Implementation of Self Organizing Map (SOM) as decision support: Indonesian telematics services MSMEs empowerment

Implementation of Self Organizing Map (SOM) as decision support: Indonesian telematics services MSMEs empowerment

E.T. Tosida, S. Maryana, H. Thaheer, Hardiani

Computer Science Department, Pakuan University, Jl. Pakuan PO Box 452, Indonesia

E-mail: enengtitatosida@unpak.ac.id

Abstract. Information technology and communication (telematics) is one of the most rapidly developing business sectors in Indonesia. It has strategic position in its contribution towards planning and implementation of developmental, economics, social, politics and defence strategies in business, communication and education. Aid absorption for the national telecommunication SMEs is relatively low; therefore, improvement is needed using analysis on business support cluster of which basis is types of business. In the study, the business support cluster analysis is specifically implemented for Indonesian telecommunication service. The data for the business are obtained from the National Census of Economic (Susenas 2006). The method used to develop cluster model is an Artificial Neural Network (ANN) system called Self-Organizing Maps (SOM) algorithm. Based on Index of Davies Bouldin (IDB), the accuracy level of the cluster model is 0.37 or can be categorized as good. The cluster model is developed to find out telecommunication business clusters that has influence towards the national economy so that it is easier for the government to supervise telecommunication business.

1. Introduction

Information technology and communication (telematics) is one of the most rapidly developing business sectors in Indonesia. The sector has strategic position in its contribution towards planning and implementation of national developmental, economics, social, politics and defence strategies. Studies that focus on development of the national telecommunication service are pretty limited despite the annual growth of the sector that reaches nearly 4% [1]. A study conducted by [1] of which focus is telecommunication service reveals that the most preferable products of telecommunication for the public as the end-users of telecommunication products are ones related to computer and internets, either direct use of computer and internets or in the form of information. Indonesia ranks the fifteenth in terms of countries with the largest internet users even though level of penetration is still pretty low (1.7%). Shifting to digital communication is launched as public information policy initiatives [2]. It is expected that the initiatives become the key contributor to spread internet access in every household. Public policy has an active role in expanding and promoting information technology diffusion that emphasizes on quality of service.

One method to prepare telecommunication industry as national priority is developing model for classifying the national telecommunication service [1]. In order to develop the classification, qualitative study that compares and harmonizes types of business groups should be conducted. The classification model has yet been completed with detail characteristics related to potential of each area and its contribution towards developing factors for telecommunication service. The data related to Indonesian telecommunication service mapping is relatively large. The approach used to deal with large data is

Knowledge Discovery in Database (KDD) approach and one of its components is widely known as Data Mining. According to [3], data mining is more meaningful as a means of gaining knowledge (data) from interdisciplinary subject. One of data mining techniques implemented to explore characteristics of particular business sector is clustering.

Implementation of clustering related to business development strategy has been conducted by some researchers for example clustering-based strategic formulation to explore business characteristics [4]. Information and Communication Technology (ICT) sector is closely related to innovations, clustering towards types of existing technology based on its growing patterns as well as correlations between one technological information [5] and another can be used as strategic decision support help for the sector [6]. It is expected that the clustering can help stakeholders understanding characteristics of the industry or existing business groups more particularly telecommunication sector as well as analyzing process of innovations within the business groups.

Studies related to identification of Indonesian telecommunication service characteristics and potentials have been conducted by [1]. The studies are limited to qualitative and descriptive approach. Thus, deeper exploration that involves data mining is of necessity. The concept should be equipped with visualization of business developing factor inter-correlation as well as potentials of each area using clustering process with Artificial Neural Network (ANN) with SOM algorithm.

2. Indonesian Telematics Services MSMEs Competences

Indonesian telematics businesses are grouped into the hardware industry; the software industry; and industrial and non-industrial telematics services. The results of Multi Sectoral Qualitative Analysis (MSQA) concluded the level of interest of the development of the telematics industry types in Indonesia [7]. Specifically the study of the classification of telematics services business in Indonesia resulted in three major groups namely business services, communications services and education services [1]. This classification model is obtained through a comparative study and followed by expert acquisition using the Delphi method. The basis of the development of this model is taken from the document of the World Trade Organization / WTO [8], as it relates to the preparation of Indonesia towards the liberalization of the field of telematics. These fields have a subfield and in detail derived by following the numbering systematically arranged following the numbering system. The numbering system is known as Group of business scope standard in Indonesia (ISIC).

