

AMPM₃ Criteria of Algorithm Summation for Classifying Datamining of Software Quality Management

Kattiya T. Yangyuen^{1*}, Tipaporn Suppamit², Surasak Mungsing²

¹Rattana Bundit University, 306 Ladproud Rd., Bangkok, Thailand.

²School of Information Technology, Sripatum University, 2410, 2 Phahonyothin Rd., Senanikhon, Chatuchuck, Thailand.

* Corresponding author. Tel.: 063-652 4642; email: kattiya.kty@gmail.com

Manuscript submitted July 24, 2018; accepted March 8, 2018.

doi: 10.17706/jsw.14.1.36-46

Abstract: This research has the objective to present the algorithm summation method of AMPM₃ criteria for reducing attributes in data mining classification of software quality management. Moreover, the programs used to analyze in this research are WE-KA and MATLAB and the techniques used to predict the equation and the accuracy are Decision Tree, Rule-Based, Naïve Bayesian and KNN. Besides, the results of analyzing the Algorithm with AMPM₃ criteria are to reduce the attributes from searching from the relationship of regression analyzing and the regression analysis 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) Preparation of Information 2) Selection of Information 3) Practicing Information Set and Test 4) Processing of Information 5) Creation of Model for analyzing the information relationship and effectiveness measurement to reduce attributions.

Key Words: Reducing of attributes, classifying of information, software quality management.

1. 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 software project executives can't notice for the clear effects. Currently, it has much difficulty of operation for software project management or the intangible product although it has given the definition of the software including of the software Procedure for Paper Submission documents when comparing to the building project or the automatic manufacturing project as the tangible project to see it clearly. Thus, the executives of the software project won't notice for the obvious effects. According to this reason, the researcher will propose for the new concept idea for reducing the attributes of classification for the data mining of software quality management. However, according to the Algorithm Summation of AMPM₃ criteria with analyzing to difference measurement or the dispersing of information, it was found that the Search Algorithm and AMPM₃ criteria 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 management 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, research patterns, results and conclusion.

2. Literature Review and Relevant Researches

According to studying the theories and relevant researches, it can make understand about the principles of each theory to apply in this research. Moreover, it comprises with the software quality management , the techniques of Decision Tree, Rule-base, Naïve BayesianUnits and KNN and the relevant researches with the explanation details as this following:

2.1. Software Quality Management

The Software quality management refers to all management activities to be determined as the quality policies, objectives and quality responsibility. Then, it can bring to apply with methods, such as quality planning, quality assurance, quality controlling and quality development. [1] However, the software quality management is the way to bring the work in the part of software quality management, such as software product processes and applying resources to apply. [2] Similarly, it can determine the standards and work processes in the proper way no matter of the small or the large system. Consequently, it is vital to bring the software quality management to apply with the difference for only the work scopes. In addition, for the large system of software quality management and the activities as this following:

a) Quality Planning

The quality planning setup process for the project. Then, inside this work plan it should determine the definitions of these quality software attributes and explain these quality assessment methods to be selected by the teamwork with the suitable standard for each product and process to take assessment.

b) Quality assurance

The quality assurance is still the way to maintain the product quality for the manufacturing and the maintenance times. Then, it has operated with activities in each step to make the software test activity to be the important one for quality assurance. Moreover, the important thing of quality assurance is to determine the standard for applying in manufacturing process and bringing to apply with software products.

c) Quality control

The quality review of documents and the software processes have been operated by the teamwork with assignments. Then, the teamwork will check whether the project has been operated with the standard including of checking whether the documents and the software products have been used directed to the standard. Besides, it will check whether each value has deviated from the standard for recording information. Then, the executives can inform to the responsible people about the mistakes. Therefore, it includes of models, processes, test plans, configuration management processes, process standards, manuals and others. However, the comparison of document quality and the software products are in the good standard. Moreover, it can use the automated software tool by spending less time than the first method. On the same way, it is the method to be made with the comparison. Consequently, it should measure the software attributes to bring the gained values which are mostly in the quantity ones for comparing to the standard to take assessment of the software easier. Then, the measurement of software attributes is relevantly to the software assessment to control the software quality.

