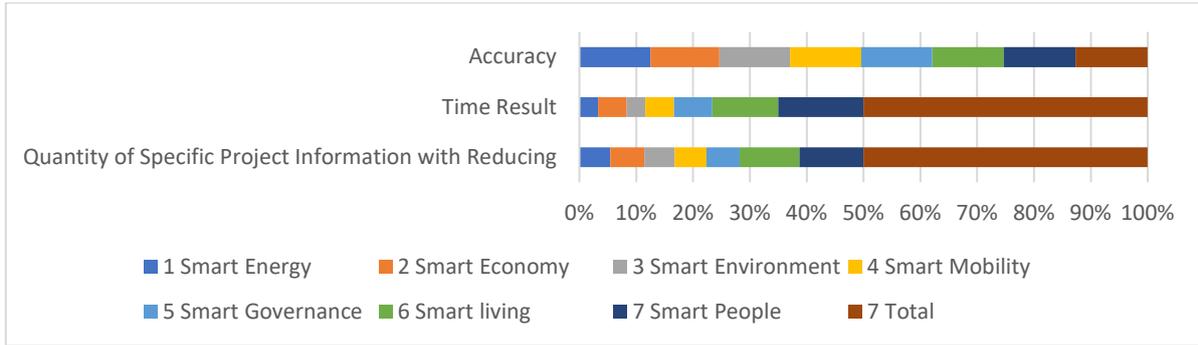


Figure 7. Project-Specific Analysis Results of Processing Time

According to figure 7, the results of the analysis of the specific processing time of the project. It is the result of data analysis from reducing the amount of information already. In order to analyze the time required for analyzing project-specific data, it was found that the time it took to process was better when the project-specific data was downgraded to the most accurate at 0:30:51 and the accuracy value at an average of 97.62% of the specifics of the projects that have already been reduced in 4702 attributes the properties (attributes).

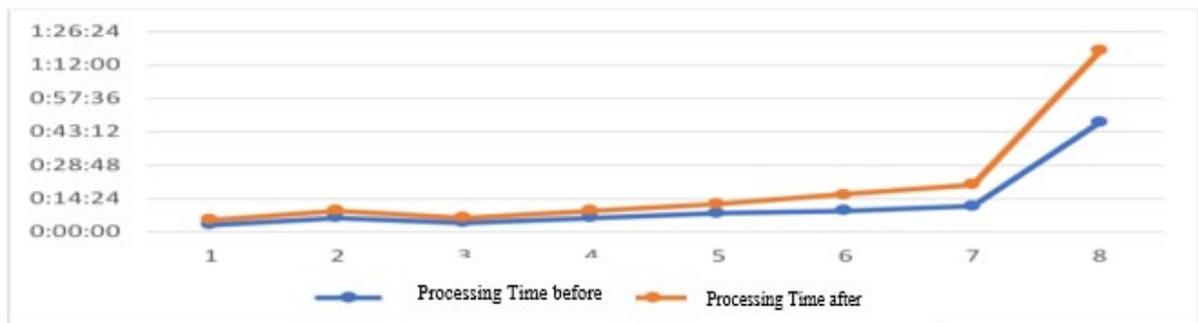


Figure 8. Compare processing times

According to figure 8, the results of the analysis of the specific processing time of the project, it was found that the time it took to process was better when the project-specific data was downgraded to the most accurate at 0:30:51.