Related to the development business of telematics services group in Indonesia, has been arranged competence map services telematics group [9]. The existing competence map remains dominated by a group of MSMEs telematics service, which are have the conditions that still needed help in the development of their business. One of the government's efforts in empowering SMEs telematics through the ICT Fund. ICT Fund which was directed to improve business ecosystems, including through venture capital assistance, with the framework of Evolution Concept of USO (Universal Service Obligation).

Indonesian telematics service business competence can be identified through the value chain framework. The results of this framework indicates that the computer consulting business is a business that has the support of both internal and external, so that it becomes easier in its development. Field of telematics education efforts will still need affirmative action to strengthen the competence of external support. Publishing Software, Computer Programming, Information Technology and other market potential is still vast, but the internal capabilities need to be improved. Computer Repair Enterprises, Web Portal, Hosting, and Special Design, began to show saturation. This matter may be attributable to the increasing number of businessmen, the increasingly fierce competition, and the carrying capacity of the human resources inadequate [10].

3. Method

This research resulted a model that includes empowering of national telematics services MSMEs, as one of the strategies to increase competitiveness, which is done through clustering models. This model is

supported by Data Mining process. Research carried out in stages using the stage as shown in Figure 1. The main data used are economic census sample data from the Central Bureau of Statistics (BPS), with the scope of entities and attributes that will be tailored to the needs of stakeholders.

