LEARNING VECTOR QUANTIZATION IMPLEMENTATION TO PREDICT THE PROVISION OF ASSISTANCE FOR INDONESIAN TELEMATICS SERVICES SMEs

[1a]Eneng Tita Tosida, [1b]Fajar Delli Wihartiko, [1c]Indra Lumesa

Computer Science Department, Pakuan University, Jl. Pakuan PO Box 452, Bogor, Indonesia
[a]enengtitatosida@unpak.ac.id, [b]fajardelli@unpak.ac.id, [c]indralumesa@gmail.com

Abstract : Implementation of Learning Vector Quantization(LVQ) Algorithm for classification of Indonesia telematics service is designed and created as a classification system to support the decision of grant aid for Small Medium Enterprises (SMEs). Based on the test results, the LVQ algorithm has the best accuracy (93.11%) when compared with ID3 algorithm (64%) and C45 (62%) for telematics data of National Census of Economic (*Susenas* 2006). The data is still valid and relevant for use in this research because in Indonesia census data is done every 10 years and there is no update of data until now.LVQ implementation results are applied to a web-based decision support system to predict the provision of assistance for Indonesian telematics services SMEs. Unlike the C45 and ID3 algorithms, the LVQ algorithm generates the weight of a neural network where it difficult to know which attributes are most influential for decision making. But in this study LVQ able to show good performance through the analysis of the relevance of existing conditions by comparing it with the weight value produced by the model that are implemented in a web-based decision support system

Keyword : Decision Support System,Learning Vector Quantization, Small Medium Enterprises

INTRODUCTION

Indonesia has a tremendous opportunity to win the competition in the Asian Economic Community (AEC). This is supported by the number of Small and Medium Enterprises (SMEs) in the field of telematics that can survive even in the global economic crisis. SME data on telematics is available at the Central Bureau of Statistics (BPS), in the format of the National Economic Census (*susenas*2006). But this data has not been used optimally by the government to support the decision to empower SMEs. In fact, *susenas* data related to SME telematics has complete variables and has conformity with the process of providing assistance conducted by the Ministry of Cooperatives and SMEs Indonesia (Tosida et al. 2015). In addition, the 2006 *susenas* data can be used in this research because the data is still in accordance with the current condition and there is no update data from BPS until 2017.The telematics industry (Information and Communication Technology - ICT) itself is one of the priority industries that will and will be developed by the Government through the National Industrial Development Policy. The telematics industry itself is currently a rapidly growing industry in the world with 6.9% growth per year. In 2004 the world's ICT market reached US $ 533 billion, while Asian ICT market was US $ 42 billion with 23% growth per year. In Indonesia alone, the market was recorded at only US $ 1.3 billion with growth in 2004 and 2005 of 9.8% and 22.1%, respectively. Of that amount, it is estimated that US $ 0.5 billion to US $ 0.75 billion is absorbed by the banking sector. The telematics industry consists of a group of goods and services, including the Device Industry, Infrastructure / Networks (access, nodes, transport & support) and software including applications (content). For developing countries software and services generally have greater opportunities because they do not require large investments in research and production support equipment. This is mainly due to more software based on knowledgeable workforce (Tosida et al. 2016).

Based on previous research by Tosida et al (2015) the system has not built a system of determining the provision of telematics services assistance limited to SMEs data visualization of each region. On the feasibility of assistance for Indonesian telematics services Micro Small Medium Enterprises (MSMEs) involve complex criteria consisting of 21 criteria (Tosida et al. 2016). The relationship between the criteria for the feasibility of the aid is non-linear, Indonesian telematics services SMEs therefore it can be approximated by artificial neural network method. The assistance scheme for telematics SMEs has a character that is almost the same as the credit scheme for others SMEs in bank. Therefore research that can be used as a reference among others related to credit scoring mechanism for business owners including SMEs(Ince, Aktan 2009; Sadratasoul et al. 2013).

