Customer Classification using Learning Vector Quantization Neural Network

Ari Basuki
Industrial Engineering Dept., University of Trunojoyo Madura, Indonesia.
aribasuki@trunojoyo.ac.id

***Abstract*—The application of customer relationship for Small-Medium Sized Enterprise (SME) is still at the developmental stage. Knowledge obtained from Customer Relationship Management (CRM) can help SMEs to estimate the profitability of individual accounts. This paper presents a prediction model to identify the customers who are likely to purchase the offered product of a company based on their past purchasing history. Experiments using Learning Vector Quantization Neural Network were conducted to classify the potential customers into the purchasing and non-purchasing categories. The results of experiments reveal that the best parameter to model customer classification using this data set has high accuracy.**

***Index Terms*—Customer Classification; Customer Relationship Management (CRM); Learning Vector Quantization (LVQ); Neural Network (NN).**

I. Introduction

In Customer Relationship Management (CRM), there are three basic problems to solve: how to get new customers, how to retain the existing customers and how to maximize the customer's spending. Getting new customers or retaining the existing customers is crucial to ascertain the sustainability of a company. In other words, a company should be able to identify which customers are attainable and which customers are kept for short/long term as this information can influence the companies' life. Therefore, the core of analyzing their customers is to classify them.

Understanding customers is the secret of successful selling. Existing customers are an important source of information; hence, the more a company knows about the customers, the easier the company finds a strategy to increase the customers' spending. Moreover, this information is essential to find targeted new customers. The information collected depends on company's type of business. For example, if the company is selling a product to individual consumers, the company needs to know the customers' age, gender, location, spending habits and income. When the company is selling a product to other businesses, it is necessary for the company to identify what sector they are in, how big they are, how much they spend and what other suppliers they use. This kind of information can be obtained by analyzing the sales records, talking to the customers, and conducting surveys.

This research investigated customers' characteristics using customer classification techniques at the Batik SME in Bangkalan, Indonesia. As a business entity, this company has a goal to earn high profit. Therefore, it should obtain new customers to gain high profit. For that reason, the SME develops a market analysis by grouping its potential customers into two categories: the purchasing and non-purchasing customers.

The classification problems can be solved by data mining approach, such as Learning Vector Quantization (LVQ), Genetic Algorithm (GA), Decision Tree, etc. LVQ is one of classification algorithms that has a fast diagnosis speed, high accuracy and strong generalization ability [1].

The main goal of this work is to study the customers' classification of Batik SME in Bangkalan, Indonesia by applying the Learning Vector Quantization (LVQ) approach. This method is used to classify and to predict the customers who frequently respond to the product offered based on the previous purchased historical data. The experimental study showed the best parameter that fits perfectly to the given data set based on this algorithm.

II. Customer Classification

Customer classification is an important issue in real world marketing. It is believed that a company with strong understanding of its customer behavior patterns has a better chance to develop effective marketing strategies. In commercial operation, using the membership card system management is considered as the most superior method to help the businessmen to accumulate their customers' information. This system is helpful to either collect customers' information or offer corresponding service for different card-rank users. Therefore, it can enhance customers' loyalty to the company.

In order to develop a model to effectively differentiate purchasing customers from the non-purchasing customers, all possible factors, such as customer demographics and other supporting information were collected. The selection of reasonable classification variables is the basis of correct and effective customer classification. Consequently, the supporting information from experienced domain experts were collected to support the selection process. Table 1 summarizes some of the recent works done in this sphere by different authors about customer classification.

Based on analyzing and summarizing the existing literatures, the features designed in this research include age, marital status, number of child, and profession. The description of the data is presented in Table 2. Two hundred customer profiles were collected and used in this research, including the purchasing and non-purchasing customers.