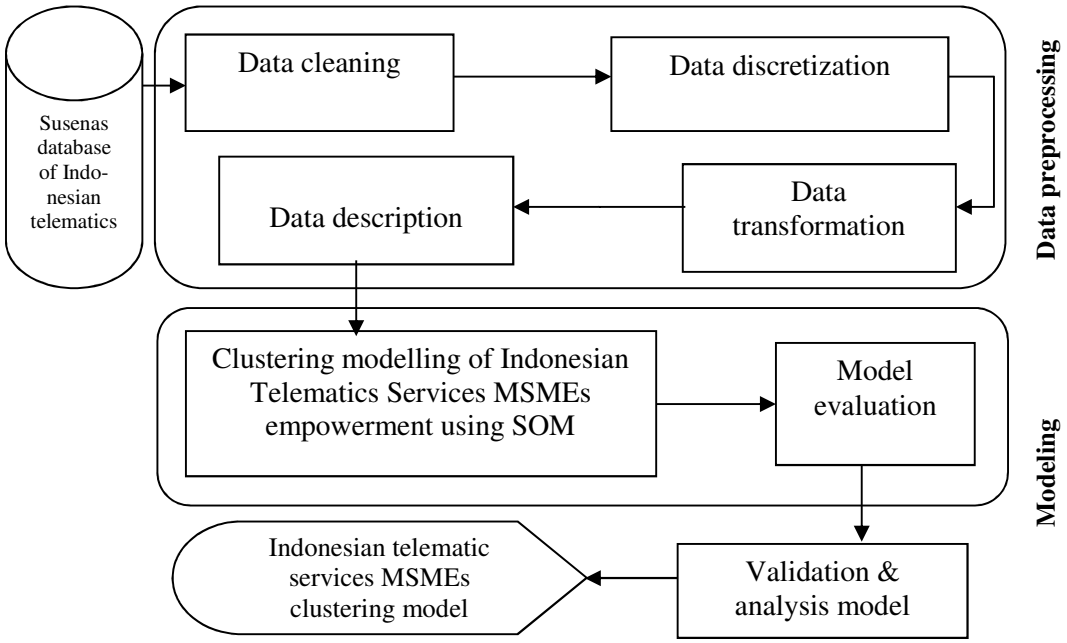

Figure 1. Research method

4. Result and Discussion

4.1. Analysis and Design Clustering Model used SOM method

Analysis of system and network architecture SOM algorithm performed in order to determine the work process SOM algorithm. This was done to evaluate and give an overview of clustering models Indonesia telematics services MSMEs to be created. SelfOrganizing Map (SOM) or what is often called a Kohonen neural network topology is one form of Unsupervised Artificial Neural Network (ANN Unsupervised, where in the training process does not require surveillance (target output). Kohonen network / SOM is used for grouping (clustering) data based on the characteristics / features of the data [11]. SOM algorithm is shown in Figure 2 .

SOM architecture consists of two layers that are the input and output. Inter- layers are connected by a certain weight are often referred to as representation vector or vector code book or reference vector [11]. In the one-dimensional architecture, input layer (X) will take the form of linear and similarly the output (Y). In this architecture, the adjacent unit would have been more different than the farther unit. Weights (W) is used as one of the components to determine the distance of the output layer. The shortest distance will become references clustering data. If there is a distance to the second output is the closest then the vectors are included will be placed in the second output class, and then the weight that is to be updated by considering the inclusion of the input vector to the output layer of the second class. This new weight then the next will be a reference for further input.

Architecture of SOM in general is shown in Figure 3. Case of Indonesian telematics services MSMEs clustering model based on the architecture as shown in Figure 4. The results of interviews with experts and refers to the data Susenas 2006, determined 10 factors that affect the empowerment of SMEs telematics services Indonesia [12]. Therefore, the inputs used in the architecture of SOM in these cases totaled 10. Determination of the number of clusters based on the type of Indonesian telematics services

MSMEs that was consist of 10 types. Type of MSMEs encompasses Rental Services Computer and VCD, Repair and Computer Service, Printing Services, Photo Copy, Training Service and Private Computers, Service Setting Computer and Internet Services Data Processing and Analysis Statistics, Consulting Services Software, Services Procurement and Assembling Computer, Buy-Sell Computers and accessories and Computer Services other. Thus the output of architecture SOM for this case amounts to 10. The SOM architecture for Indonesian telematics services MSMEs clustering models shown in Figure 4 .

