

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

Tipaporn Supamid¹, Surasak Mungsing²

¹*School of Information Technology, Sripatum University,*

²*School of Information Technology, Sripatum University,*

Abstract

In this paper, we proposed a new approach to create new pattern-sampling by using regression analysis form correlation coefficient equation. In this research, we used more than one variable of researching relationships for finding the similarities. Although, the regression analysis has been an outstanding successful classification in a wide diversity of application where balance class dataset partition coefficient is assumed. However, the regression analysis is ineffective whereby the minor-class instance far outnumber the original minor-class instance. Our new approach, called liner to pattern-sampling, can deal with imbalanced to effectively recognize new classification instances for better classification with the regression analysis form correlation coefficient. The results of our class prediction experiments using the proposed approach demonstrated better performance than the existing smote, Decision tree, Rule-based, Naïve Bayesian and KNN methods in terms of the k fold cross validation.

Keywords: Pattern Sampling, Reducing of Attributes, Classifying of Information, Software Quality, Smart Mommunity.

I. INTRODUCTION

Nowadays, it has seen the difficulty of the software manufacturing management project obviously in the form of software or intangible product. However, the definitions of software including of software document can be seen clearly when comparing to the building project or the automobile manufacturing project as the tangible product. Thus, the executives of the software project won't notice for the obvious effects.

The regression analysis is an extremely successful classification proposed by kelvin [23] under the presumed condition of the relationship between a dependent variable and one or more independent variables of balanced data distributions among classes. However, the regression analysis is ineffective when mining data with unbalanced class. An unbalanced dataset in which the representation between classes is not approximately equal. Then, we have the presentation theory of the regression analysis form correlation coefficient is a statistic used to show how the scores from one

measure relate to scores on a second measure [27] for the same group of individuals. A high value (approaching +1.00) is a strong direct relationship, a low negative value (approaching -1.00) is a strong inverse relationship, and values near 0.00 indicate little, if any, relationship. There are many applications in real-world domains that have innately unbalanced datasets including implementing software in a controlled and scientific [24], the various classification and clustering methods using supervised learning and unsupervised learning [28].

In the paper, we focus on the creation of model for the relationship analysis with effectiveness measurement of reducing attributes for reducing the attributes of classification for the data mining of software quality. However, the regression analysis form correlation coefficient to difference measurement or the dispersing of the information. thus, the regression analysis and regression correlation coefficient, it was found that this research has the objective to present the algorithm summation method of regression analysis form correlation coefficient for reducing attributes in data mining classification of software quality. Moreover, the programs used to analyze in this research are WE-KA and Matlab program and the techniques used to predict the equation of the correctness, reliability, performance are Decision Tree, Rule-Based, Naïve Bayesian and KNN. Besides, the results of analyzing the regression analysis form correlation coefficient are to reduce the attributes from searching from the relationship of regression analysis form correlation coefficient that are the analyzing from the relationships between 2 variables for finding the similarities between questions and model documents by searching from the patterns of designing. Then, it has the steps as these following: 1) data description 2) the creation of model for the relationship analysis of information and the effectiveness measurement of reducing attributes. 3) Preparation of Information Set to practicing and testing 4) processing of Information can indicate that the characteristic of information has the relationship pattern relevantly to the reducing of attributes. On the other hand, the reducing of attributes hasn't affected to the software quality although it has proved from the ordinary homogenous linear equation on rank n by considering the answer of it that $y = c_1y_1 + c_2y_2$. In addition, according to the answer of the homogenous equation of $c_2 = 0$, it results to the answer from multiplying with the constant values to gain the same answer

that y_1, y_2, \dots, y_m as the answer of the ordinary homogenous linear. Moreover, the value of c_1, c_2, \dots, c_m are the constant values of this equation answer with $y = c_1y_1 + c_2y_2 + \dots + c_my_m$, as same as the answer of the above homogenous linear equation. This research has the structures as this following: literature review and relevant researches, method, experimental analysis and results and conclusion.

