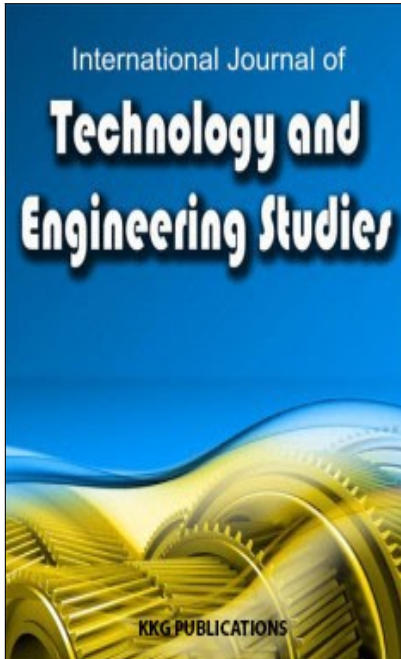
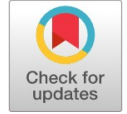


This article was downloaded by:
Publisher: KKG Publications



Key Knowledge Generation

Publication details, including instructions for author and subscription information:

<http://kkgpublications.com/technology/>

The Forecast Career Path of the New Graduates of Computer Sciences Program at Suan Sunandha Rajabhat University by Decision Tree Technique

THITIPORN SITTICHAILAPA ¹, WATCHARA PANTASAK ², RATTANATHIP RATTANACHAI ³

^{1,2,3} Suan Sunandha Rajabhat University, Bangkok, Thailand

Published online: 10 August 2016

To cite this article: T. Sittichailapa., W. Pantasak and R. Rattanachai, “The forecast career path of the new graduates of computer sciences program at Suan Sunandha Rajabhat University by decision tree technique,” *International Journal of Technology and Engineering Studies*, vol. 2, no. 4, pp. 110-116, 2016.

DOI: <https://dx.doi.org/10.20469/ijtes.2.40003-4>

To link to this article: <http://kkgpublications.com/wp-content/uploads/2016/2/Volume2/IJTES-40003-4.pdf>

PLEASE SCROLL DOWN FOR ARTICLE

KKG Publications makes every effort to ascertain the precision of all the information (the “Content”) contained in the publications on our platform. However, KKG Publications, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the content. All opinions and views stated in this publication are not endorsed by KKG Publications. These are purely the opinions and views of authors. The accuracy of the content should not be relied upon and primary sources of information should be considered for any verification. KKG Publications shall not be liable for any costs, expenses, proceedings, loss, actions, demands, damages, expenses and other liabilities directly or indirectly caused in connection with given content.

This article may be utilized for research, edifying, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly verboten.

THE FORECAST CAREER PATH OF THE NEW GRADUATES OF COMPUTER SCIENCES PROGRAM AT SUAN SUNANDHA RAJABHAT UNIVERSITY BY DECISION TREE TECHNIQUE

THITIPORN SITTICHAILAPA^{1*}, WATCHARA PANTASAK², RATTANATHIP RATTANACHAI³

^{1, 2, 3} Suan Sunandha Rajabhat University, Bangkok, Thailand

Keywords:

Career Path
Decision Tree Technique
Data Mining
Computer Science

Received: 11 June 2016

Accepted: 16 July 2016

Published: 10 August 2016

Abstract. The objective of this research was to develop the career path model of the graduates of the Computer Sciences Program of Science and Technology Faculty at Suan Sunandha Rajabhat University. This model is gathering results of academic curriculum and the survey employed applications. The satisfaction criteria of grade were 2.75, 2.75 up, and less than 2.75. By decision tree technique, we developed the subject structure to affect the career paths of graduates who followed the computer sciences program in 2006. After that, the model projected a career path to the new graduates and checked their information. This study found an important problem in data of a new graduate; it has many records and no information about the first occupation; this is important for this research because it develops confidence. However, a model from completely new data is built. Finally, we checked the reliability of the model discovered to be 70 percent correct and should have developments in the prototype. It aims that our students have to be ready to goal on the future and stick to a strong curriculum for the new technology. However, it should not be ignored that 30 percent has missing value (for example, age, salary, expense, the progress in corporate, etc.). In the future, the researcher is finding out the other factors' effect.