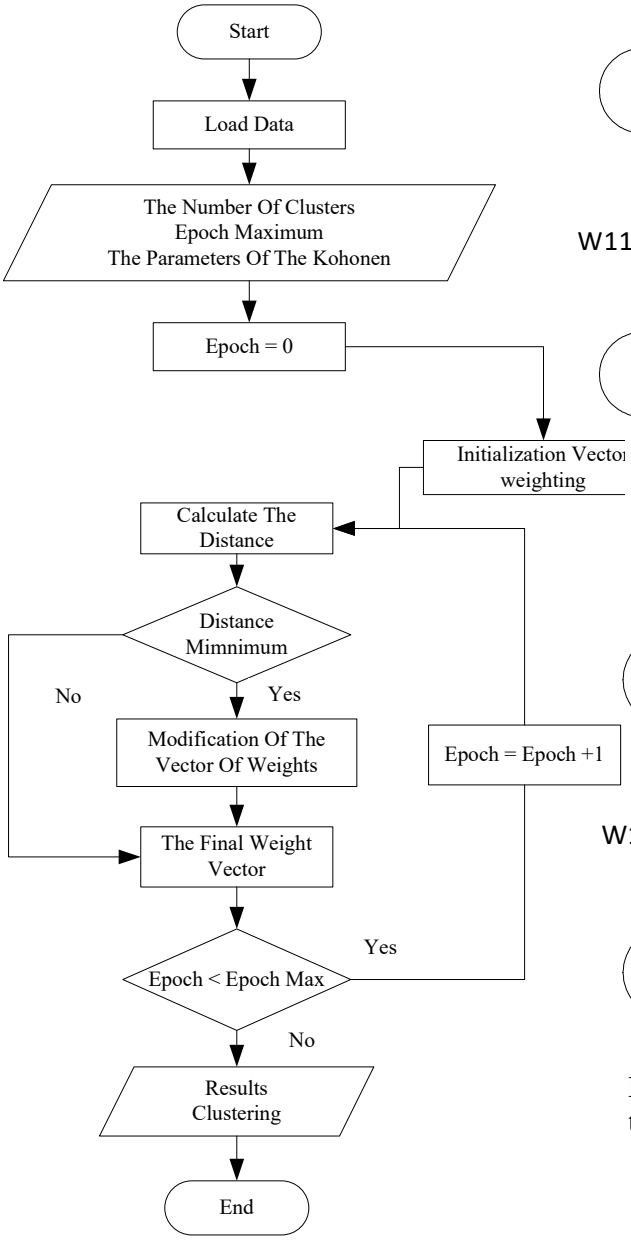

Figure 2. SOM algorithm

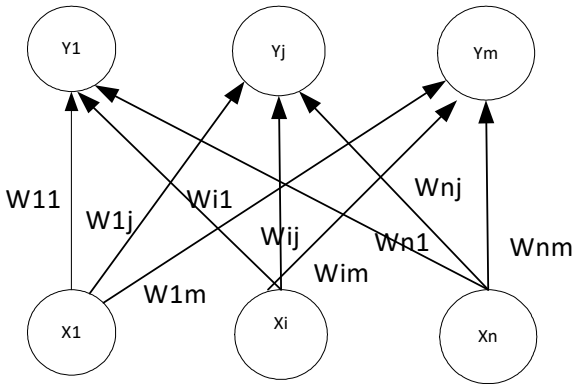

Figure 3. SOM architecture in general

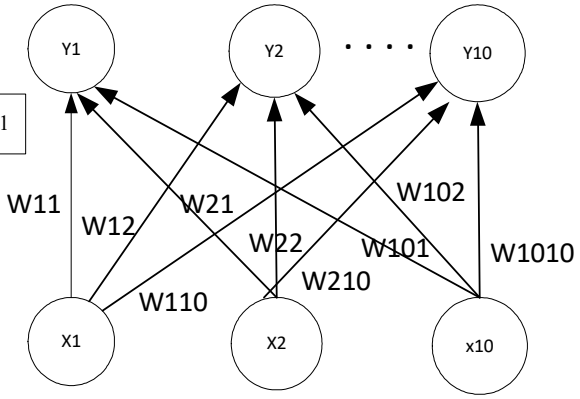

Figure 4. SOM architecture of Indonesian telematics services MSMEs clustering model

4.2. Clustering Indonesian Telematics Services MSMEs Model

Model of clustering is developed base on the data from the 2006 National Economic Census. The number of samples are 2704 that is divided into two, 70% for training data and the remaining 30% for testing data. Clustering model outlook is developed using *plotsomhits* (Matlab 2010a). The outlook makes it easier for users to find out total number in each cluster. Testing was conducted using different

clusters, namely 10, 7, and 5 clusters. The basis of using the clusters in the testing related to [5]. Epoch used are 100, 80, 50, 30 while learning rate is 0.01, 0.03, 0.04, 0.05, 0.07, 0.08 and 0.09. The clustering model is conducted using the SOM algorithm and is evaluated using Index of Davies Bouldin (IDB). Index of Davies Bouldin aims at maximizing inter cluster and minimizing intra cluster in one cluster. The index is used to measure validity from result of clustering in order to obtain the most suitable IDB, which is minimum IDB (Zaki dan Meira, 2013).

The results of modelling using training data is shown in Figure 5, while the results of modelling using testing data shown in Figure 6 .