I I . LITERATURE REVIEW AND RELEVANT RESEARCHES

Software quality is products or services that the organization will hand in to the customers to gain the good quality of software. [1] Then, it conforms to the terms for applying with the requirements or the agreements. Thus, the software should have the correct property, good credit, easiness to use, simplicity to maintain, changeability, application with works and others. Currently, it has the development of software measurements for software product attributes with processes. Therefore, the planning steps and the controlling of software development with the values to make comparison to the objectives are based on to the standard to quality assessment. Besides, it is in the way of the saying of Lord Kelvin [23] that "when you want to indicate in what you say with numbers, it may not perform in what you want. "Thus, the management and the satisfaction are the beginning of the missing knowledge for developing the software to apply with the higher step of science. Then, if you can't indicate the result, you can't manage with it. However, the errors of software have the important role to produce software with quality. [24] On the other hand, although the software pattern still has the errors, it can be brought to develop in the high level as well as maintenance, costs assessment and quality reduction. In addition, it can specify the errors of software in the beginning until the ending of software development. After that, it will know about the occurrence problem to reduce the failure of software development and then it can increase the confidence to the customers that these products have the quality with the requirements. On the contrary, it can classify the information as the big learning as two types that are the learning with teachers and the learning without teacher. Moreover, it can apply the information for each class to make comparison with the learning without teacher, such as using clustering, grouping and others. On the other hand, the classifying of learning with teachers is one important technique, and it is popular to classify the information set that is called Decision trees. Similarly, it is one method to take estimate of function with not continuous value. Additionally, the Decision tree will classify the information set in each instance of each node for each attribute until gaining its own value as the variable set or attribute-value pair. After that, it will predict the type of Decision tree beginning with the root node for testing the variable value of each node and then continue to the branches of tree for determining the value before connecting to the next node. Consequently, the testing will be performed until meeting the leaf node for expressing the one prediction result or more than one prediction result which is called Decision trees. Then, it has the studying of scope to memorize the patterns and the learning tools with the work selection as the learning scope with this following: 1.

J48 Algorithm as the technique of Decision trees by using the selection of J48 and 2. CART Algorithm as the technique of Decision trees with using the selection of CART or Classification and regression tree. What's more, the Decision trees is in the form of binary or the creating to classify the node into the repetitive two nodes for determining the beginning root 3. BFTree Algorithm as the technique of Decision tree to apply with the selection of C4.5 4. Random Forest Algorithm as the technique of Decision trees as the selection of Random Forest and 5. Naive Bayesian Algorithm as the technique of Bayes to apply with the selection of Naive Bayesian. However, the technique for data classifying namely naïve bayes random forest (NBRF). [17] The decision process starts by extracting the difficult designate data, if any and follows by a construction of learning model for reclassifying them. There are research performed the first step by using two algorithms; the naïve bayes paralleled with the random forest. The data getting different answers from the two algorithms will be selection as the difficult designate data. These data are lying around the border of the different classes. The experiments performed on 10 benchmarks; 6 sets are synthesized data including clus 1000, clus200, rand1000, rand200, pat1 and pat2 data, and 4 sets are real world data including vowel, hepato, iris and kla-azar data. The result showed that the NBRF technique outperformed the existing model. The average performance of NBRF is better than those of fuzzy c-mean random forest 4.81%, fuzzy c-mean decision tree 13.57%, random forest 3.91%, naïve bayes 17.15%, and FCM 35.54% and the efficiency of stock trading signal prediction is an enhancement factor for the investor to get more gain from their decision making. [18] Applied the computer intelligence approach for trading signal prediction generating which the method of discovered the hidden pattern cluster, which represent the characterization of trading significant, in dynamical multi-dimensional phase space by genetic algorithm is the proposed approach in this paper. The hidden pattern is a set of data, lying which in the hyper sphere radius of which the present data point is the center. Instead of searching overall area of phase space, using the last point of data as the center of hyper sphere is the method that improved the better solution for stock prediction problem and saving the access time for machine operation as the same time. This approach is so called "locality-based genetic algorithm". In local pattern evaluation process, the k-nearest neighbor algorithm and applied singular value decomposition with principal component regression are introduced to reduce the noise and obtain more resilience for the prediction solution. The result shows that these methods can outperform the accuracy rate of prediction; the average rate is more than 60% of trading decision making. The model of automated web services composition by using rule-based engine. [19] Rules are modeled according to user's preferences and learnt from business constraints. The services are composed automatically by these rules. Consequently, it generates BPEL document which match exactly with user's requirements and is able to proceed on BPEL engine. The method to data classification of malaria in plasmodium vivax by texture features and k-nearest neighbor (KNN) [20]. The basic idea is to segment blood cell image and to classify a species of malaria in plasmodium vivax. The segmentation of blood cell image used texture

