CLASSIFICATION OF INDONESIAN TELEMATIC SERVICES MSMEs FEASIBILITY ASSISTANCE,USINGJ 48 ALGORITHM

Eneng Tita Tosida[1], Sufiatul Maryana[1], Hermawan Thaheer[1], Okta Hairlangga[1]

[1]*Computer Science Department, Pakuan University, Jl. Pakuan PO Box 452, 16143, Bogor,Indonesia*
enengtitatosida@unpak.ac.id, sufiatul.maryana@unpak.ac.id, hermawanmohdThaheer@yahoo.com, oktahairlangga.062@gmail.com

Abstract

Feasibility assistance for Micro, Small and Medium Enterprises (SMEs) Indonesia telematics services need to be assessed for compliance. suitability parameters assistance can be seen from the characteristics of MSMEs. Telematics Services MSMEs characteristic data fully available on the National Economic Census data (Susenas). Optimization of Susenas as the basis for the feasibility study can be done through the implementation of a decision tree approach particularly algorithm J 48. The use ofJ 48 algorithms to determine the feasibility of Indonesian telematics services MSMEs capable of providing an alternative way to describe the characteristics of MSMEs are eligible to receive assistance. This is particularly useful for ministries such as the Ministry of Cooperatives and SMEs, in terms of selecting eligible SMEs are given assistance. Criteria for aid feasibility is based on the general condition of SMEs, access to information technology, economic conditions, partnerships and development plans. Evaluation of the performance of the system showed that the algorithm J 48 is able to achieve an accuracy of 64% on the training data and at 57% on the test data, resulting in a decision rule that representative. The low accuracy is caused by data that is not balanced, so it is a potential for the development of further research involving balancing data.

Keywords : Classification, Decision Tree, J 48 Algorithm, Telematic Services.

I. INTRODUCTION

Telematics become the primary concern in the development and economic growth of Indonesia today. Increasing the quantity of MSMEs telematics is a potential that must be supported to have the competitive edge, especially in the ASEAN Economic Community (AEC). But there are many difficulties in the decision making aid for feasible MSMEs. One reason is the variety of data as well as qualification standards for deciding eligibility assistance to SMEs telematics.

Previous research has been carried out by [1]. The research has developed a data visualization of Indonesian telematics services MSMEs, spatially and graphics for each region. This application was not equipped with decision support aid feasibility assistance. therefore, these systems need to be developed to support the elections in the provision to the MSMEs.

Selection of SMEs (especially telematics services MSMEs) that deserve assistance can be done through assessments characteristics. Characteristics Indonesian telematics services MSMEs can be identified through the data Susenas 2006. One of the characteristics of SMEs telematics services that can be used as a reference in the feasibility of providing assistance was whether the MSMEs receive assistance from other parties.This data can be used as a determinant for the establishment of a classification class of the other characters in an effort to determine assisstance feasibility.

Another previous research about Indonesian telematics services MSMEs supported by [2]. This study explain about clusteringmodel of empowerment of Indonesian telematics services MSMEs using the Susenas database has performed quite well, with the achievement of a minimum value of IDB which revolves around the value of 0.36.Some relief for the classification process technology based SMEs have been done through a decision tree approach. But the aid is meant only for the provision of credit and special light given to technology-based start-up [3] and [4].

Therefore this research proposed, that related to the classification of the provision of assistance to telematics services MSMEs based on Susenas 2006. This study was developed considering the potential for Indonesia telematics services MSME has a broader scope with very diverse characteristics.

The method proposed in this study is the algorithm J 48. The research results in the form of rules or the eligibility rules of telematics services help SMEs Indonesia. User interface system is built through a web format using Yii framework.

The method proposed in this study is the algorithm J 48. User interface system is built through a web format using Yii framework [5]. Yii is a framework PHP-based components, high-performance for the development of large-scale Web applications. Yii provides maximum reusability in Web programming and can improve development speed significantly. The advantages of yii is simple usage, for example in the case of a relational database or creates the form, open-source and accelerate to the creation of a system with features supporting [6].