Research on feasibility of assistance for SMEs through credit risk in Supply Chain Finance (SCF) has been done by Zhu et al. (2016), which predict SMEs credit by six machine learning methods. There are one individual machine learning (i.e. decision tree), three ensemble machine learning methods (i.e. bagging, boosting and multiboosting) and two integrated ensemble machine learning machines methods (i.e. RS-boosting and multiboosting). Blanco et al. (2013) applies credit scoring using non parametric method for Peruvian microfinance industry. The model used is neural network multi layer perceptron (MLP) which compared its performance with linear discriminant analysis (LDA) approach, quadratic discriminant analysis (QDA) and logistic regression (LR).MLP performance shows a higher degree of accuracy than the other three approaches. Therefore in this research will be done visualization development using artificial neural network of Learning Vector Quantization (LVQ) algorithm to make the rule of eligibility rule of Indonesian telematics services SMEs and compare result with C45 algorithm and ID3 algorithm. Actually, The selected classification techniques J48, CART and ID3 have equality in the inference process, and each has deficiencies and advantages (Sohn, Kim 2012).

MATERIAL AND METHODS

The data set of the Telematics Services business obtained from the National Census in 2006 consisted of 8798 Telematics Services SMEs with attributes of 21 attributes (4 numeric attributes and 17 categorical attributes). Data for this classification have undergone pre-processing stages of data mining such as data cleansing and data integration conducted by previous research by Tosida et al. (2016). The available data is unbalanced data on the key attribute. In this case 93% is data receiving assistance and 7% is data not receiving assistance. Here is a description of the data (Table 1), LVQ algorithm flowchartand the architectureused (Fig1). In this classification system validation test using confusion matrix method(Han et al. 2012) and alpha value (α) and alpha (Dec α) with 2 different scenarios with balancing data and without balancing data to find alpha value (α) and alpha (Dec α) with the highest percentage to be implemented into the application.

RESULT

A comparison table of test results aimed at Table 2. It can be concluded that the first scenario with balancing data with 49.545% percentage value and alpha value changes and Dec α does not affect the percentage results. The second scenario without balancing the data with the best percentage of 93.112% and Alfa and Dec α changes with more than 150 iterations affect the percentage results. The smaller the Dec α value and the more iterations the percentage value will decrease. The comparison with the ID3 and C45 algorithms using the same scenario with 20 attributes without balancing data can be seen in Fig. 2. The output of the LVQ algorithm is then implemented in the form of a web-based decision support system to be accessible to MSME owners to predict whether it is possible to receive assistance. The look of the decision support system for this problem can be seen in Fig. 3.

DISCUSSION

Implementation of Learning Vector Quantization (LVQ) algorithm for the classification of Indonesian telematics services SMEs assistance has been successfully designed and built. Implementation of this system using Matlab software to build Learning Vector Quantization (LVQ) algorithm. Adobe Dreamweaver is used to build web pages with PHP programming language stored in the *Yii* framework. The *Yii* framework itself has advantages of fast and easy design process, then to design using Bootstrap Template for web view to be responsive, and database design using MySQL). The research phase begins with the analysis of the system that is looking at the description of data to be used as training data, database design is done with ERD (Entity Relationship Diagram) and DFD (Data Flow Diagram). The model base describes the flow of the Learning Vector Quantization (LVQ) algorithm. System validation test using confusion matrix.

The total data used is 8798 data, and 2 output classes (got assistance and no assistance). Data divided by 2 is 80% as training data and 20% as test data. Results from the data use 2 scenarios. The first scenario got 21 attribute with unbalanced data gets 93,112% accuracy. The second scenario got 21 attribute with biaxial data gets 49,545% accuracy. From the scenario can be concluded that the scenario with 21 attributes with data unbalanced the best level of accuracy is 93.112% to be implemented into the system.

Implementation of LVQ for optimization of the prediction process of providing assistance for Indonesian telematics services SMEs generate weight that is easy to be traced. The weighting details for the two specified classes are shown in Table 3. Based on Table 3 shows that the attribute that has the highest weight to receive assistance is the repayment services attribute. It can also be interpreted that according to the LVQ model the most influential attribute to the prediction of aid is the repayment services attribute. The attributes of repayment services or remuneration are also influenced by labor conditions in the field of telematics services in general have a value of remuneration or a higher level of salary compared to other fields. This condition is in accordance with various strategic efforts undertaken by the government to strengthen the start-up of the field of

telematics through the strengthening of human resource competence(McGuirck et al. 2015). The condition of repayment services for Indonesian telematics services SMEs is also closely related to the educational conditions of SME owners. The type of business of telematics services is strongly influenced by the level of innovation and high technology. Therefore, high ownership education is expected to encourage and become a catalyst for the development of Indonesian telematics services SMEs (Lee 2011; Dhewanto et al. 2012).
The LVQ model built on Susenas 2006 also shows that the attributes of business group types have a high impact in the relief process. This condition is in line with the results of research Tosida et al. (2017) which states that one of the characteristics of business clusters of telematics services that are eligible to be provided has a characteristic of high business improvement. This is shown from the type of business group Computer Settings Services and Internet businesses. In 2006 those are telematics service business that is growing with excellent business prospects. Of course this is no longer relevant to the current conditions, but these findings indicate that the optimization of the prediction process using the LVQ model based on Susenas 2006 data has a high degree of relevance to the conditions at that time. So this further strengthens the performance of decision support models for the provision of assistance for Indonesian telematics services SMEs.