features and fuzzy c-mean. The texture features in segmentation are derived by using gray level co-occurrence matrix. The data classification used texture features and k-nearest neighbor. The texture features in classification is derived by law's texture. Resulted from law's texture have 25% , so the sequential backward selection which is suboptimal searching techniques for selecting texture features for law's texture. The species of plasmodium vivax is classified by k-nearest neighbor. In our experiment, 32 images are tested and compared with expert physical. The results show that our technique performs to data classification correctly more than 81%. The optimal inventory of the short-shelve life products is very important. These products are required on accurate sale forecast; in other words, maintaining the inventory for products those responses to the customer demand [21]. Since the product inventory affects the financial flow of the company both from revenue of products out of date. This study, the apply rule - based expert system to improve the accuracy of forecast compared to those in the literature using some statistical techniques alone. The demand of green cabbage for the modern trade store is used for our case study. The results have shown that the proposed rule based forecasting technique can simulate the product demand more accurate and thus can be used as a decision support tool to increase the operational profit of the company.

II.I Smart Community

Intelligent Community is the development of communities with an intelligent community management system. [30] Which genius means Bringing information and communication technology, Internet of Things technology, and Big data collected by the system to develop various aspects of the community. For effective control Including increasing security and meeting international standards Which has the following components of the Smart Community.

- (1) Smart Energy is a balancer of energy systems in that society. Especially Clean energy such as Solar and wind energy, which the energy can be stored in the form of batteries for backup and can be used immediately when in shortage. In addition, in that society Also have the said energy monitor center for proper use according to principles demand-supply at that time.
- (2) Smart living is to improve the quality of life of the people in the community. Nowadays, there are a lot of elderly people in the society and due to the change of lifestyle Many seniors have to live at home alone while their children are out and working. Smart Living platforms should be developed. By focusing on the creation of a complete ecosystem to help solve the problem of caring for the elderly Efficient service And to sustainably strengthen the health services of the population.
- (3) Smart Economy focuses on the development of community enterprises that can be driven through cooperative systems by using ICT as a tool for

productivity development. And expanding the market By increasing distribution channels and complete distribution of products and services.

- (4) Smart Environment, city that takes into account The impact on the environment and climate change by using technology to help manage systematically, such as water management, weather management, waste management And disaster monitoring, as well as increase public participation in the conservation of natural resources.
- (5) Smart People is a way of life that people will continually change. With modern technology being used in daily life continuously Digital and technology will be important mechanisms for development in various fields, which must be constantly learning. Must be an analytical person Have the ability to solve problems Have skills in information technology Creative and innovative skills Knowing the integration of science and knowledge in various fields to benefit And able to adjust to live with others and live in a digital society happily.
- (6) Smart Mobility is a traveling around. With internet technology, enabling data to be linked in various fields Such as passenger, transportation, traffic, etc. In addition, the data combined into Big Data becomes important data that can be analyzed to search for Insight and provide solutions to help solve problems directly. This group of solutions is called Smart Mobility. It is a technology that helps solve problems in various forms such as the geographical location of the vehicle, road conditions, convenience conditions. And passenger safety In order to manage the said information, it is necessary to use a computer with high efficiency and to help manage the said information, it needs the computer with high efficiency.
- (7) Smart Governance is a city that develops public service systems To facilitate For the people who have the right to access to government information by focusing on transparency and participation And continuously updated through the application of service innovations.

II.II Research related to smart community operations

Analysis and design smart indicator for smart community in Thailand [SukSawat Natthawutisit, Thanasukwaree and Thitaporn Sinjaroonsak, 2018]. This research presents the concept and form of Som Communicate Universal to apply to the Economic and National Development Plan 2015 – 2018 Community development to be a genius at a local level still lacks standard and acceptable indicators.