Table 1
The summary of customer classification from literatures

Reference	Factors	Method
Mutanen, A.,et.al [2]	Electricity connection information (location, supply voltage, fuse size, number of phase), customer class, consumption (high tariff consumption, low tariff consumption), additional information	ISODATA (Iterative Self-Organising Data Analysis Technique)
Patil, N.,et.al [3]	Yearly income, Number of Children, Number of Cars, Marital Status, House Owner,Country, Membership card	Decision Tree (C 5.0, CART)
Neethu, B. , et.al [4]	Account type, Age, Tax, Customer Type, Qualification, Income, Loan Sanction	Naive Bayes Classifier
Xiao, J., et. al [5]	Customer level, charges for range call, charges for domestic long-distance call, charges for intra-regional call, charges for international call, monthly fee, monthly total fee, average times out of service in 3months, average expenditure in 3months	Dynamic Classifier Ensemble
Zhu, Q., & Zhang, Y. [6]	Bike buyer, Age, Education, Yearly income, Total Children, Marital Status, Cars Owner, Commute Distance	Naive Bayes Classifier
Xinwu, L. [7]	Post code, Birthdate, Sex, Qualification, Occupation, Monthly Income, Marital Status	Particle Swarm Optimization
Abdillah, G, et.al [8]	Profession, number of occupant, total usage	K-Means

Table 2
The description of case features

Feature	Data Type	Content
Age	Integer	Range [1 - 70]
Marital Status	Integer	No = 0, Yes = 1
Number of Children	Integer	Range [1 - 8]
Profession	List	Range [1 - 10]

There are various data mining classification algorithms. This research applied the Learning Vector Quantization in neural network to classify potential customer into two categories, either purchasing or non-purchasing customers. Data derived from real world Indonesian SME were collected from early 2015 until early 2016 and they were separated into training and testing group for model construction.

III. Neural Network

The human brain can be expressed as an interconnected web of neurons carrying detailed patterns of electrical signals. The input signal is received by dendrites and based on those inputs as an axon transmits output signal. Similarly, an Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way the brain processes information [9]. The key element of this paradigm is the novel structure of the information processing system. It is composed of a vast number of highly interconnected processing elements (neurons) working in unison to solve specific problems. The figure of real neural network and the artificial one is illustrated in Figure 1.

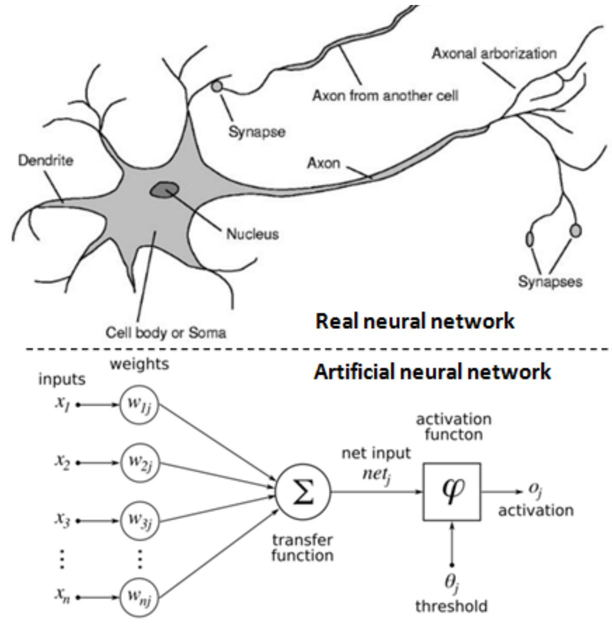

Figure 1: The real neural network and artificial neural network

The neuron has two modes of operation: the learning mode and the testing mode. In the learning mode, the neuron can be trained for particular input patterns. In the testing mode, when the system is given an input, then its associated output becomes the current output.

Moreover, neural networks are best at identifying patterns or trends in data. They are well suited for prediction or forecasting needs including sales forecasting, industrial process control, customer research, data validation, risk management and target marketing.

A. Learning Vector Quantization

Learning Vector Quantization (LVQ) is a supervised classification algorithm based on centroids or prototypes [10]. It can be interpreted as three layers competitive neural network. The first layer is only an input layer. The second layer is where the competition takes place. The third layer performs the classification. Each neuron in the competitive layer has an associated numerical vector of the same dimension as the input examples and a label indicating the class they will represent. These vectors are the ones that, at the end of the adaptive process, will contain information about the classification prototypes or centroids. There are different versions of the training algorithm. However, the one which is used in this article will be described.