Another attribute that has a real effect on the process of providing assistance is the type of business legality. The dominance of the type of legality of SMEs telematics services until 2006 is the type of individual business. This type of business activity is relatively small and simple, resulting in relatively low organizational costs, relatively flexible management, and also has a low risk business (Marcelino et al. 2014). Because of limited investment capability so that business development is limited, as well as with managerial skills. Based on this, the model results are in conformity with the real condition, because the Indonesian government is intensively developing and developing start-up based technology, especially start-up based on information and telecommunication technology (telematics). This strategy is carried out to strengthen the telematics field built from the SMEs telematics able to have a strong competitiveness (Daniel et al. 2015), especially facing the AEC(Tosida et al. 2016).

Finally, this decision support model base LVQ has a usefulness as a classification of the feasibility of Indonesian telematics services SMEs, receiving assistance using rules or rules that have been made and store data classification and its decisions. This research can be developed by using all the attributes that exist without selection to increase the accuracy and input values that can use the *csv* file to facilitate the input of many data and also the implementation of dynamic algorithm in the web so that the rules can be changed when the test data in the modification.

ACKNOWLEGMENTS

1. DRPM Ristek Dikti, 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

Blanco, A., R. Pino-Mejias, J. Lara, S. Rayo. 2013. Credit scoring models for the microfinance industry using neural networks : Evidence form Peru. Expert Systems with Application 40 (2013) 356-364. Science Direct, Elsevier. http://dx.doi.org/10.1016/j.eswa.2012.07.051.

Daniel Palacios-Marqués, Pedro Soto-Acosta, José M. Merigó. 2015. Analyzing the effects of technological, organizational andcompetition factors on Web knowledge exchange in SMEs. J. Telematics and Informatics 32 (1) : 23–32. Elsevier,http://dx.doi.org/10.1016/j.tele.2014.08.003.0736-5853/@2014.

Dhewanto W, Prasetio EA, Ratnaningyas S, Herliana S, Cherudin R, Aina Qorri, Bayuningrat RH, Rachmawaty E. 2012. Moderating effect of cluster on firms innovation capability and business performance : A conceptual framework. Procedia-Social and Behavioral Science Vol. 65: 867-872. doi: 10.1016/j.sbspro.2012.11.212.

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

Ince H, Aktan B. 2009. Comparison of data mining techniques for credit scoring in banking : A managerial prespective. *Journal of Business Economics and Management*, 10(3) : 233-240. Publiser : Taylor & Francis.DOI: 10.3846/1611-1699.2009.10.233-240
Lee H, Lee S, Byungun Y. 2011. Technology clustering based on evolutionary patterns : the case of information and communications technologies.Journal Technological Forecasting & Social Change, 78 (6) : 953-967. Elsevier. doi:10.16/j.techfore. 2011.02.002.

Marcelino-Sadaba S., Perez-Ezcurdia A, Lazcano AME, Villanieva P. 2014. Project risk management methodology for small firms. *International Jounal of Project Management* Vol. 32 (2014) 327-340. Elsevier.http://dx.doi.org/10.1016/ j.ijproman. 2013.05.009.

McGuirk H, Lenihan H, Hart M. 2015. Measuring the impact of innovation human capital on small firms propensity to innovative. *J. Research Policy, Vol.44 (4)* : 965-976. Elsevier. doi:http://dx.doi.org/10.1016/j.respol.2014.11.008.

Sadatrasoul, SM, Gholamian MR, Siami M, HajimaohammadiZ. 2013. Credit scoring in banks and financial institutions via data mining techniques : A literatur review. *Journal of AI and Data Mining*. Vol. 1, No. 2, 2013, 119-129. [Diunduh pada 2015 Jan 5]. Link : http://jad.shahroodut.ac.ir/article_124_0.html
Sohn SY, Kim JW. 2012. 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.