The research objective is to develop a classification system of assistance to Indonesian telematics MSMEs using algorithms J 48. This research resulted in the rules or the eligibility rules of telematics services help SMEs Indonesia. This research resulted in the decision rules for Indonesia telematics services MSMEs aid feasibility.

II. METHOD

The methodology used is the approach to data mining or also called Knowledge Discovery and Data (KDD)

[7].As a series of processes, data mining can be divided into several stages which are presented in Fig. 1. The stages are interactive, engaging users directly or through knowledge base.

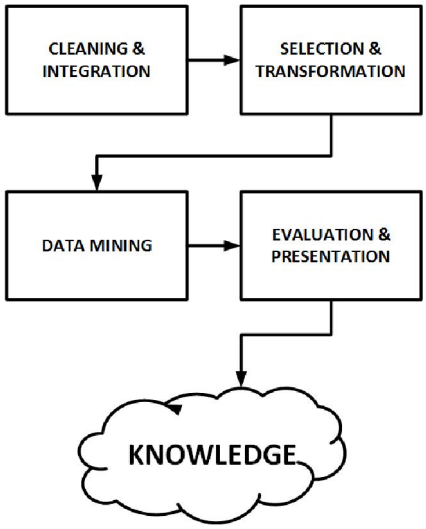

Fig.1. Data Mining Step [6]

A. Cleaning and Integration Data

Data cleaning is the process of removing noise and inconsistent data or data irrelevant. Data integration is the combination of data from different databases into the new database. For this phase of the study refers to [1].

B. Selection and Transformation Data

Existing data in the database is often not all used, therefore only the appropriate data to be analyzed to be retrieved from the database. Data Transformation is the process of changing data or merge data into a format suitable for further processing in data mining. Selection of attributes in this study using entropy-based method. This selection serves to sort out the attributes of the most influential in the classification process to be more concise and optimal [5].

C. *Mining Process*

This process is the main process of finding valuable and hidden knowledge from data. The method used for this study is the J 48 algorithm (Fig. 2).

D. Pattern Evaluation

This stage is able to identify interesting patterns into knowledge based alerts. In this stage the results of data mining techniques in the form of distinctive patterns and prediction models were evaluated to assess whether the hypothesis that there is indeed reached. If it turns out the results obtained do not fit the hypothesis there are several alternatives that can be taken such as making a feedback to.

E. *Knowledge Presentation*

This stage is the visualization and presentation of knowledge about the methods used to obtain the knowledge acquired. The last stage of the data mining process is how to formulate a decision or action of the analytical results obtained. In this presentation, the results in the form of meaningful charts, contains some rules implemented in the form of web-based systems. To improve the process of data mining, or accept these results as a result of the unexpected that may be useful. Evaluation of this study using the confusion matrix.

III. RESULT AND DISCUSSION

Results from this study is the classification system of SME support telematics services using the algorithm J 48 then the algorithm is executed using the application Rstudio. Prorporsi training data and test data is 80%: 20% of the 8798 data. Having obtained the output in the form of a decision tree, rule or rules of the decision tree is implemented using adobe dreamweaver then applied in yii framework and database designed and made in the application of MySQL. Here is the view classification system classification SME support telematics services using the algorithm J 48. Because of the decision tree produced too complex, then do generalize each attribute and select the attributes to be the most important attribute 10 through entropy-based (TABLE I).

TABLE I.
Attribute Rating Based on Entropy Value

Rating	Attribute Name	Value
1	Has reveived training	4592,23
2	Planning	3408,06
3	Adversity	2442,09
4	Partnership	1923,90
5	Business prospect	579,46
6	Legal entity	404,74
7	ɔerative membership	328,07
8	Marketing	266,78
9	Sale	245,45
10	Business group	206,17

Prunning tree is also to be cut decision tree to be more concise. But this resulted in a decreased level of accuracy of about 10% to 30%. Here is a description of the data that has been generalized and in the selection are presented in TABLE II. Class for this classsification process consist of two classes (TABLE III).