Figure 5. Experimentation result of clustering models using training data

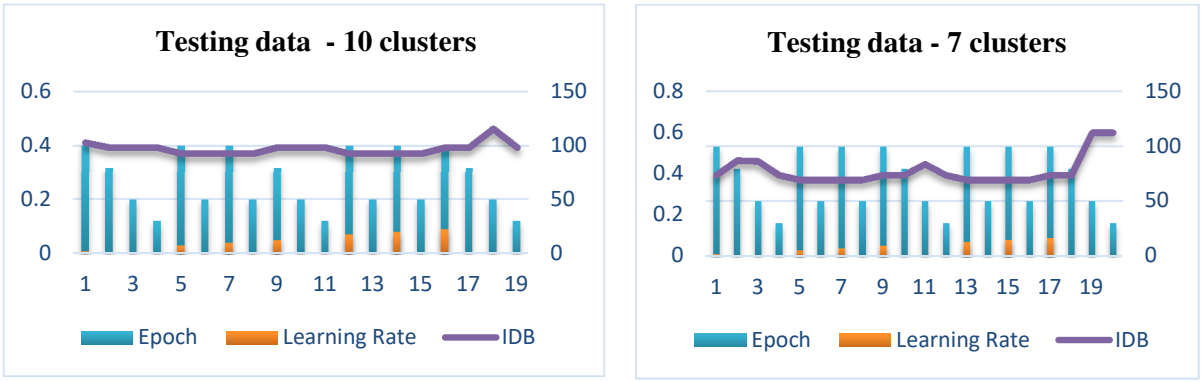

Figure 6. Experimentation result of clustering models using testing data

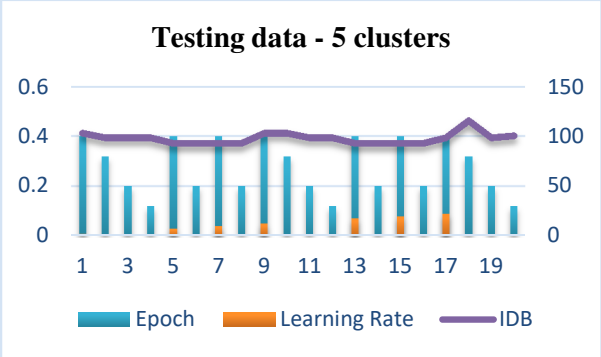

Figure 6. Experimentation result of clustering models using testing data (continued)

Based on Figure 5 experimentation clustering models using training data IDB has produced a minimum of 0.35. The experimental results of the training data indicated that the number of clusters and Learning Rate (LR) did not affect the achievement of the minimum IDB. The epoch affect the achievement of the IDB. The experimental results from the training data, shows that the minimum value of IDB was in the epoch 80. These results form the basis for the learning process is done for modelling using testing data (based on Figure 6). Experiments to testing data showed similar characteristics with the results of training data, which changes the number of clusters and LR did not affect the achievement of the minimum IDB. IDB minimum value obtained at 0.37 value and occurs in epoch 80. Value IDB experimental results of a more stable testing data at each change of LR and epoch, when compared with the experimental results of the training data. Clustering models obtained in this study as a whole has achieved fairly good accuracy. This is demonstrated by the achievement of IDB relatively small value nearly zero.

Further analysis of the experimental results show interesting findings. One example in cluster 5 the modelling results using 10 clusters with optimal condition of the IDB, has found the characteristics of MSMEs conditions unique telematics services. The condition in question include the following: the amount of capital has an average of 4 million, with an average spending of 1 million, and the average income of over 2 million. Business entities that exist in this cluster dominant following the guidance or counselling and the state of the company and the company's prospects were very good. Based on this information, the model of clustering using SOM method can be used as a decision support for the government in the process of empowerment of Indonesian telematics services MSMEs. Based on 2006 Susenas, cluster 5 is referred to the types of Computer Settings Services and Internet businesses. In 2006 those are telematics service business that are growing with excellent business prospects. Based on the model results clustering exposure conditions, the government can direct the kind of empowerment that is more suitable for this type of business. These telematics services business is dominated by businesses that have received guidance and counselling, as well as the conditions were relatively good profit margins. Therefore, empowerment which was more suitable for these business type was increase the competence of human resources or marketing competitive regulation. So that in the future these types of telematics services business is expected to have a stronger competitiveness.

Exposure to other conditions may be identified through this clustering models. It shows that through the Susenas database, SOM method can be used as decision support for the government in determining the type of empowerment of SMEs telematics services. Disadvantages of this study include data used still have data that is not balanced on a number of factors or parameters used. Therefore, further research requires the balancing process the data, so the expected value obtained IDB will become more optimal.