Smart cities and the future internet: towards cooperation frameworks for open innovation. [schaffera, 2011] This research is presented today. Cities face complex challenges to meet objectives related to economic and social development and quality of life. The concept of "smart city" is to respond to challenges. This article explores "Smart City" as an open and driven innovation environment for users to test and

validate future internet-enabled services. From the current landscape analysis of the smart city pilot project, future internet-driven research, and projects in the Living Labs domain, common resources related to research and innovation can be identified that can be shared in an open innovation environment. The efficient sharing of these common resources for the purpose of creating innovative ecosystems in cities and regions requires sustainable cooperation and stakeholder collaboration strategies.

Smart Cities: Definitions, Dimensions, Performance, and Initiatives [vito albino, umberto berardi and rosa maria dangelico, [2015]. This research presents the concept of smart cities becoming increasingly relevant for both scholars and politicians. Policy Despite this, there is still some confusion about what intelligent city is like, many people tend to use it interchangeably. This article aims to explain the meaning of "smart" in the context of the city through an in-depth review of relevant literature, including official documents from international institutions. It also specifies dimensions and key elements that characterize intelligent cities. Different indicators of intelligence in a city have been examined to show the need for a shared definition of what is considered intelligent city.

II.III McCabe and Halstead complex measures.

Halstead's complex measures are software indicators proposed [26] by Maurice Howard Halstead in 1977 as part of the empirical science of software development. Halstead notes that software metrics should reflect their use. The use or expression of the algorithm in various languages that researchers have studied. But will not depend on usage under a particular platform, as these metrics are calculated statically of the code. However, The goal of Halstead is to identify software measurable properties and the relationships between things that are similar to specifying the measurable properties of matter. (Such as volume, mass and gas pressure) and the relationship between classes (Similar to gas equations), therefore, the measurement is not just a measurement complexity.

II.IV McCabe's cyclomatic complexity.

The Cyclomatic complexity Report (or the McCabe Complexity Report) presents Cyclomatic complexity (a broad measure of soundness and confidence for a program) [25] for the selected project entity. Cyclomatic complexity measures the number of linearly-independent paths through a program module.

III. METHOD

This research has presented about the factorial procedure analysis to develop the software quality management. Then, it has applied the relationship rule of regression analysis for algorithm as the detail of requirements and behaviors of the executives in the software project. Besides, it has the

procedures in relationship analysis and making comparison to reduce the attributes as this following:

III.I Pre-Processing

According to the information from questionnaires of the basic source and PHP programs, it can analyze this information to prepare the pre-processing. Then, it can prepare for process the result with correctness and accuracy. Besides, the researcher has focused on the part of factorial information of software quality with the steps as this following: 1) The extraction of missing data to gain the most important information only 2) The separation of information from the questionnaires with the same subject or in the similarities as the partition data. 3) It can make data transform to give the correct answer equally to "1" as the main reason. Similarly, it can cut the non-relevant answer in the form of "Don't care". In addition, other answers can indicate to the denying which isn't beneficial to find the relationship of information. Then, it results to the dividing of 2 main ones with 7087 records as the main function for applying the software quality management. Significantly, it has 7 attributes with the required groups and 262 attributes of the software quality for general behaviors.

III.II Information Selection

Information Selection is the selection in the pattern of selecting column with quite full information. Moreover, in each column it should have the value for every row as the same one. Besides, the value in each column shouldn't be repetitive, but it should edit information to be correct and perfect. Thus, it should adjust information to be suitable for making decision and setup the information group to reduce the dispersing of information. Therefore, it can find the Means to collect samples of questionnaires prior to find the Means to estimate the values of population or μ . As the result, it should determine the Confidence Interval on the Mean: Variance known is $l \leq \mu \leq u$ When l is Lower limit and u is upper limit, we are called as "Two-sided Confidence interval" by finding the Median of populations inside limit from the solution is $P(L \leq \mu \leq U) = 1 - \alpha$. thus, In this case α is Error risk or the value to refer to the risk from prediction the value of μ with error, the reliability value and the risk value are $100(1 - \alpha)\%$ or called as the percent of reliability. Similarly, it has the value called as One-sided confidence interval with the solution is $\leq \mu$. So that, according to the above solution, we are called as lower-confidence interval with the value of l or called as lower limit with this following $\mu \leq u$ and according to the above solution, it is called as upper-confidence interval by the value of u is the value of upper limit. Thus, the value $100(1 - \alpha)\%$ is called as the reliability of parameter m . When we refer to the theory of Sampling distribution for mean, it has the median value of m and the dispersing value equally to σ^2/n . Therefore, it has the value is $Z = \bar{x} - \mu / \sigma / \sqrt{n}$ When disperse value as Normal, the value α should be divided by 2. So that In the case of finding Two-sided confidence interval from the figure, it can be concluded as this following $P\{-Z\sigma_{1/2} \leq Z \leq$