d) Management oversight

Improvement procedures are in place and utilized. Lessons learned are regularly examined and used to improve documented process. Projects are given high value within the organization, thus there is high visibility

to the individuals who are actively involved in projects on a regular basis. All project, changes, and issues are evaluated based upon efficiency and effectiveness metrics and management takes an active role in management oversight and executive sponsorship of projects within the organization.

e) Awareness and support

An improvement process is in place to continuously improve management awareness and support of corporate project and their need. Lessons learned are captured and used to improve the monitoring and control efforts. Projects are managed with consideration as to how the project performed in the past.

f) Involvement

An improvement process is in place to continuously improve management's involvement in the process of managing projects. Lessons learned are captured and used to improve the monitoring and control efforts. Management uses the data obtained from the efficiency and effectiveness metrics for projects to make decisions.

2.2. Search Algorithm

It has searched for the finding of similarities between questions and model documents with the pattern of designing. Moreover, it has selected the regression analyzing and regression analysis from correlation by calculating of similarities. In addition, each procedure of regression analyzing is replaced by Means of independent variable and the information of the document is replaced by Means of dependent variable for index. Similarly, the independent variable is matched with the dependent variable and it has calculated for finding the similarities value between questions and index of document.

a) Regression Analysis of Algorithm

Regression relationship analysis is the relationship analysis between 2 variables that are X and Y variables with linear relationship. Moreover, it has the regression equation that is $Y = \alpha + \beta x$. Thus, Y is the Means without being each value of Y because the simple regression analysis should have X variable to be determined the value first. Similarly, Y value will change with X variables. Then, when bringing all X and Y values to plot on the axis of X and Y and line up between the dot appearing on the graph line as the straight line, it shows of the relationship between the Means of X and Y variables. In addition, it is the regression line, so it can write the regression equations simply by finding a and b values from these at $a = \bar{Y} - b\bar{X}$ and $b = \sum xy / \sum x^2$

b) Regression Analysis from Correlation Co-efficient

It has predicted to find the dependent value from Y variable with the changing of independent value. Furthermore, it can analyze the regression analysis to predict from 2 variables of r-correlation coefficient. Then, it can make the prediction of regression analysis without depending on constant value of a and the regression coefficient. However, it has the necessity to know about the standard deviation of 2 variables in order to find the scores of prediction the dependent variable from the equations with these following:

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

2.3. Selection of the Property on the Basis of Relationship

The selection of attribution property to be considered on the basis of relationship (Correlation-based feature selection: Cfs) [9] is to find the attribute group for taking assessment for the values of the expectation ability. In addition, it is selected to classify the information type for managing with not involved attributes. Thus, it makes the effectiveness of type classification to be low and it can find the equation as this following:

$$Merit_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}$$

$Merit_s$ is Heuristic as the group of Attribute S comprising of k attribute to be selected.

r_{cf} is the Means of the group of Attribute to be selected with the relevant relationship of information type.
 \bar{r}_{ff} is the Means of the group of Attribute to be selected from relevant relationship inside the group of Attribute.

a) AMPM3 Criteria

Apply model project management maturity model (AMP_{M3}) criteria is the way to present the method of Algorithm summation technique to indicate of the relationship between information. Besides, it should measure the differences or disperse information. Then, if the information is varied much, the values to be calculated will be higher. On the other hand, if the information is in similar way, the value will be in the low way. On the other hand, with the calculating of AMP_{M3} criteria Algorithm it is the calculation to find the value with the beginning node. Additionally, it has applied as the tree roots between Information GAIN and Cellular automata. Therefore, the researcher can make the assumption that the value of AMP_{M3} criteria can give the accuracy value with the highest value of equation. Thus, it can find the relationship to reduce the information of attributes. As the result, the researcher can present the Algorithm of AMP_{M3} criteria by searching the rules of relationship and it has the equations to apply in AMP_{M3} criteria as these following:

The first equation of Information Gain is as this following:

$$\text{Entropy}(y, S) = \sum_{c_j \in \text{dom}(y)} - \frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_2 \frac{|\sigma_{y=c_j} S|}{|S|}$$