Tosida ET, Maryana S, Thaheer H, Damin FA. 2015. Visualization model of Small Medium Enterprises (SMEs) telematics services potentiality map in Indonesia. ***Published*** **in:International Conference on** Information & Communication Technology and Systems (ICTS), 2015. Added to IEEExplorer. January 14th 2016. **DOI:** 10.1109/ICTS.2015.7379890.

Tosida, E. T., S. Maryana, and H. Thaheer. 2016. Atribut selection of Indonesian Telematic Services MSMEs Assisstance Feasibility, using AHP. Kursor 8 (2), 2016. DOI: http://dx.doi.org/10.21107/kursor.v8i2.1299.

Tosida ET, S Maryana, H Thaheer. 2017 Implementation of Self Organizing Map (SOM) as decision support : Indonesian telematics services MSMEs empowerment. IOP Conference Series : Materials Science and Engineering 166 (1), 012017. IOP Publishing. http://iopscience.iop.org/article/10.1088/1757-899X/166/1/012017/meta.

Zhu, Y., C. Xie, Gang-Jin Wang, Xin-Guo Yan. 2016. Comparison of individual, ensemble and integrated machine learning methods to predict China's SME credit risk in supply chain finance. Neural Comput & Applie. The Natural Computing Applications Forum 2016. CorssMark, Springer. DOI 10.1007/s00521-016-2304-x.

Table 1.Description of data

No	Attribute	Type	No	Attribute	Type
1	Province	Categorical	11	Labor	Numerical
2	Education owner	Categorical	12	Repayment services	Numerical
3	The form of legal entity	Categorical	13	Difficulty	Categorical
4	Year Start of Operation	Categorical	14	Cooperative	Categorical
5	Computer users	Categorical	15	Partnership	Categorical
6	Internet user	Categorical	16	Receive training	Categorical
7	Business Group	Categorical	17	Marketing	Categorical
8	Sales	Numerical	18	Company condition 3 months ago	Categorical
9	Total assets	Numerical	19	Estimated 3-month business Prospects will come	Categorical
10	Capital	Numerical	20	Plan	Categorical

Table 2. Comparison of Test Results

Training			Data	
Alfa (α)	Dec α	Epoch	Balancing	Without Balancing
0,1	0,1	50	49,545%	93,112%
	0,01	100	49,545%	93,100%
	0,001	150	49,545%	78,415%
0,2	0,2	50	49,545%	93,112%
	0,02	100	49,545%	93,112%
	0,002	150	49,545%	90,134%
0,3	0,3	50	49,545%	93,112%
	0,03	100	49,545%	93,112%
	0,003	150	49,545%	93,089%

Table 3. The Attributes Weight of The Prediction Classes

No	Attribute	The Weight Value of Classes		No	Attribute	The Weight Value of Classes	
		RA*	NA*			RA*	NA*
1	Province	1.938	-5.231	11	Labor	2.203	-6.199
2	Education owner	1.186	-2.615	12	Repayment services	30.254	-2.615
3	The form of legal entity	4.886	-11.471	13	Difficulty	1.336	-2.615
4	Year Start of Operation	1.065	-2.615	14	Cooperative	1.892	-2.615
5	Computer users	1.916	-7.846	15	Partnership	1.524	-4.411
6	Internet user	1.088	-2.908	16	Receive training	2.896	-2.615
7	Business Group	5.922	-12.204	17	Marketing	1.405	-3.287
8	Sales	2.349	-8.212	18	Company condition 3 months ago	2.489	-9.092
9	Total assets	2.367	-9.956	19	Estimated business prospects next 3-month	1.938	-2.994
10	Capital	1.058	-3.287	20	Plan	1.219	-3.249

Notes : RA = Receives Assistance; NA = No receives Assistance

FIGURES.