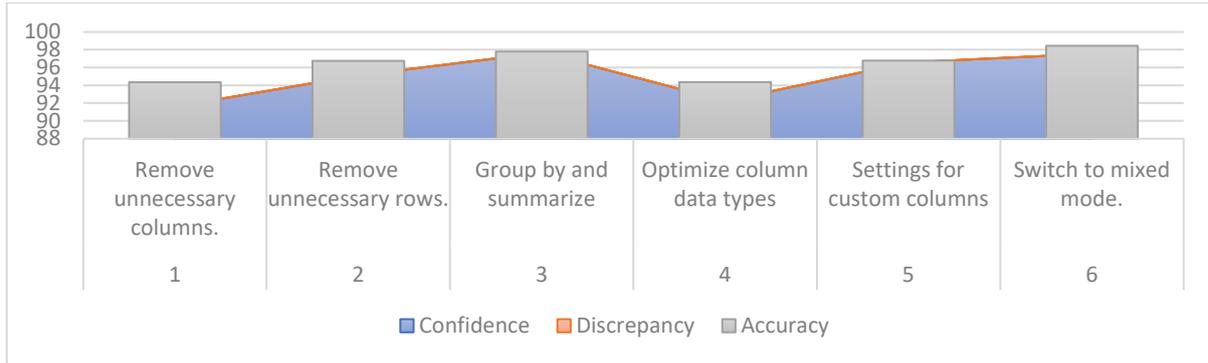


Figure 9. Reducing Method Effects of Data on the Characteristics of the Smart Community Project

According to Figure 9, the results of the reduction method on the characteristics of the Smart Community project. According to the result of data analysis and volume reduction method from the comparison of classification methods, the researchers compared the six data segmentation methods: delete unnecessary columns, delete unnecessary rows, group by and summarize, optimize column data types, and settings for custom columns and switch to mixed mode. The results showed that the follow-and-summation method gave the most accuracy at 98.45%.

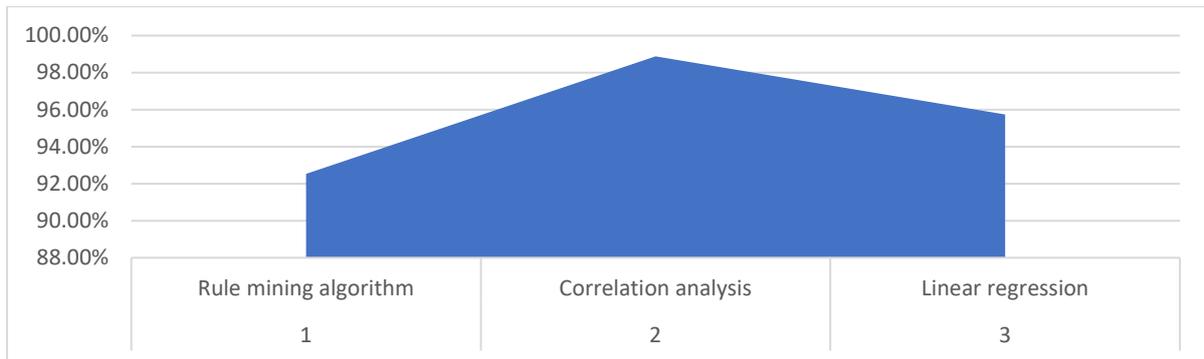


Figure 10. Reduction Analysis Results of Characteristics for the Smart Community Project from the Data Correlation

According to Figure 10, it represented to the results of reduction analysis from the characteristics of the Smart Community Project with finding of correlation data. Besides, it is the result of data analysis from the comparison of the relational value with the specific information. In this case, the researchers have compared the relationship to classify the segments with three ways for rule mining algorithm, such as rule mining, correlation analysis (Correlational), and linear regression (regression analysis). Thus, it was found a correlation analysis (Correlational) to gives the highest accuracy at 98.88%.