Figure 2 represents the architecture of LVQ network. W1 is a weight vector that connects each neuron within input layer to the first neuron at the output layer, while W2 is a weight vector that connects each neuron within input layer to the second neuron at the output layer. The activation function F1 will project y_in1 to y1 = 1 if |X-W1| < |X-W2| and y2 = 0. Similarly, the activation function F2 will project y_in2 to y2 = 1 if |X-W2| < |X-W1| and y1 = 0 [11].

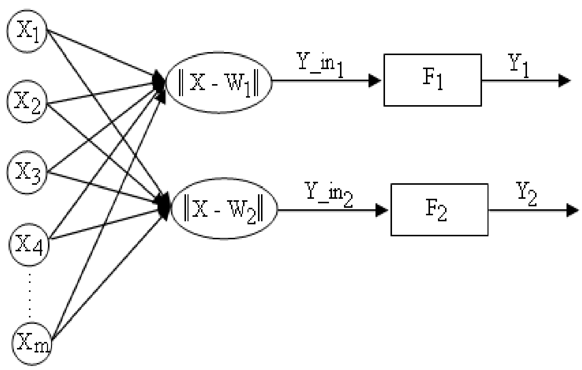

Figure 2: LVQ network architecture

The flows of the algorithm are:

1. Set the initial weight, MaxEpoch (maximum number of iterations), and Learning Rate (α, sufficient small value) and Eps (minimum error)
2. For each training vector with (iteration ≤ MaxEpoch) and (α ≥ *Eps*), find J so that D(J) is minimum. This calculation uses Euclidean Distance formula as (1).

$$D(J) = \sqrt{\sum \left(X_i - W_j\right)^2} \quad (1)$$

where: X_i = ith training vector
W_j= jth weight vector
Choose the minimum D(j) and represent as Cj
3. Update the weights of the J neuron (Wj) as follows:
T = Cj then Wj = Wj + α(Xi –Wj) (2)
T ≠ Cj then Wj = Wj - α(Xi –Wj) (3)
4. Reduce learning rate α

The implementation of Learning Vector Quantization towards research dataset is as follows:

1. Set the initial parameter. In this research, the parameters are set as MaxEpoch (200), Learning Rate (0.05) and Eps (0). The initial weight is set randomly as presented in Table 3.

Table 3
Initial weight

W1	0,2	0,1	0,3	0,2
W2	0,5	0,4	0,2	0,3

2. Calculate the euclidian distance using Equation (1). Before proceeding the data, normalizing the data within the standardize range is needed. Min max method is used to do this, and as a result, the data range would be 0 – 1 [12]. The normalization min max method uses the formula (4) as follows:

$$\sum_{i}^{n} f(x) = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Table 4 presents the normalized data, while Table 5 presents the euclidean distance for the first iteration. The minimum value of Euclidean Distance (Cj) is written in bold.

Table 4
The normalized value from raw data

Cust. ID	Age	Marital Status	# of Child	Profession	Purchasing Potential
1	0,684211	1	0,33	0,375	Y
2	0	0	0	0,875	N
3	0,578947	1	1	0,125	Y
4	0,736842	1	0,33	0,625	Y
5	0,473684	1	1	0,875	Y
6	1	1	0,33	0,875	Y
7	0,894737	1	1	0,5	Y
8	0,052632	0	0	0,125	N
9	0,052632	0	0	0	N
10	1	1	1	0,5	Y
...	...	...	...	...	...
196	0,105263	1	1	0,875	Y
197	0,947368	1	0,67	0,625	Y
198	0,526316	0	0	0,5	Y
199	0,789474	1	1	1	N
200	0,473684	1	1	0,75	Y

Table 5
Euclidean Distance (1st iteration)