INTRODUCTION

The researcher is an instructor in the Computer Sciences Program of Science and Technology Faculty at Suan Sunandha Rajabhat University. We are interested in historical data, may be to predict something for us. Data Mining is the technique for research, algorithm example: association, classification, artificial neural network, etc. It uses historical data to research knowledge, a pattern of hidden data, etc. And we have a lot of data, it makes sure to have high confidence and trust in output.

The career path of the Computer Sciences Program in the market place includes a Programmer, Database Administrator, System Analyst, Graphic Designer, Web Programmer, Support System, System Administrator, etc. The first occupation is almost a programmer but somebody might not like coding. So the objective of this paper is finding patterns of the first occupation by grade in each subject in the Computer Sciences Program of a new graduate in the Computer Sciences Program of Science and Technology Faculty at Suan Sunandha Rajabhat University. And then we can apply it, to predict the sophomore year or more levels before it is finished. If they know their career path then they are planning the future in remaining years before finished, we hope that. The result of this case is expected to be a prototype model and this concept can apply to another program and another place. Of course, we may improve

outcome to the computer sciences program for quality assurance program.

The output of this conceptual model just predicts that the students are excellent for which path. We have imported history data from 2006th to 2015th, they have a transcript and occupation survey record of the new graduates. We have gathered data in 2 sections; designed data structure in ARFF format and transformed to run by WEKA application by decision tree algorithm. We expect rules base about the career path from result of subjects in Computer Sciences Program and bring the rules base to update the Computer Sciences Program, it is a part of quality assurance of academic program. And guide the students about how to be expert for each future path before finished. The best way should be recommended to them at sophomore level.

LITERATURE REVIEW

The Development of Information Systems for the traditional Occupation of the Local Community [6]. Just developed Web Application for the traditional occupation of the local community. It sparked idea to find out the future occupation of new graduates in computer sciences program at Suan Sunandha Rajabhat University. They developed the web base type for

*Corresponding author: Thitiporn Sittichailapa

†Email: thitiporn.si@ssru.ac.th, koi.thitiporn@gmail.com

dynamic concept. [1] analyzed decision tree technique by cross-fold (5, 6, 7) for finding out accuracy based on cross fold ratio of a classifier by ID3 algorithm. The title name is an Analysis of Data Mining Classification with Decision tree Technique.

Stepping Ahead with Rolling Forecast

Can Bangladesh look Beyond?. [5] tried to implement rolling forecasting in the budget operations of the companies. It is in business field, so it has factors that cannot be controlled. The implementation is divided in two phases, model development phase and model implementation phase like the process of data mining.

Book recommendation system for digital library based on user profiles by using association rule [2] and Efficient adaptive-support association rule mining for recommender systems [3]. Two papers find out data of association by data mining technique (association rule algorithm).

METHOD AND MATERIALS

[4] suggests that:

“How can I analyze these data?” - Data Mining Concept is a knowledge mining from data.

The knowledge discovery process is an iterative sequence of the following steps:

- Data cleaning (to remove noise and inconsistent data)
- Data integration (where multiple data sources may be

combined)

- Data selection (where data relevant to the analysis task are retrieved from the database)
- Data transformation (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)
- Data mining (an essential process where intelligent methods are applied to extract data patterns)
- Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interestingness measures)
- Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users)

Classification is one of the techniques of data mining. It is a form of data analysis that extracts models describing important data classes [8-11]. Such models, called classifiers, predict categorical (discrete, unordered) class labels [7].

Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node. A typical decision tree is shown in Figure 1.