$Z_{\sigma/2}\} = 1 - \alpha$ When replace with the value z, it has the value as this following:

$$P\left\{-Z_{\sigma/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq Z_{\sigma/2}\right\} = 1 - \alpha$$

According to the setting up of the new solution, it has the value as this following:

$$P\left\{\bar{x} - Z_{\sigma/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\sigma/2} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

III.III The creation of model for the relationship analysis of information and the effectiveness measurement of reducing attributes.

After, it has passed the procedure of preparation, selection and processing of information from PHP programs. The researcher has created the procedure model for analyzing and measuring effectiveness. Thus, it can reduce the attributes with 3 parts of presentation as this following:

1. The information from the result of pre-process to classify in the information parts in the procedure of factorial analysis to develop the software quality. After that, it can bring the information into the procedure to find the relationship rule of regression analysis. Then, it can classify this information as the result dividing into the minor parts as the main function to apply the software quality. Additionally, it includes with the required groups and behaviors of applying the software quality. Furthermore, the researcher has made the assumption that the result from this procedure can indicate with the reducing of the attribute numbers with better effectiveness. Thus, the equation will be available in written form below.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \textcircled{1}$$

2. The procedure to reduce attributes can be applied with the method of regression analysis form correlation coefficient. Significantly, it is the technique to indicate of the relationship between information to measure the differences or the dispersing of information. Then, if it has more differences of the value, the value to be calculated will be higher. Oppositely, if the information has many similarities, the values will be low. Therefore, the information combined to this procedure is in the same information with the heading of 3.2.2 because this procedure is the way to find the result from the characteristic of each attribute or the dimension of information and it should remember about the differences of each attribute. Therefore, the equation will be available in written form below:

$$Y = \bar{Y} + r \frac{S_Y}{S_X} (X - \bar{X}) \quad \textcircled{2}$$

3. The development of regression analysis from correlation to the measurement procedure and the comparing of effectiveness between the regression analysis and regression analysis form correlation coefficient. Then, the relationship rule of regression analysis from correlation coefficient is the way to reduce the dimension of information. Therefore, from

equations 3.2.1 and 3.2.2 can be rewritten as the following equation.

$$Y = [\beta_0 + \beta_1 X_i + \varepsilon_i] \left[\bar{Y} + r \frac{S_Y}{S_X} (X - \bar{X}) \right] \quad \textcircled{3}$$

III.IV Preparation of Information Set to practicing and testing

The preparation setup to gain the most perfect information which can bring the preparation information to divide into 2 groups that are the practicing set and testing set with these details:

- 1 Practicing Set : it can divide into 2 of 3 ratios of all information.
- 2 Testing Set: it can divide into 1 of 3 ratios of all information.

III.V Processing of Information

The processing of information for both practicing sets and tests have applied WE-KA and Matlab program to make comparison of the accuracy and the correctness in making prediction of the software quality of presentation as this following:

- Decision Tree by using J48 algorithm in We-Ka program and selecting Tree Menu and J48
- Rule-Based by using Rule-Based with C4.5 technique in We-Ka program and selecting Rule and part.
- Naïve Bayesian in We-Ka program by selecting Bayes and Naïve Bayesian.
- KNN in We-Ka program by selecting Lazy Menu and Kstar.
- Regression Analysis form correlation coefficient by selecting matlab.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

The results of this research can predict the quality of the Smart Community Project Decision Making software by hypothesis processing. In addition, we use the WE-KA and Matlab program. The results of the analysis and comparison of the research results can be divided into 3 parts as follows.