No	Euclidean1	Euclidean2	Potential Customer	
			Prediction	Actual
1	1,0374	**0,646016**	Y	Y
2	**0,77177**	0,88353	N	N
3	1,20384	**1,018262**	Y	Y
4	1,13134	**0,734505**	Y	Y
5	1,35297	**1,153827**	Y	Y
6	1,38085	**0,97898**	Y	Y
7	1,36845	**1,093534**	Y	Y
8	**0,35685**	0,656326	N	N
9	**0,40214**	0,700099	N	N
10	1,42478	**1,135782**	Y	Y
…	...	....	...	...
196	1,32838	**1,219197**	Y	Y
197	1,29755	**0,939969**	Y	Y
198	0,5445	**0,490604**	Y	Y
199	1,51244	**1,25451**	Y	N
200	1,29515	**1,096901**	Y	Y

3. Update the weight using Equation (2) or (3). The updated weight from Table 5 is presented in Table 6. After that, the learning rate is reduced.

Table 6
The updated weight after 1st iteration

W1	4,253158	2,52	6,176667	4,29
W2	10,01316	8,28	4,256667	6,21

4. Conduct step 2 and step 3 until the (iteration =MaxEpoch) or (α =*Eps*). Then, the last updated weight is set as the weight parameter for testing case. The testing process procedure is similar to step 1 and step 2 of the training process procedure.

IV. RESULT AND ANALYSIS

This research aims to obtain the best parameter that results in high accuracy for the testing case. Therefore, two experimental scenarios were created: (1) finding the best parameter of the number of data training and learning rate; (2) finding the maximum iteration for better accuracy. These parameters were then used for the testing case.

In the first scenario, the initial learning rate adjusted start from 0,1 until 1. This experimental set up is described at Table 7. The result of this experiment is shown in Figure 3.

Table 7
The 1st experimental set up

Series Name	#Training Set	Percentage (Total: 200 records)
1	150	75%
2	100	50%
3	50	25%

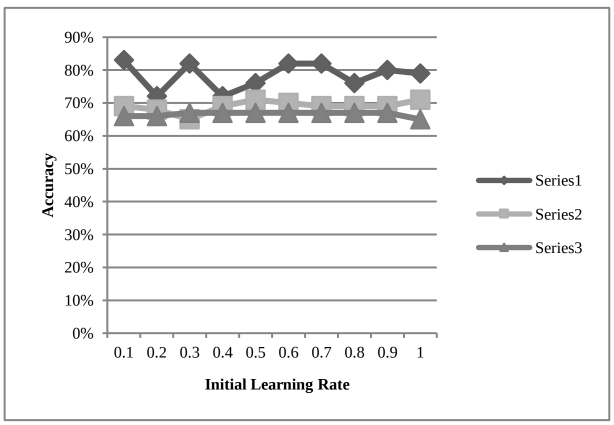

Figure 3: The relationship between initial learning rate and accuracy

The accuracy associated with the initial learning rate to the adjusted number of training data set can be observed in Figure 3. As shown in Figure 3, the differences between Series 1 to Series 2 and Series 3 are even more evident. Moreover, the best accuracy is acquired in Series 1. Therefore, series 1 (75% training data set) is selected as the best parameter for testing case. Based on Series 1, the accuracy is higher when learning rate = 0,1; 0,3; and 0,6. Thus, these parameters were then used in the second experimental scenario.

The second experiment was aimed to get the maximum iteration needed to get the best accuracy. This information is required in order to have effective iteration. This second experiment used the best parameter from the 1st experiment. Table 8 presents the relationship between the initial learning rate with the accuracy.

It can be observed in Table 8 that higher maximum iteration is likely get higher accuracy. When the max epoch is set = 200, the accuracy is still the same with max epoch = 500. Therefore, this indicates that in order to iterate effectively with high accuracy, it only needs to set the max epoch = 200 for testing case.