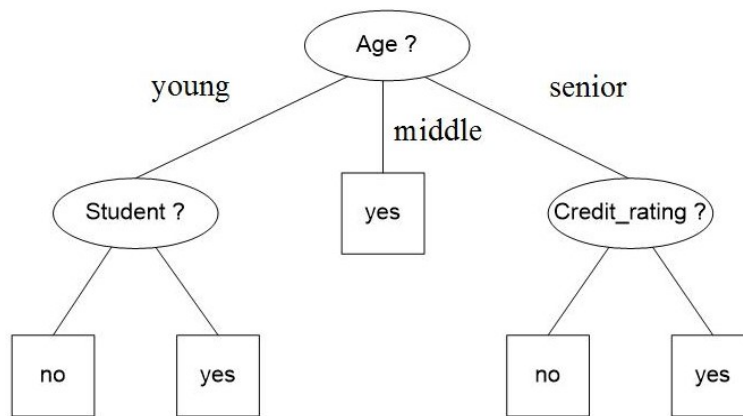


Fig. 1 . A decision tree for the concept buys computer, each internal (non-leaf) node and each leaf node [4]

Algorithm: Generate_decision_tree

Generate a decision tree from the training tuples of data partition, D.

Input

Data partition, D, which is a set of training tuples and their associated class labels; Attribute_list, the set of candidate

attributes; Attribute_selection_method, a procedure to determine the splitting criterion that best partitions the data tuples into individual classes. This criterion consists of a splitting_attribute and, possibly, either a split-point or splitting subset.

Output

A decision tree.

Method

- Create a node N;
- if tuples in D are all of the same class, C, then
- Return N as a leaf node labeled with the class C;
- if attribute_list is empty then
- Return N as a leaf node labeled with the majority class in D;
- apply Attribute_selection_method (D, attribute_list) to find the “best” splitting_criterion;
- label node N with splitting_criterion;
- if splitting_attribute is discrete-valued and multiway splits allowed then
 - attribute_list ← attribute_list - splitting_attribute;
 - for each outcome j of splitting_criterion
 - let D_j be the set of data tuples in D satisfying outcome j;
 - if D_j is empty then
 - attach a leaf labeled with the majority class in D to node N;
 - else attach the node returned by Generate_decision_tree (D_j , attribute_list) to node N; endfor
 - return N;

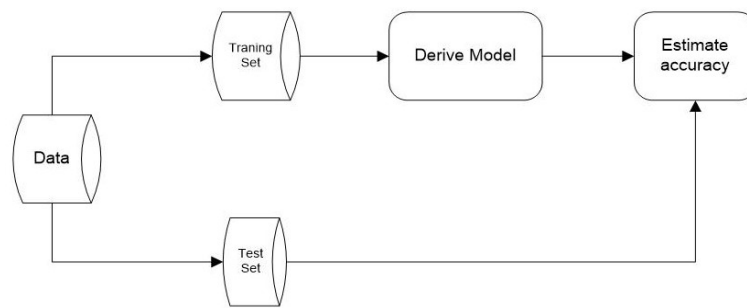


Fig. 2 . Estimating accuracy with the holdout method [4]

This figure is a process flow of data: separate 2 parts are training data set and test data set. The training data set is analyzed by a classification algorithm and test data set is used to estimate the accuracy of the classification rules. If the

accuracy is considered acceptable, the rules can be applied to the classification of new data tuples, followed by Figure 2. So formula of measure can be seen in Table 1.

TABLE 1
TABLE OF EVALUATION MEASURES [4]

Measure	Formula
Accuracy, recognition rate	$TP + TN$ $P + N$
Error rate, misclassification rate	$FP + FN$ $P + N$
Sensitivity, true positive rate, recall	TP P
Specificity, true negative rate	TN N
Precision	TP $TP + FP$
F, F_1, F -score, harmonic mean of precision and recall	$2 \times \text{precision} \times \text{recall}$ $\text{precision} + \text{recall}$
F_β , where β is a non-negative real number	$(1 + \beta^2) \times \text{precision} \times \text{recall}$ $\beta^2 \times \text{precision} + \text{recall}$

The meaning of measure term:-

- TP - True positives: These refer to the positive tuples that were correctly labeled by the classifier. Let TP be the number of true positives.
- TN - True negatives: These are the negative tuples that were correctly labeled by the classifier. Let TN be the number of true negatives.
- FP - False positives: These are the negative tuples that were incorrectly labeled as positive. Let FP be the number of false positives.
- FN - False negatives: These are the positive tuples that were mislabeled as negative. Let FN be the number of false negatives.