TABLE II.
DESCRIPTION OF THE DATA ATRIBUT AFTER THE SELECTION PROCESS ATTRIBUTE

No	Atribut	Value Range			
1	Legal entity	1.	CV	5.	Personal
		2.	Firma	6.	Limited Company
		3.	Special permission from the relevant agencies	7.	Founda-tion
		4.	Cooperative		
2	Business group	1.	Telecomuni-cation Services	3.	*Software Consul-tant*
		2.	*Hardware Consultant*	4.	else
3	Sale	1.	Micro	3.	Medium
		2.	Small		

TABLE II.
DESCRIPTION OF THE DATA ATRIBUT AFTER THE SELECTION PROCESS ATTRIBUTE (Continued)

No	Atribut	Value Range			
4	Adversity	1.	Material	6.	Capital
		2.	Energy	7.	No adversity
		3.	Skilled labor	8.	Transporta-tion
		4.	Else	9.	Wage of laboes
		5.	Marketing		
5	Cooperative membership	1.	Yes	2.	No
6	Partnership	1.	Yes	2.	No
7	Has reveived training	1.	Yes	2.	No
8	Marketing	1.	Export	3.	Province
		2.	District	4.	Outside the province
9	Business prospect	1.	Better	4.	Equally bad
		2.	Worse	5.	Can not be compared
		3.	Equally good		
10	Planning	1.	No	4.	Yes, improve skills
		2.	Yes, expand a site of business	5.	Yes, else
		3.	Ya, open branches		

TABLE III.
Class of Classification

Attribute	Value Range	
Has recieved assisstance	1. Yes	2. No.

A. *Mining Process*

The data mining process is done by algorithms J 48. Before performing the data mining process is divided 2 into training data and test data. Training data as much as 80% and 20% for the test data. The test data is used as a validation of the decision tree that was created with the original data pattern. If the rules are too complex then obtained another process is required in order to obtain a more concise rules that are so easy to implement and improve system performance. Fig. 2 shows the flowchart for algorithm J48. A common step algorithm J 48 in building a decision tree is as follows:

a. Selecting attribute as root

To select an attribute as a root node, based on the value of the highest Gain existing attributes. To get first need to calculate Gain Entropy. Entropy is a parameter to measure heterogenenity or (diversity) of a sample set of data. If the sample set of data is becoming more heterogeneous, the greater the entropy value [7]. Mathematically, entropy formulated in Equation (1) and Gain defined in Equation (2). Where c is the number of values in the target attribute (the number of grade classification) and Pistating the number of samples for the class i.

$$Entropy(\mathrm{S}) = \sum_{\mathrm{t}}^{\mathrm{c}} -\mathrm{Pi}\ \log_2 \mathrm{Pi} \quad (1)$$

$$Gain\ (A) = Entropy(S) - Entropy_A(S) \quad (2)$$

With the gain obtained from the reduction of the overall results of entropy with entropy attribute A.

b. Create a branch for each value

Once the entropy values obtained from a sample set of data, it can measure the effectiveness of an attribute in classifying data called gain ratio. Gain ratio is calculated based on the Split Information defined in Equation (3).

$$Split\ Information\ (\mathrm{S}, \mathrm{A}) = \sum_{\mathrm{i}=1}^{\mathrm{c}} -\left|\frac{\mathrm{S_i}}{\mathrm{S}}\right| \log_2 \left|\frac{\mathrm{S_i}}{\mathrm{S}}\right| \quad (3)$$

By S denote the set of data samples and Si to Sc declared sub sample set of data which is divided by the number of variations in the value of the attribute A.

c. Dividing the cases in the branch

Dividing the cases in the branch is performed by calculating Gain Ratio formulated in accordance with Equation (4).