And, the second equation of Cellular automata is as this following:

$$S_i^{t+1} = f_i (s^t_{\text{neighborhood}})$$

Thus, it can setup the pattern of AMP_{M3} criteria that is ①+② as the equation with this following:

$$\text{AMP}_{M3} \text{ criteria } (S, V) = 1 \left[\sum_{c_j \in \text{dom}(y)} - \log_2 \left(\frac{|\sigma_{y=c_j} S|}{|S|} \right)^2 \right] + f_i (s^t_{\text{neighborhood}}) + q(s)y$$

However, Form equation Thus, it can be formatted in a manner of linear equations, and differential equations level second this following:

$$y'' + p(x)y' + q(x)y = r(x)$$

Or

$$\frac{d^2y}{dx^2} + p(x) \frac{dy}{dx} + q(x)y = r(x)$$

The equation is as follows: If it is the case. $r(x) = 0$. It is A homogeneous equation (Homogeneous) and it can write an equation this following:

$$\frac{d^2y}{dx^2} + p(x) \frac{dy}{dx} + q(x)y = 0$$

And the consideration can be written as ordinary differential equations level n . The equation will be available in written form below.

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

2.4. Decision Tree is the Step of Decision Tree

method as one technique to estimate the Discrete-value function. Then, it has comprised with the sets of regulations, such as if-then by creating the model of decision tree to select the attributes with the relationship of class at most. In addition, it can be the top tree (root node). After that, it can find the next attribute continuously for gaining the relationship of attribute to measurement called as Information Gain (IG). Furthermore, this calculated value has come from the equation as this following:

$$\text{entropy}(c_1) = -p(c_1) \log p(c_1) - p(\bar{c}_1) \log p(\bar{c}_1)$$

It is the possibility of c_1 and then we will try to calculate for each attribute comparing with class to find the attribute with most IG value following by root of decision tree.

2.5. Rule-Based Is the Decision Tree to Be Created in Large Size

Then, it results in the complexity of interpreting. Additionally, with this cause the extraction of the decision tree rule is designed to create in the form of "IF-THEN" to make understand easily. Besides, it should create one rule from each branch of tree beginning from the root node to the leave node. Similarly, each attribute in the branch is in the part of "IF" and leave node is in the part of "THEN", respectively.

2.6. Naïve Bayesian Is the Method to Solve the Problems

with classification for making prediction of results and explain the Naïve Bayesian technique to gain the good and rapid result with simplicity and rapidness by using the equation as this following:

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

2.7. KNN, K-Nearest Neighbor

is the method to find the similarities by using the theory to measure distance. Then, it should consider from the least distance of 1-2 suitable numbers. After that, it can bring to solve the problem of Non-Parametric Estimation with classification. As the result, this dispersing information is setup to consider that which point is closed with any point at most by using the equation as this following:

$$\text{dist} = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

3. Methodology

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:

3.1. Pre-processing

According to the information from questionnaires of the basic source, 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 825 records as the main function for applying the software quality management. Significantly, it has 6 attributes with the required groups

and 15 attributes of the software quality for general behaviors.

3.2. 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 $P\{-Z_{\sigma/2} \leq Z \leq Z_{\sigma/2}\} = 1 - \alpha$ 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:

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

3.3. Preparation of Information Set to Practicing and Testing

It is 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.

3.4. Processing of Information

The processing of information for both practicing sets and tests have applied WE-KA program and MATLAB program to make comparison of the accuracy and the correctness in making prediction of the software quality management.

3.5. 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 questionnaire sources, 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:

It can bring 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 management. After that, it can bring the information into the procedure to find the relationship rule of regression. Then, it can classify this information as the result dividing into the minor parts as the main function to apply the software quality management. Additionally, it includes

with the required groups and behaviors of applying the software quality management. 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.

The procedure to reduce attributes can be applied with the method of algorithm summation as AMPM3 Criteria. 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 3.5.2 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.

The measurement procedure and the comparing of effectiveness between the regression Algorithm relationship rule and AMPM3 Criteria is the way to reduce the dimension of information. Moreover, this procedure is as the result of the heading 3.5.1 and the heading of 3.5.2.