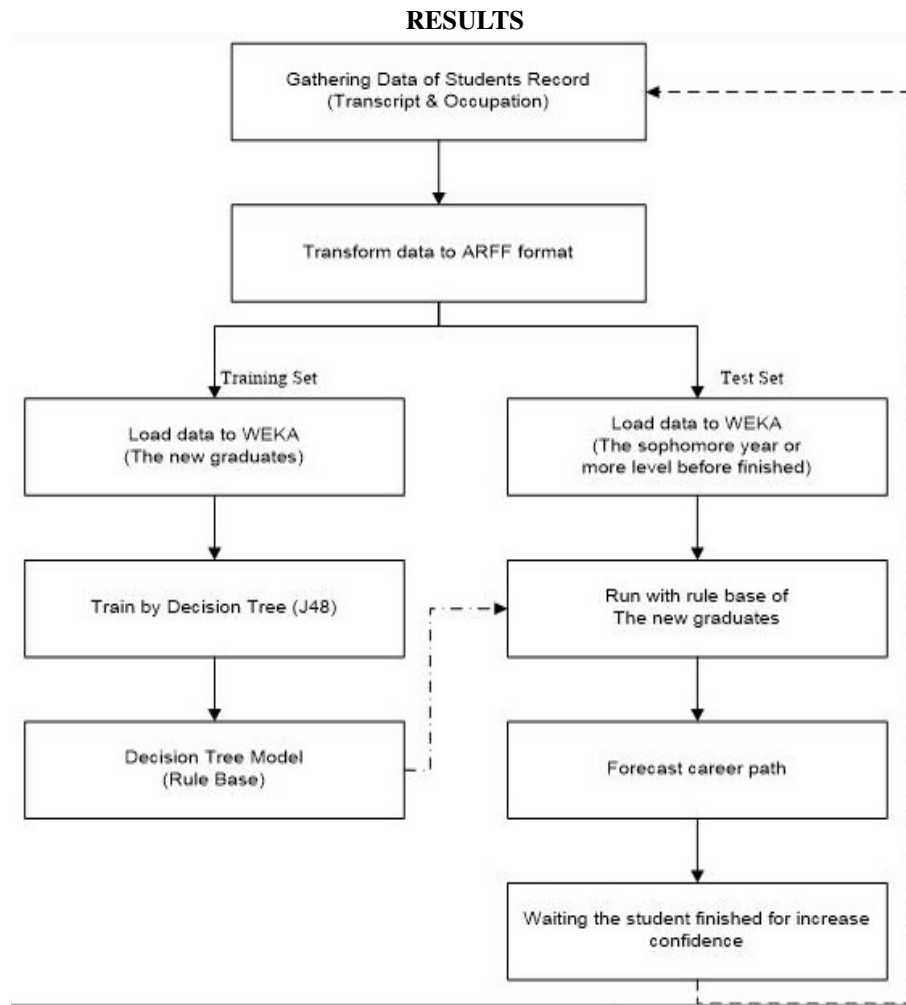


Fig. 3 . Work flow diagram for mining occupation of the new graduates of computer sciences program at Suan Sunandha Rajabhat University by decision tree technique

We prepared training data set from a graduate 2010th 2013th (student code 2006th - 2009th) and test data set got from a graduate 2014th 2015th (student code 2010th - 2011th) but a lot of data is not complete. So the model may be inaccurate in the prediction. After that they are transformed into ARFF structure followed by figure 4. The structure of training data set has 25 attributes representing grade of subjects (2 level:

1 - grade 2.75 4.0, 0 grade less than 2.75). And the label attribute is represented by occupation (“occu”) for predicting. The career path interested in this research has 7 careers: these are Programmer, System Analyst, Web Developer, Graphic Designer, Support, Software Engineer, System Administrator and plus “Other” (Out of 7 career paths) in attribute “occu”. All the process can be seen in Figure 3.