5. Conclusion

Model clustering 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. The experiments show that the diferent number between cluster and Learning Rate (LR) did not have any influence towards optimum IDB scores. Optimum clustering model takes place when epoch is 80, with IDB score of 0.37. Different occurence happens towards the training data resulting in IDB score of 0.35. The clustering characteristics of testing data is similar to the model of training data clustering. It shows that the model works well. Prominent clustering model takes place in cluster 5 that shows dominating number of telecommunication MSMEs that less worthy to get grants compared to other clusters. Such domination causes bias in the result of cluster as the effect of imbalance number of data in one of the types of telecommunication MSMEs. However, the model results clustering using SOM has been able to provide information of interest and can be used as decision support for the government in determining the type of empowerment of MSMEs telematics services. This information was obtained from the results of further analysis of the characteristics of the cluster formed from the model . Disadvantages of this study are associated with the data that is not balanced is expected to be overcome with the preprocessing of data that involves the balancing process of the data. Thus the model obtained will be more accurate with the IDB a more optimal value.

Acknowledgment

1. DP2M DIKTI Educational and Cultural Ministry 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 and user requirement.

References

[1] Tosida ET, Harsani P, Hermawan, Setyaningsih S. 2012. Classification models of information technology services bussiness in Indonesia. Proceeding International Seminar on Science and Technology Innovations 2012. University of Al Azhar Indonesia, October 2-4th 2012, ISBN 978-602-95064-5-7, UAI Press. 304-309.

[2] Sourbati M. 2011. The digital switchover as an information society initiative: The role of public policy in promoting access to digital ICTs. Journal of Telematics and Informatics, Volume 28, Issue 4, November 2011, Pages 295–30.

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

[4] Hadighi SA, Sahebjamnia N, Mahdavi I, Shirazi MA. 2013. A framework for strategy formulation based on clustering approach: A case study in a corporate organization. Knowledge-Base Systems, 49 (2013), 37-49. http://dx.doi.org/10.1016/j.knosys.2013.04.008.

[5] Lee H, Lee S, Byungun Y. 2011. Technology clustering based on evolutionary patterns: The case of information and communications technologies. J. Technological Forecasting & Social Change, 78 (2011) 953-967. DOI: 10.1016/j.techfore.2011.02.002.

[6] Tsui Chia-jung, Ping W, Fleischann KR, Oard DW, Sayeed AB. 2010. Exploring the relationships among ICTs: A scalable computational approach using KL divergence and hierarchical clustering. Proceedings of the 43rd Hawaii International Conference on System Sciences. 978-0-7695-3869-3. Added to IEEExplorer. DOI: 10.1109/HICSS.2010.203

[7] Setyaningsih S, Hermawan, Tosida ET. 2013. Competency mapping of human resources in the field of Indonesian telematics industry: As an investment policy. Proceedings of the National Conference of Mathematics and Natural Sciences: MIPA as Platform Creation and Technological Innovation. IPB International Convention.251-258.

[8] WTO. 1991. Services sectoral classification list [Internet]. [Downloaded 2012 Feb 10]. Available at https://www.wto.org/english/tratope/serve/mtn_gns_w_120_e.doc.
[9] Tosida ET, Maryana S, Thaheer H. 2014. The Potencies of Indonesian Telematics Services Enterprises Group. Proceedings of National Conference of Information, Communication & Management Proceeding. Bina Darma University, Palembang, August, 2014, 45-52.
[10] Tosida ET, Thaheer, Maryana S. 2015. Strategy to enhance competitiveness through the value chain of framework for the Indonesian telematics business service competence. J. Postal and Informatics Research. Vol. 5(1) 2015. 1-18.
[11] Fausett L. 1994. Fundamentals of neural networks, architectures, algorithms, and applications. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, ISBN: 0-13-334186-0.
[12] 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.
[13] Zaki MJ, Meira Jr W. 2013. Data mining and analysis: Fundamental concepts and algorithms.