$$Gain\ Rasio(\mathrm{S}, \mathrm{A}) = \frac{\mathrm{Entropy\ (S)} - \sum_{\mathrm{vcValue(A)}} \left|\frac{\mathrm{S_v}}{\mathrm{S}}\right| \mathrm{Entropy\ (S_v)}}{\mathrm{Split\ Information}} \quad (4)$$

With V expresses a possible values for attribute A, Values (A) is the set of values that is likely to attribute A. Sv is the number of samples for the value of v, and S is the total number of data samples. Entropy (Sv) is the entropy for samples that have value v.

d. Repeat the process for each branch until all cases the branches have the same class.

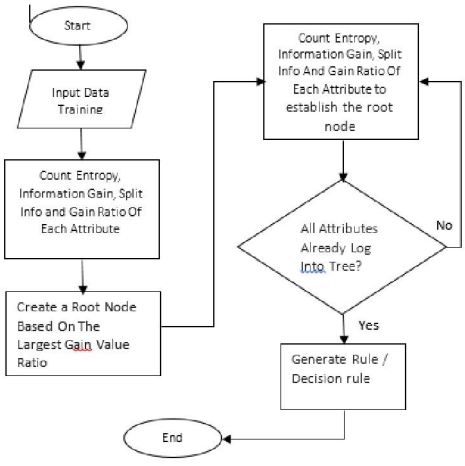

Fig.2. Flowchart of J 48 Algorithm [7].

B. *Pattern Evaluation*

Pattern evaluation for this system using the confusion matrix. Experimental results of this study using three different scenarios in order to know the level of accuracy and the number of rules that are formed. If the rules are too complex or large created the necessary process in order to rule created prunning more concise and make the system lighter and easier to implement. Here is presented in Fig. 3, 4 and 5.