TABLE 2
TABLE OF EVALUATION MEASURES [4]

Attribute	Description	Value
COM1101	Foundation of Computer Science	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM1102	Computer Programming I	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM1109	Data Communication	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2401	Business Information System	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2103	Computer Programming II	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2104	Operating System	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2105	Data Structure	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2107	Software Architecture	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2108	Database System	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2110	Discrete Mathematics	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM2112	Object-Oriented Methodology	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3106	System Analysis and Design	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3119	Internet Technology	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3404	Management Information Systems	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3111	Software Engineering	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3113	Computer Graphics	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3114	Artificial Intelligence	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3115	Object-Oriented Programming	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3205	Human-Computer Interaction	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3209	Database Programming	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3210	Database Management System Design	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3215	Data Warehouse and Application	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3237	Object-Oriented Analysis and Design	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3212	Multidimensional Database	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
COM3216	Data Mining	1 - grade 2.75 - 4.00, 0 - grade less than 2.75
Occu	Career path of computer sciences program	ICT1 - Programmer ICT2 - System Analyst ICT3 - Web Developer ICT4 - Graphic Designer ICT5 - Support ICT6 - Database Administrator ICT7 - System Administrator OTH - Other

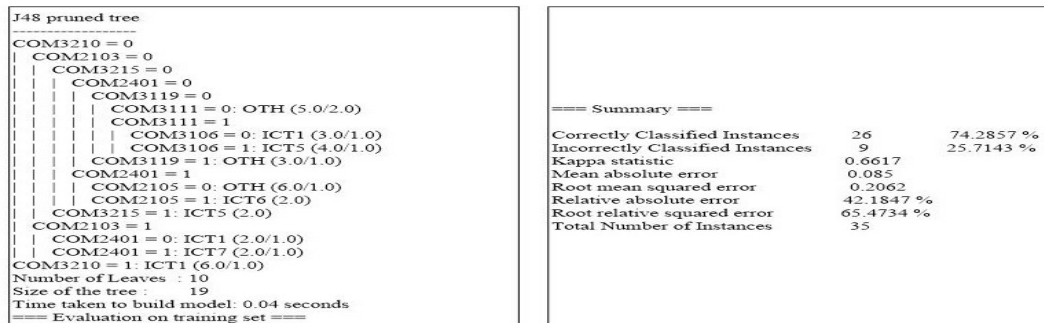


Fig. 4. The output after run training set by WEKA with classify by J48

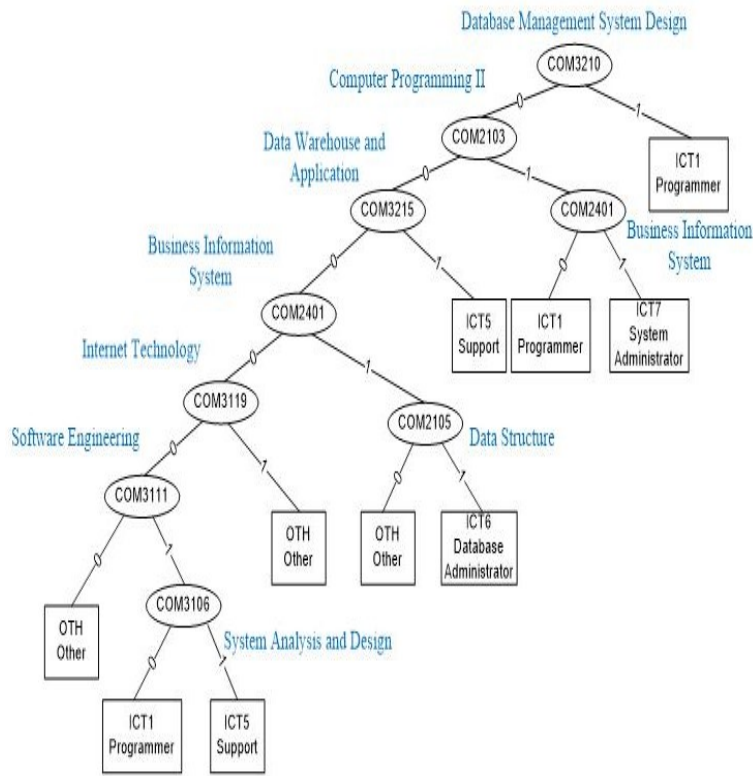


Fig. 5 . The career path model of the new graduates of computer sciences program at Suan Sunandha Rajabhat University