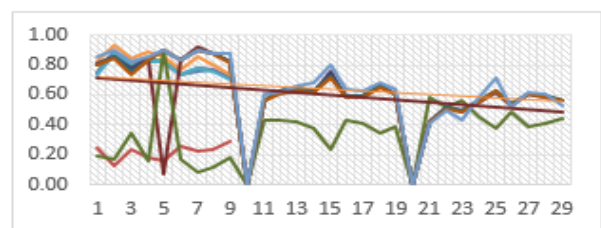


Figure 1. Comparison of datamining for correctness

Figure 1, the comparison of the data mining for correctness in applying among the techniques of Decision Tree, Rule-Based,

Naïve Bayesian, KNN and correlation coefficient. Besides, it was found that the correctness, the overall of effectiveness and ROC values with applying the techniques of Decision Tree and Rule-Bases were in similar way. However, according to the result of Rule-Based value, it was found the high level with the Means of 89 percent and the increasing Means with 90 percent. Besides, according to the selection of information totally 2 in 3 because balanced information will make learning on each floor. Can show results about predicting the best results.

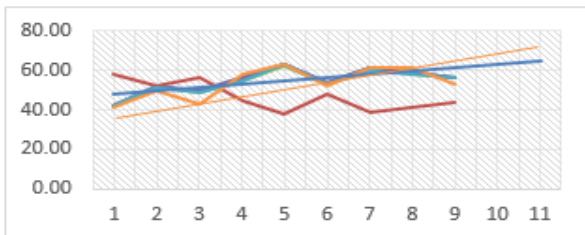


Figure 2. Comparison of datamining for reliability

Figure 2, the comparison of the data mining for reliability in applying among the techniques of Decision Tree, Rule-Based, Naïve Bayesian, KNN and correlation coefficient. Besides, it was found that the reliability, the overall of effectiveness and ROC values with applying the technique of Decision Tree were quite high with the Means of 68 percent and the increasing Means with 69 percent. Besides, according to the selection of information totally 2 in 3 because balanced information will make learning on each floor. Can show results about predicting the best results.

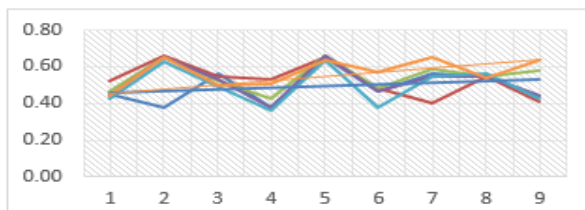


Figure 3. Comparison of datamining for performance

Figure 3, the comparison of the data mining for performance in applying among the techniques of Decision Tree, Rule-Based, Naïve Bayesian, KNN and correlation coefficient. Besides, it was found that the performance, the overall of effectiveness and ROC values with applying the technique of Decision Tree were quite high with the Means of 78 percent and the increasing Means with 80 percent. Besides, according to the selection of information totally 2 in 3 because balanced information will make learning on each floor. Can show results about predicting the best results.

V. PROOF OF MODEL FOR INFORMATION AND REGRESSION ANALYSIS FORM CORRELATION COEFFICIENT

It can bring the ordinary homogeneous differential equation in n rank to write as the answer to be proved as this following:

$$y(x) = c_1y_1(x) + c_2y_2(x)$$

If it can prove that y_1 and y_2 is the answer of equation that c_1 and c_2 are the constant value, it will result in as this following:

$$y = c_1y_1 + c_2y_2$$

Because y_1 and y_2 are the answer of equation, it will result in as this following:

$$a_0(x)y^n + a_1(x)y^{n-1} + \dots + a_{n-1}(x)y' + a_n(x)y = 0$$

Thus, it can replace by the value (a) into the equation of

$$a_0(x)y^n + a_1(x)y^{n-1} + \dots + a_{n-1}(x)y' + a_n(x)y = 0$$

With this following:

$$a_0(x)(c_1y_1 + c_2y_2)^n + \dots + a_{n-1}(x)(c_1y_1 + c_2y_2)' + a_n(x)(c_1y_1 + c_2y_2) = 0$$

$$= a_0(x)(c_1y_1^n + c_2y_2^n) + \dots + a_{n-1}(x)(c_1y_1' + c_2y_2') + a_n(x)(c_1y_1 + c_2y_2)$$

$$= c_1[a_0(x)y_1^n + \dots + a_{n-1}(x)y_1' + a_n(x)y_1] + c_2[a_0(x)y_2^n + \dots + a_{n-1}(x)y_2' + a_n(x)y_2]$$

$$= c_1(0) + c_2(0) = 0$$

Thus, $y = c_1y_1 + c_2y_2$ is the answer of homogenous equation as the special and interesting case of homogeneous equation that is if $c_2 = 0$. Then, the result can be brought to multiply with the constant value for gaining the same answer that is y_1, y_2, \dots, y_m , as the answer of homogenous equation and c_1, c_2, \dots, c_m are the constant values as the answer of the equation as this following:

$$Y = c_1y_1 + c_2y_2 + \dots + c_my_m,$$

Completed for the proof.