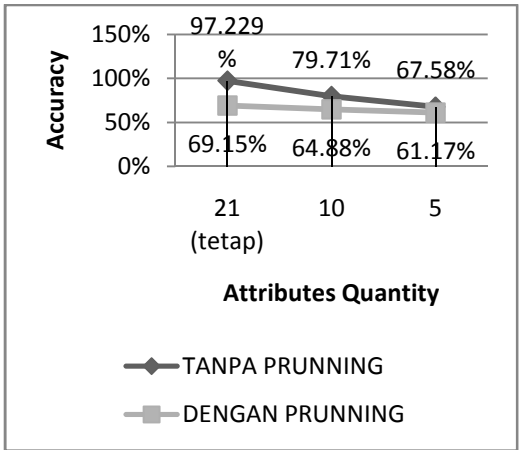

Fig.3. Accuration model of training data

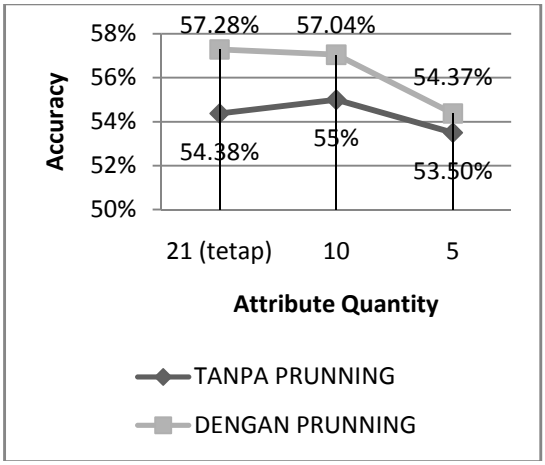

Fig.4. Accuration model of testing data

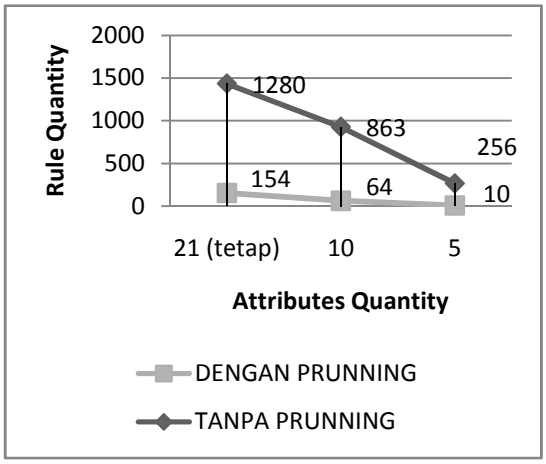

Fig.5. The number of rules generated

Fig. 3 showed the comparison of the accuracy of training data from all three scenarios are created. The highest accuracy value obtained by the scenario 21 attributes without using processes prunning tree that is 97% and also the smallest of accuracy that scenario 5 attributes with prunning process.Fig. 4 showed that comparison of the accuracy of the testing data created three scenarios. The highest accuracy value obtained by the scenario 21 attributes using prunning process tree that is 57% and also the smallest accuracy is 5 attributes scenario without prunning process.Fig. 5 shows the number of rules generated from the 3 different scenarios. In scenario 21 attributes without prunning produce as much as 1280 that rules for its implementation is more complex and the least is the scenario prunning 5 attributes with which produced only 10 rules.

Based on Fig. 3, 4 and 5 it can be concluded that the level of accuracy of training data is greater than test data, due to variations that exist in the training data more and in this research were chosen scenario of 10 attributes to the process prunning because its implementation is not too complex, but the rule in accordance with SMEs Kemenkop following rules presented more details on the presentation of knowledge.

C. *Knowledge Representation*

The resulting decision tree visualization of the most optimal scenario is shown in Figure 6. The rule for the classification of Indonesian telematic services MSMEs assisstance based on the main attributes is that training.Fig. 6 shows that the aid will be granted to SMEs have had training. These conditions appropriate delivery mechanisms in The Ministry of Cooperative and SMEs, stating that assistance will be granted to SMEs who have done the training organized by the institution [8].

SMEs who have not attended the training will be reprocessed by the partnership. If SMEs are to have a partner it will be given support, if not then processed again by the SME business plan and so on. Decision tree is then implemented into the web-based system [9]. The following is the interface that has been created.

***1) Display Index* / Home**

On this page contain the initial menu in this classification system where we can begin the classification of Indonesian telematics services MSMEs assisstance, which is shown in Fig. 7.Fig 6 show that the aid will be granted to SMEs have had training. These conditions appropriate delivery mechanisms in The Ministry of Cooperative and SMEs, stating that assistance will be granted to SMEs who have done the training organized by the institution [8].This finding is in line with the results of research conducted by [9]. One of the crucial factors for the assessment of aid SMEs are the basic conditions related to the ability and ownership of telematics infrastructure.

SMEs who have not attended the training will be reprocessed by the partnership. If SMEs are to have a partner it will be given support, if not then processed again by the SME business plan and so on. Decision tree is then implemented into the web-based system [10]. The following is the interface that has been created :