The results were drawn by decision tree structure, seen of final year students (student code 2012). See in Table 3. in Figure 5. After that, we proved this model by the transcripts

TABLE 3
THE TABLE OF DATA FOR EVALUATING MODEL

TRUE OCCU.	COM3210	COM2103	COM2401	COM3215	COM2401	COM2105	COM3119	COM3111	COM3106	PREDICT.
ICT1	1									ICT1
ICT1	1									ICT1
ICT1	0	0		0	1	1				ICT6
OTH	0	0		0	1	0				OTH
ICT5	0	0		0	1	1				ICT6
ICT1	1									ICT1
ICT4	0	1	1							ICT7
ICT1	1									ICT1
ICT1	1									ICT1
ICT1	1									ICT1

From the result, we found the truth in highlighted records, to be exactly 7:10 (confidence value is 0.7) in ICT1: Programmer (factors of this occupation are Database Management System Design, Computer Programming II and Business Information System). All example data tests have missing value 0.3 percent.

CONCLUSION AND RECOMMENDATIONS

This research just predicts the computer science career path by transcript of the student. And we got the training set

from the first position of the new graduates, so that is uncertain in the career path. There are many factors to affect a career path of the new graduates and growth in the career path; example: the personality, the opportunity, the activeness, etc. So this research has a disadvantage. We should be gathering the attitude test, age, gender, behavior at work etc. that add to the training set for learning and then we will get the reliability model.

However, the result of this research is brought to be developed for quality assurance in computer sciences program.

In the future, researchers plan to modify structure of

training data set for forecasting with high confidence.

Declaration of Conflicting Interests

There are no competing interests associated with this study.

Acknowledgments

The authors gratefully acknowledge the financial subsidy provided by Suan Sunandha Rajabhat University.

REFERENCES

- [1] D. Singh, N. Choudhary and J. Samota, "Analysis of data mining classification with decision tree technique", *Global Journals Inc.*, vol.13, no. 13, 2013.
- [2] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," in *Fourth International Conference on Innovative Computing Technology (INTECH)*, pp. 130-134, August, 13-15, 2014.
- [3] W. Lin and S. Alvarez, "Efficient adaptive-support association rule mining for recommender systems," *Data Mining and Knowledge Discovery Journal*, vol. 6, no. 1, pp. 83-105, 2004.
- [4] J. Han, M. Kamber and J. Pei, *Data Mining Concept and Techniques*, 3rd ed. Burlington: VT, Morgan Kaufmann, 2012.
- [5] M. Nodee, "Stepping ahead with rolling forecast: Can Bangladesh look beyond?" *Business Economics, Social Science and Humanities*, vol. 48, no. 5, pp. 1-5, 2016.
- [6] S. Janpla, K. Kularbphetong, R. Rattanachai and K. Tangtastham, "The development of information systems for the traditional occupation of the local community," in *Proceeding of Social and Behavioral Sciences*, vol. 197, pp. 2526-2529, 2015.
- [7] I. C. Yeh and C. H. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2473-2480, 2009.
- [8] G. S. Linoff and M. J. Berry, *Data mining techniques: For marketing, sales, and customer relationship management*. New Jersey, NJ: John Wiley & Sons, 2011.
- [9] O. P. Rud, *Data mining cookbook: Modeling data for marketing, risk, and customer relationship management*. New Jersey, NJ: John Wiley & Sons, 2001.
- [10] G. R. Murray, and A. Scime. "Data Mining", *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary*, pp. 1-15, 2015.
- [11] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong. "Data mining for the internet of things: literature review and challenges," *International Journal of Distributed Sensor Networks*, vol. 11, no. 8, pp.431-437, 2015.

— This article does not have any appendix. —