VI. CONCLUSION

The comparison of decision tree techniques, rules, Naïve Bayesian, KNN and correlation coefficient, using WE-KA and Matlab program. In addition, it is found that the analysis of the algorithm for regression analysis from correlations is used for the classification of data excavation of software quality from the decision making system in selecting the community smart project for initiating projects based on the concept of Smart City. After that, according to the analysis of each technique to get very accurate results with the accuracy of the software quality should be equal to 90 percent, 80 percent and 69 percent respectively. However, from the regression analysis from correlation analysis to measure the difference or distribution of the data, it was found that the data of the search algorithm and regression analysis from correlations indicate the characteristics of the data in the form of related relationships to reduce the features. Therefore, reducing the features will not affect the quality of the software. Therefore, when the proof of ordinary homogeneous equations in things from the consideration of the second answer can be treated as $y = c_1y_1 + c_2y_2$ in the form of the homogenous equation answer with $c_2 = 0$ or the result from multiplying the answer with the constant value to gain the same answer with y_1, y_2, \dots, y_m , as the answer of homogeneous equation and c_1, c_2, \dots, c_m with the constant value or the answer of equation

as $y = c_1y_1 + c_2y_2 + \dots + c_my_m$ in the same way as the above answers for linear homogeneous equations. Lastly, additional research is the Time Management Model for Software Project Management for Executives and to Offer Software Quality Measurement from the decision making system in selecting the community smart project. For initiating projects based on the concept of Smart City.