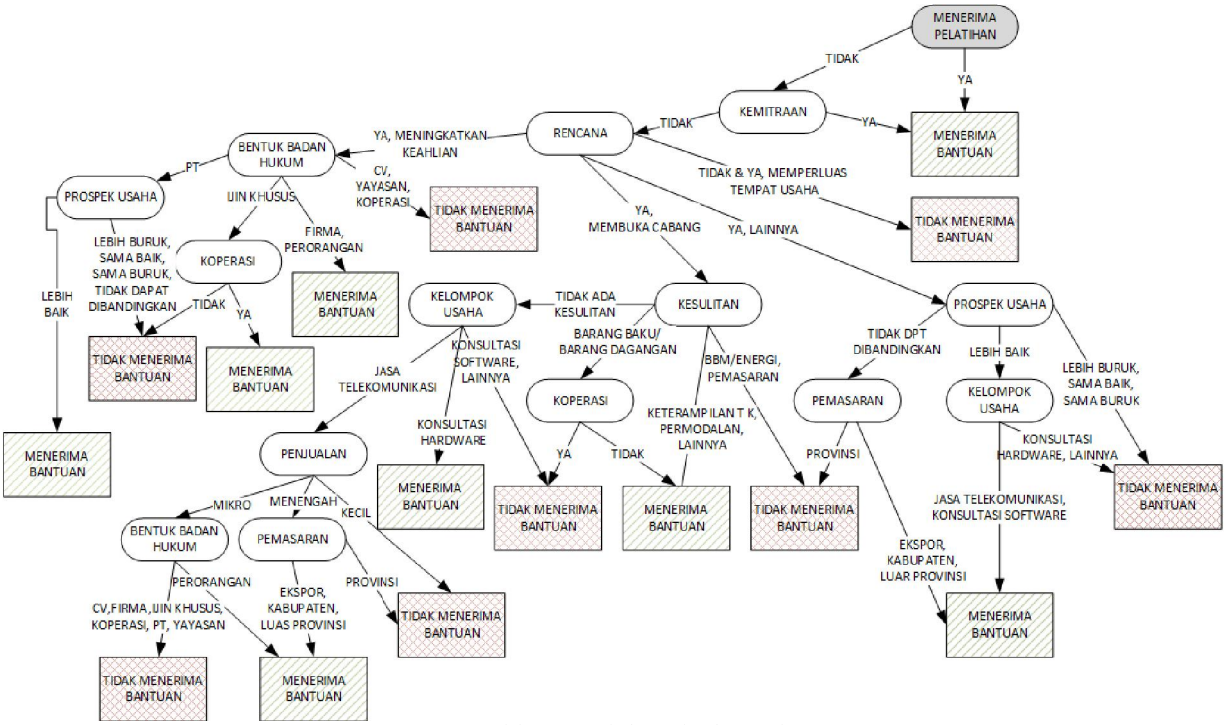

Fig.6. Decision Tree of the optimal scenario

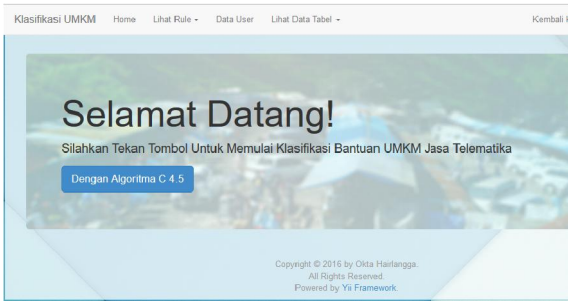

Fig. 7. Display Index

2) Display of Login form

This display serves as an authentication user who has the right to delete or edit the existing data on the web. Here is presented in Fig. 8.

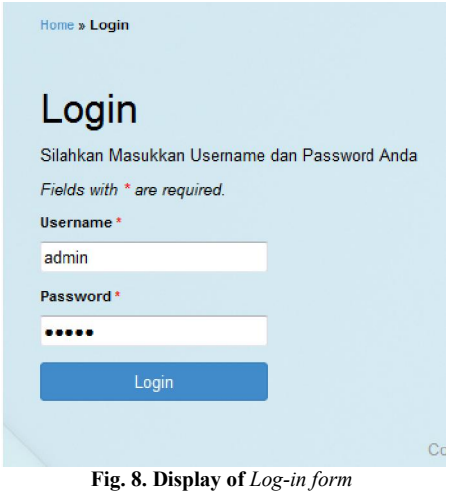

Fig. 8. Display of *Log-in form*

a. *Classification Form*

In this menu contains a form of data that SMEs will be classified. The output is feasibility of SMEs in receiving assistance. Form input is shown in Figure 9 and the data view classification results shown in Figure 10.

Fig. 9. Display of Classification *Form*

Fig. 10. Display of Classification result

b. Display of Classification Data

This menu contains data that has been entered into a classification system which has been in previous input. For admin can directly edit or delete the data. The display shown by Fig. 11.