4. Result

According to the result, this research can make prediction of the software quality management by processing with the assumption. Besides, it has applied WE-KA program and MATLAB program which the result to be analyzed and compared the result on this research can divide into 2 parts as this following:

Software quality Management

Comparison of Technique and Main Factor to Reduce Dimensions of Information

Part A: Software quality Management

The assessment for testing the search Algorithm and AMPM3 Criteria. Then, the researcher will present the Algorithm Summation to take assessment and discuss with the result as this following:

According to the testing of relationship for information by testing the search algorithm, it can indicate with the direction to develop the quality software management. Moreover, it has the information to emphasize on 21 attributes

Part B: Comparison of Technique with the Main Factor to Reduce Attributes

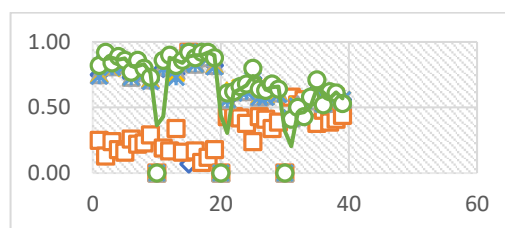


Fig. 1. Information of quality planning.

According to the explanation of Fig. 1, the comparison of the data mining for quality planning in applying among the techniques of Decision Tree, Rule- Based, Naïve Bayesian and KNN. Besides, it was found that the accuracy, the validity, the reliability, 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 92 percent. Besides, according to the selection of information totally 2 in 3, it affected to the learning information without equality for each class. Thus, if in each class it has the practicing information numbers more than another class, it reflects to that class to have the best result of test.

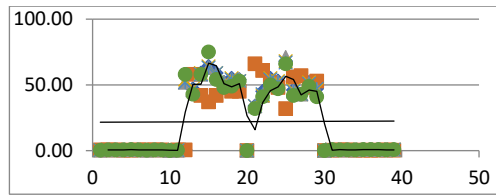


Fig. 2. Information quality assurance.

According to the explanation of Fig. 2, the comparison of the data mining for Information quality assurance in applying among the techniques of Decision Tree, Rule-Based, Naïve Bayesian and KNN. Besides, it was found that the accuracy, the validity, 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 75 percent. Besides, according to the selection of information totally 1 in 3, it affected to the learning information without equality for each class. Thus, if in each class it has the practicing information numbers more than another class, it reflects to that class to have the best result of test.

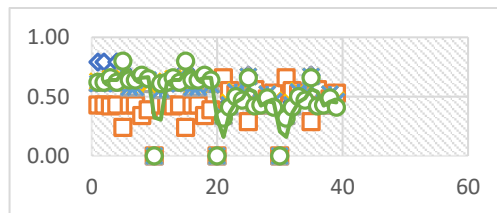


Fig. 3. Information of quality control.

According to the explanation of Fig. 3, the comparison of the data mining for information quality control in applying among the techniques of Decision Tree, Rule-Based, Naïve Bayesian and KNN. Besides, it was found that the accuracy, the validity, the reliability, 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 1 in 3, it affected to the learning information without equality for each class. Thus, if in each class it has the practicing information numbers more than another class, it reflects to that class to have the best result of test.

5. Research

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%

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.

6. Conclusion

According to the comparison of technique Decision Tree, Rule-Based, Naïve Bayesian and KNN, it has applied WE-KA and MATLAB programs. Moreover, it was found that the analysis of Algorithm summation for AMPM₃ criteria is used for the data mining classification of the software quality management. Then, according to the analyzing of each technique to gain much accuracy result with the correctness of the software quality management, it should have the Means of 63percent, 45percent and 85 percent, respectively. However, according to the AMPM₃ criteria analyzing to measure the differences or the dispersing of information, it was found that the information of Search Algorithm and AMPM₃ criteria indicates about the characteristic of information as the pattern of relevant relationship to reduce the attributes. Thus, the reducing of attributes can’t affect to the software quality management. Thus, as The proving of the ordinary homogenous equation in n rank from the consideration of the answer in the second rank can take the action 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 of the linear homogenous equation. Lastly, the further research is the model to measure the maturity for administering the software project for the executives and to present the measurements for the further software quality management.