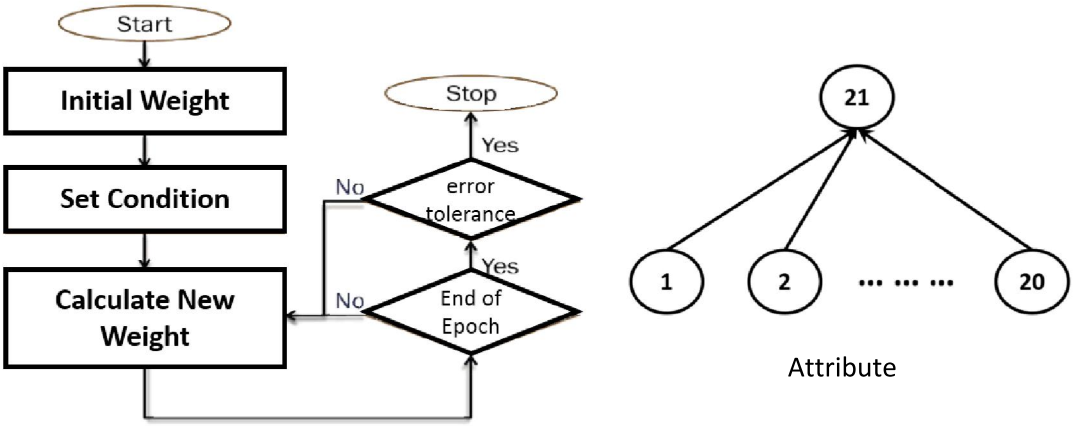

Fig. 1. LVQ algorithm flowchart and the architecture used.

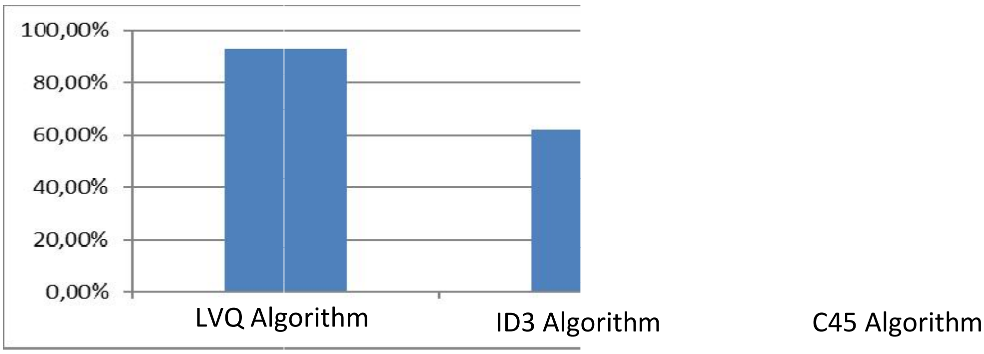

Fig. 2. Comparison of accuracy between algorithms.

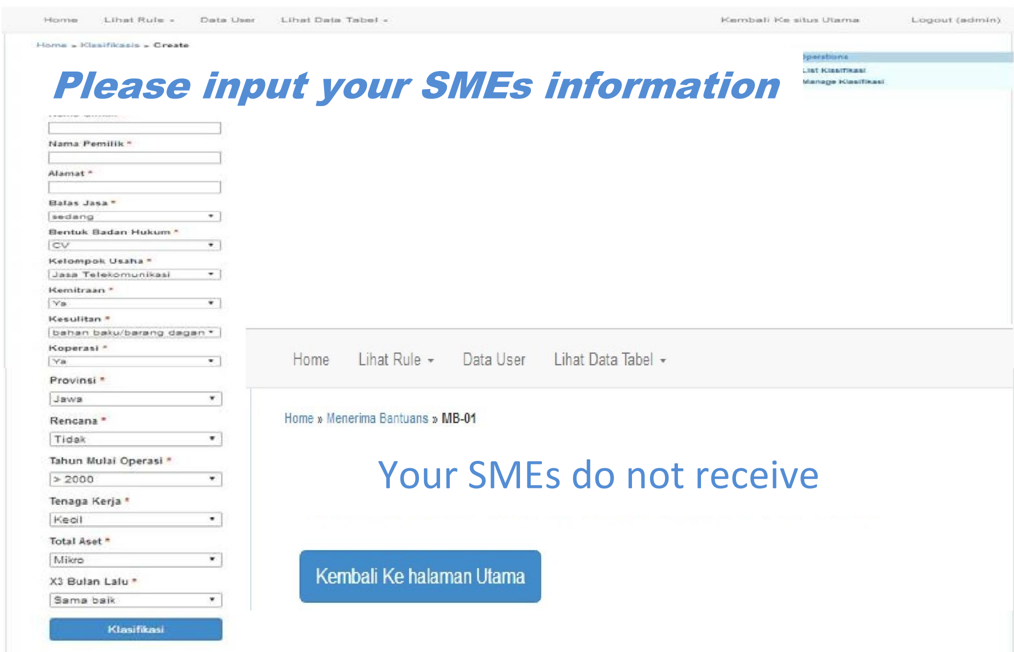

Fig. 3. The interface of decision support system for Indonesian telematics services SMEs