Table 8
The accuracy of 2nd experiment

Learning rate	Maximum Iteration (max epoch)			
	50	100	200	500
0,1	70%	83%	83%	83%
0,3	63%	69%	82%	82%
0,6	61%	69%	82%	82%

Through the experimental result, it is found that the best accuracy for the model could be reached by setting the parameters on 75% of utilized data set as the training data, and the learning rate was set as 0,1. In addition, the maximum iteration was set minimum 200 epoch. As a result, accuracy of the model obtained by using these parameters is 83%. This represents that the using of LVQ neural network for customer classification on this data set is appropriate with good performance.

V. CONCLUSION

Various parameters were applied for competing paradigms of LVQ neural network. The results show that higher number of data training leads to higher accuracy. Moreover, based on the parameter found during the experiment, the accuracy was 83%, which can be considered as excellent for practical problems.

The SME can plan effective marketing of its products by selecting the target customers. By picking the right customers, it can use Learning Vector Quantization algorithm which correctly fits the data set by using the correct parameters. This technique would help the marketing department to identify the respondents so that they would be targeted for particular campaigning activity. It also prevents wasteful expenditure of sending promotion offers to the non-purchasing potential customers.

This study was limited to CRM in Small Medium Enterprise's transaction. This work could be enhanced by building a total solution for CRM implementation for SME. All business units in the SME should be integrated in small adjustable method. Moreover, this work could be further enhanced by developing a new or hybrid algorithm which would classify the data with high accuracy and minor error rate.

REFERENCES

[1] S. Ding, X.H. Chang, Q.H. Wu, H.F. Wei, & Y.L. Yang, "Study on wind turbine gearbox fault diagnosis based on LVQ neural network". *Journal of Modern Electronics Technique*, 2014 vol 37, pp. 150-152.

[2] A. Mutanen, M. Ruska, S. Repo, and Jarventausta, P, "Customer classification and load profiling method for distribution systems,". *IEEE Transactions on Power Delivery*, vol. 26 ED-3, pp. 1755-1763, 2011.

[3] N. Patil, R. Lathi, and V. Chitre, "Comparison of C5. 0 & CART classification algorithms using pruning technique," In *International Journal of Engineering Research and Technology*, vol. 1 ED-4, pp. 1-5, June 2012.

[4] B. Neethu,L.T. Priyanka, "Customer Classification And Prediction Based On Data Mining Technique," In *International Journal of Emerging Technology and Advanced Engineering*, Vol. 2 ED-12, pp. 314-318, December 2012.

[5] J. Xiao, L. Xie, C. He, and X. Jiang, "Dynamic classifier ensemble model for customer classification with imbalanced class distribution," in *Expert Systems with Applications*, vol. 39 ED-3, pp. 3668-3675, 2012.

[6] Q. Zhu, and Y. Zhang, "Construction of customer classification model based on Bayesian network", in *Journal of Computers* vol. 8 ED-5, pp. 1200-1206, 2013.

[7] L. Xinwu, "Application Research of Improved Particle Swarm Algorithm in Online Trading Customer Classification", in *Information Technology Journal*, vol. 12 ED-14, pp. 3030-3033, 2013.

[8] G. Abdillah, F.A. Putra, F. Renaldi, "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru di PDAM Tirta Raharja Menggunakan Algoritma Kmeans", in *Seminar Nasional Teknologi Informasi dan Komunikasi*, pp 498-506, Maret, 2016.

[9] Z. He, X. Wen, H. Liu, and J. Du, "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region", in *Journal of Hydrology,* vol. 509, pp. 379-386, 2014.

[10] T. Kohonen, "Improved versions of learning vector quantization," in *IEEE International Joint Conference on Neural Networks*. pp. 545-550, 1990.

[11] D. Harjunowibowo, S. Hartati, R. Ariyuana, & A. Budianto, "Pattern Recognition on Paper Currency's Feature Using LVQ Algorithm," in *Proceeding of the Electrical Engineering Computer Science and Informatics*, vol. 2 ED-1, pp. 48-53, 2015.

[12] L. Al Shalabi, and Z. Shaaban, "Normalization as a preprocessing engine for data mining and the approach of preference matrix", In *IEEE International Conference on Dependability of Computer Systems*, pp. 207-214, May 2006.