References

- [1] Crawford, J. K. (2007). Project management maturity model second edition. Auerbach Publications.
- [2] Harold, K. (2005). Using the project management maturity model strategic planning for project management. Wiley Publications.
- [3] Harold, K. (2003) Advanced project management: Best practices on implementation. 2nd Edition. Wiley Publications.
- [4] Daniel, A. C., & Ricardo, M. B. (2012). Project management and software development processes: integrating RUP and PMBOK. *Proceedings of the 2012 International Conference on Systems Engineering and Modeling*.
- [5] Liujian, C. L., Huayou, S, J., & Yongjian, R. (2015). Research on prediction method for pivotal indicator of hospital medical quality using decision tree. *Proceedings of the 6th IEEE International Conference on Software Engineering and Service Science*.
- [6] Tan, L. M, Phang, W. S., & Chin, K. (2015). Rule-based sentiment analysis for financial news. *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*.
- [7] Supakpong, J., & Nuanwan, S. (2014). Hierarchical text classification: A restructured approach, *Transactions on Engineering, Computing and Technology*.
- [8] Nittaya, K. (2008). Comparative study of techniques to handle missing values in the classification task of data mining. *Proceedings of the 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*.
- [9] Azuraliza, A., Zuriyah, K., Salwani, A., & Mazrura, S. (2011). Predictive models for dengue outbreak using multiple rulebase classifiers. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*.
- [10] Aslandogan, Y. A., & Gauri, A. M. (2004). Evidence combination in medical data mining. *Proceedings of the International Conference on Information Technology: Coding and Computing*.
- [11] Fathima, K. B., & Nazreen, M. B. (2015). Feature subset selection based on filter technique. *Proceedings of the 2015 International Conference on Computing and Communications Technologies*.
- [12] Irin, B, Dennis, B, Anirudh, K., & Joydeb, R. (2014). Tactile sensing based softness classification using machine learning. *2014 IEEE International Advance Computing Conference (IACC)*.
- [13] Wang, X., Jiang, Z., & Yu, D. (2015). An improved KNN algorithm based on kernel methods and attribute reduction. *Proceedings of the 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*.
- [14] Shweta, T, Charu, G, Sakshi, A., & Veni, J. (2015). MFZ-KNN-A modified fuzzy based K nearest neighbor algorithm. *Proceedings of the 2015 International Conference on Cognitive Computing and Information Processing (CCIP)*.
- [15] Simone A. L. (2015). Analyzing gene expression data: Fuzzy decision tree algorithm applied to the classification of cancer data. *Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.
- [16] Nikhil, R. P., Shrish, V., & Om, P. V. (2017).. Fuzzy rule-based approach for software fault prediction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.
- [17] Dacha, S, Kumron, S., & Siraput, C. (2008). Naïve Bayes random forest: high classification efficiency model. *Proceedings of the 5th International Joint Conference on Computer Science and Software Engineering*.
- [18] Tunpana, S. (2008) Prediction stock trading signals using locality-based genetic algorithm. *Computer Engineering and Intelligent Systems*.

- [19] Ukkanat, N., & Banjaporn, L. (2008). Automated web service composition using rule-based engine. *Proceedings of the 5th International Joint Conference on Computer Science and Software Engineering*.
- [20] Wilaiporn, I., & Krit, S. (2008) Plasmodium vivax image classification by texture features and k-nearest neighbor. *Proceedings of the 5th International Joint Conference on Computer Science and Software Engineering*.
- [21] Siriporn, P., Kanongnit, K., & Nagorn, I. (2008). Finding optimal inventory using rule-based forecasting technique. *Proceedings of the 5th International Joint Conference on Computer Science and Software Engineering*.



Kattiya T. Yangyuen is an assistant professor. Dr. Kattiya T. Yangyuen is the head of Department in information technology of Rattana Bundit University, Thailand. Her primary research interests are in the areas of system analysis and design, software engineering, and database system.



Surasak Mungsing is an assistant professor. Dr. Surasak Mungsing is the director of Ph.D. (information technology) program in the School of Information Technology, Sripatum University, Thailand. His primary research interests are in the areas of software engineering, system analysis and design, database system, and computer networks.



Tipaporn Supamid works as a lecturer at Rattana Bundit University and a Ph.D. candidate (information technology program) in the School of Information Technology, Sripatum University, Thailand. Her research interest is in the area of software engineering.