Discussion

The results of the research were divided into 4 categories: (1) The results of teaching comparison of classification methods, including K-Nearest Neighbors (KNN), Neural Networks and Support Vector Machine (SVM). The most accurate value of 97.88%, the Kenia Raseber data segmentation method that gave the highest accuracy was not old, used with Supamid and Mungsing (2020). described the Project-specific information conference rooms and project studies from the study of data segmentation methods, as well as Datta, Mali and Roy (2017), say that organizing links are based on the importance of knowledge discovery. (2) Analysis of data reduction of characteristics of the smart community project in 6 ways. It was found that researchers were able to use a variety of methods to obtain the most accuracy in reducing the amount of agency-specific information with Datta, Mali and Roy (2017). Efficient data and cleaning results that saddle up of data. It uses design aggregation estimation to reduce data size and automated validation methods, all single and multi-model data in reducing the amount of data. (3) Project-Specific analysis results of processing time. It was found that the time it took to perform the effect was 16.69, from the data before the reduction 7087 attribute, the time spent 47.20 downgraded the data volume 4702, the attribute (the attribute), the time used 30:51. According to Mutani, Todeschi and Nuvoli (2019), each attainment test is based on pattern data and different volumes of data and the degree of cleanliness. Availability of information and a wide range of response times. 4) Reducing Method Effects of data on the characteristics of the smart community project. The data was divided into six unnecessary elimination methods such as remove unnecessary columns, remove unnecessary rows, group by and summarize, optimize column data types, settings for custom columns and switch to mixed mode. The researchers can use a variety of methods to obtain the most accurate measurements. Reduce the amount of information for the specific project, it was found that the method for grouping according to and sentences gave the most accuracy at 97.80%. In accordance with Nattawutthisit, Sukvaree and Sincharoonsak (2016), described the testing methodology for the construction of train track indicators and the patterns of the train track and smart city, as well as Rudolf and Haindlmaier (2010), said that monitoring and manipulation to mitigate more or less conflicts depends on the information and response times on which the display is available different. (5) The reduction analysis the results of the feature for the smart community project from data correlation were found that the correlation analysis gave the most accurate at 98.88%. explains how the use of technology/tools to understand the needs of the community for training, systems, development and implementation. According to the results of the study for the project-specific factors, it represented to the analysis of 4702 attributes with showing of seven attribute features to be divided into seven groups, and each of different features has been shown in figure 6. According to the reduction results for the data

amount, it represented to the factors used in the analysis with the data characteristics for reducing of the data amount as shown in figure 5. According to the results of the processing time before and after the reduction as the data amount characteristics, it was found the good significance. According to the results of the analysis accuracy before and after reducing for the data amount characteristics, it showed of no significance including of the future research to be analyzed with the quality for the relationship rule.

Suggestion

1. Suggestion for Research Utilization

The data analysis to reduce the amount of information on the characteristics of community projects, it is involving with the initiating projects based on the concept of Smart City. In addition, it was found that the community project selection method for initiatives is based on the concept idea of Smart City by K-Nearest Neighbors method give the most accuracy to the selection and the resulting in speed and convenient for the operator and users of the program to decide on a smart community project So that the user can use the model most effective.

2. Suggestion for Future Research

In future research, the researcher would like to develop a system for analyzing and characterizing the smart community project. The system provides feedback and recommendations based on project-specific features that can enable users to immediately improve project documents in the event that the user's project does not use Smart Community 100 percent and researchers want to develop a system to receive and analyze project data in real-time analytics.

Reference

- Rudolf, G. and Haindlmaier, G. (2010). Smart cities ranking: an effective instrument for positioning of the cities. *ACE: Architecture, City and Environment*, 4, 7-26.
- Mutani, G. Todeschi, V. and Nuvoli, G. (2019). Energy Communities in Piedmont Region (IT). The case study in Pinerolo territory. *The Proceedings of 13th the IEEE International Telecommunications Energy Conference (INTELEC)*, 07-11 October 2018 at Turino in Italy, 1-8.
- Digital Economy and Society Development. (2017). *Digital Economy Promotion Agency*. [Online]. Retrieved December 9, 2022, from: <http://www.depa.or.th>. (in Thai)
- Kolozali, S. Bermudez-Edo, M. Farajidavar, N. Barnaghi, P. Gao, F. Intizar Ali, M. Mileo, A. Fischer, M. Iggena, T. Kuemper, K. Tonjes, D. (2019). Observing the Pulse of a City: A Smart City Framework for Real-time Discovery, Federation, and Aggregation of Data Streams. *IEEE Internet of Things Journal*, 6, 2651–2668.

- Datta, S. Mali, K. and Roy, P. (2018). Ranking of association rules toward smart decision for smart city. *The Proceedings of 12th International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 22-24 March 2017 at Chennai in India, 1398-1403.
- Nattawutthisit, S. Sukvaree, T. and Sincharoonsak, T. (2016). Analysis and Designing of Measurement for Smart Community in Thailand. *The Proceedings of the 11th National and International Academic Conference at Sripatum University (SPUCON2016)*, 21 December 2016 at Sripatum University (Bangkhen Campus) in Thailand, 1729-1738. (in Thai)
- Supamid, T. and Mungsing, S. (2020). Integration of techniques with data mining for community project data analysis from project specific characteristics. *International Journal of Advanced Research in Engineering and Technology*, 13, 737-750.