Fig. 11. Display of Classification Data

IV. CONCLUSION

Implementation of J 48 algorithms for classification of Indonesian telematics services MSMEs assisstance has succeeded in engineering using yii framework. Yii has an advantage in the design process is relatively quick and easy. Bootstrap Template design process uses to display web becomes responsive.

The total amount of data used is the data 8798, consists of two classes. The proportion of training data and test data is 80%: 20%. Validation trials conducted by the confusion matrix and the accuracy of the results of this classification system reached 64% for the training data and the test data of 57. The accuracy obtained from the optimal scenario. The scenario in question is preceded attribute selection process up to 10 attributes and prunning decision tree. These systems have utility as an eligibility classification of SMEs receive help using a rule or rules that have been created and store data and their classification decision.

ACKNOWLEDGMENT

1. DP2M DIKTI, Ministry of Research, Technology and Higher Education of the Republic of Indonesia as the main sponsor, which gives us Competitive Grants Scheme
2. Computer Science Department, Mathematics and Natural Science Faculty, Pakuan University, and Research Institute Pakuan University, for supporting, coordinating and facilitating to achieve this grants.
3. Indonesian Communication & Information Ministry, Indonesian Cooperation and SMEs Ministry and Bandung Technopark for active participation in the activities of interviews anduser requirement.

REFERENCES

[1] E.T. Tosida, S Maryana, H Thaheer, F.A. Darmin, "Visualization model of small and medium enterprises (SMEs) telematics services potentiality map in Indonesia", inProc. of 2015 International Conference on Information & Communication Technology and Systems (ICTS),*Added to IEEExplorer January 14th 2016*,doi : 10.1109/ICTS.2015.7379890, 2015.

[2] E.T. Tosida, S. Maryana, H. Thaheer, Hardiani, "Implementation of Self Organizing Map (SOM) as decision support : Indonesian telematics services MSMEs empowerment", IOP Conf. Series : Material Science and Engineering 166 012017. doi:10.1088/1757-899X/166/1/012017, IORA-ICOR, August 27th 2017, IORA IOP Publishing, 2017.

[3] H.S.Kim, S.Y.Sohn,"Support vector machines for default prediction of smes based on technology credit",*European Journal of Operational Research* 201 (3) : 838-846. Elsevier. doi:10.1016/j.ejor.2009.03.036, 2010.

[4] S.Y.Sohn, J.W.Kim,"Decision tree-based technology credit scoring for strart-up firms : Korean case",*J. Expert Systems with Applications* 39 (4) : 4007-40112, Elsevier, doi:10.1016/j.eswa. 2011.09.075, 2012.

[5] Y.Y. Wibisono, "Performance Comparation of Decision Tree Algorithm, C5.0, CART, Dan CHAID: Case Status Credit Risk Prediction In Bank X", Thesis,Industry Engineering Department, Katolik Parahyangan University,2007.

[6] J.Winesett, Web Application Development with Yii and PHP, Packt Publishing, Birmingham, Inggris, 2012.

[7] J. Han, M.Kamber, J. Pei, *Data mining : Concepts and techniques.* Third Edition. Morgan Kaufmann is an imprint of Elsevier, 225Wyman Street,Waltham, MA 02451, USA, 2012.

[8]M. Sourbati,"The digital switchover as an information society initiative: The role of public policy in promoting access to digital ICTs", J.Telematics and Informatics,Vol 28, Issue 4, pp 295–30. Nov 2011.

[9] E.T. Tosida, K.B. Seminar, Y. Herdiyeni, "Attribut selection of Indonesian Telematic Services MSMEs assisstance feasibility, using AHP", J. Kursor, Vol 8, No 2, pp 97-106, doi : http://dx.doi.org/10.21107/kursor.v8i2.1299, Dec. 2015.

[10]M.J. Zaki, W.Meira Jr.,*Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, United Kingdom,2013.