REFERENCE

- [1] Crawford JK. Project management maturity model. Boca Raton, FL: Auerbach Publications; 2007.
- [2] Kerzner H. Using the project management maturity model: strategic planning for project management. John Wiley & Sons; 2019 Mar 19.
- [3] <http://www.amazon.com/Project-Management-Maturity-Equations-Research/dp/0849379458>.
- [4] Callegari DA, Bastos RM. Project management and software development processes: integrating RUP and PMBOK. In2007 International Conference on Systems Engineering and Modeling 2007 Mar 20 (pp. 1-8). IEEE.
- [5] Chen L, Bi L, Si H, Zhang J, Ren Y. Research on prediction method for pivotal indicator of hospital medical quality using decision tree. In2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS) 2015 Sep 23 (pp. 247-250). IEEE.
- [6] Im Tan L, San Phang W, Chin KO, Anthony P. Rule-based sentiment analysis for financial news. In2015 IEEE International Conference on Systems, Man, and Cybernetics 2015 Oct 9 (pp. 1601-1606). IEEE.
- [7] Supakpong J, Nuanwan S. Hierarchical text classification: A restructured approach, Transactions on engineering, computing and technology, Thailand. 2004:24-29.
- [8] Kerdprasop N, Kerdprasop K, Saiveaw Y, Pumrungrong P. A comparative study of techniques to handle missing values in the classification task of data mining. In29th Congress on Science and Technology of Thailand, Khon Kaen University, Thailand 2003 Oct.
- [9] Bakar AA, Kefli Z, Abdullah S, Sahani M. Predictive models for dengue outbreak using multiple rulebase classifiers. InProceedings of the 2011 International Conference on Electrical Engineering and Informatics 2011 Jul 17 (pp. 1-6). IEEE.
- [10] Aslandogan YA, Mahajani GA. Evidence combination in medical data mining. InInternational Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. 2004 Apr 5 (Vol. 2, pp. 465-469). IEEE.
- [11] Bibi KF, Banu MN. Feature subset selection based on Filter technique. In2015 International Conference on Computing and Communications Technologies (ICCCT) 2015 Feb 26 (pp. 1-6). IEEE.
- [12] Bandyopadhyaya I, Babu D, Kumar A, Roychowdhury J. Tactile sensing based softness classification using machine learning. In2014 IEEE International Advance Computing Conference (IACC) 2014 Feb 21.
- [13] Xueli W, Zhiyong J, Dahai Y. An improved KNN algorithm based on kernel methods and attribute reduction. In2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC) 2015 Sep 18 (pp. 567-570). IEEE.
- [14] Taneja S, Gupta C, Aggarwal S, Jindal V. MFZ-KNN—A modified fuzzy based K nearest neighbor algorithm. In2015 International Conference on Cognitive Computing and Information Processing (CCIP) 2015 Mar 3 (pp. 1-5). IEEE.
- [15] Ludwig SA, Jakobovic D, Picek S. Analyzing gene expression data: Fuzzy decision tree algorithm applied to the classification of cancer data. In2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) 2015 Aug 2 (pp. 1-8). IEEE.
- [16] Singh P, Pal NR, Verma S, Vyas OP. Fuzzy rule-based approach for software fault prediction. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2016 Feb 19;47(5):826-37.
- [17] Dacha Sanyota, KS, Siraput C. Naïve Bayes random forest: high classification efficiency model. Thailand. 2008:28-35
- [18] Tunpana, S. Prediction stock trading signals using locality-based genetic algorithm. Thailand. 2008:52-59
- [19] Sukkanat Numtong, B.L. Automated web sevice composition using rule-based engine. Thailand. 2008:202-209.
- [20] Wilaiporn L, Krit S. Plasmodium vivax image classification by texture features and k-nearest neighbor. Thailand. 2008:279-286.
- [21] Siriporn P, Kanongnit K, Nagorn I. Finding optimal inventory using rule-based forecasting technique. Thailand. 2008:396-403.
- [22] <http://promise.site.uottawa.ca/SERpository/>
- [23] Fenton NE, Pfleeger SL. Software Metrics, A Rigorous & Practical Approach. International Thomson Computer Press, London. 1997.
- [24] Monica C. Evolutionary Decision Trees and Software Metrics for Module Defects. International Journal of Computer and Information Science. 2008;2:273-277.
- [25] <http://www.cs.waikato.ac.nz/ml/weka>.
- [26] <http://promise.site.uottawa.ca/SERpository/>
- [27] National Council on Measurement in Education. http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorC
- [28] Chug A, Dhall S. Software defect prediction using supervised learning algorithm and unsupervised learning algorithm. New delhi. 2013:173-179.
- [29] Happy Sawadee Natthawutisit T, Thitaporn S. Analyze and design indicators of smart community of Thailand. Bangkok. National and international academic conferences Sripatum University. 2016;21:1729-1738.

- [30] Durongkaveroj P. Digital Economy Promotion Agency: DEPA. Minister of Digital Economy and Society. Ministry of Digital Economy and Society. 2017. Retrieved from <http://www.depa.or.th>
- [31] Secretariat of the National Library of Thailand Office of the National Economic and Social Development Board. (2018-2580). Retrieved from https://www.nesdb.go.th/download/document/SAC/NS_SumPlanOct2018.pdf
- [31] Schaffer H, Komninos M, Trousse B, Nilsdon M, Oliveira A. Smart cities and the future internet: towards cooperation frameworks for open innovation. Future internet assembly. 2011:431-446.
- [32] Henry Mintzberg and Fances. Handbook of Decision Making. Wiley. 2012:1-772.
- [33] Brinkers HS. Decision-making. The Ohio State University press All Rights Reserved. 2008:1-286.
- [34] Barlow M, Levy-Bencheton C. Smart cities, smart future: Showcasing tomorrow. John Wiley & Sons; 2018 Oct 23.
- [35] BIS. The smart city market: opportunities for the UK. Department for business, innovation and skills. London, UK, BIS Research 136. 2013.
- [36] BLACK & VEATCH. Starting a smart community evolution, the journey to Urban system resilience. 2014.
- [37] Chourabi H, Nam T, Walker S, Gil-Garcia JR, Melloul S, Nahon K, Scholl HJ. Understanding smart cities: An integrative framework. In 45Th Hawaii international conference on smart cities will be enabled by smart it. 2012:2289-2297.
- [38] Giffinger R, Gudrun H. Smart cities ran. 2010.