ISSN: 1978 - 8282

AN EXPERIMENTAL STUDY ON BANK PERFORMANCE PREDICTION BASE ON FINANCIAL REPORT

Chastine Fatichah[1]
Nurina Indah Kemalasari[2]

e-mail : chastine@cs.its.ac.id, nurina@cs.its.ac.id

Diterima : 2 Mei 2011/ Disetujui : 14 Juni 2011

ABSTRACT

This paper presents an experimental study on bank performance prediction base on financial report. This research use Support Vector Machine (SVM), Probabilistic Neural Network (PNN) and Radial Basis Function Neural Network (RBFN) methods to experiment the bank performance prediction. To improve accuracy prediction of both neural network methods, this research use Principal Component Analysis (PCA) to get best feature. This research work based on the bank's financial report and financial variables predictions of several banks that registered in Bank Indonesia. The experimental results show that the accuracy rate of bank performance prediction of PCA-PNN or PCA-RBFN methods are higher than SVM method for Bank Persero, Bank Non Devisa and Bank Asing categories. But, the accuracy rate of SVM method is higher than PCA-PNN or PCA-RBFN methods for Bank Pembangunan Daerah and Bank Devisa categories. The accuracy rate of PCA-PNN method for all bank categories is comparable to that PCA-RBFN method.

Keywords: bank performance prediction, support vector machine, principal component analysis, probabilistic neural network, radial basis function neural network

INTRODUCTION

The prediction of accuracy financial bank has been the extensively researched area since late. Creditors, auditors, stockholders and senior management are all interested in bankruptcy prediction because it affects all of them alike [7].

1. **Informatics Department, Faculty of Information Technology.**
   Institut Teknologi Sepuluh Nopember
   Gedung Teknik Informatika Lt. 2, Jl. Raya ITS Surabaya
2. **Informatics Department, Faculty of Information Technology.**
   Institut Teknologi Sepuluh Nopember
   Gedung Teknik Informatika Lt. 2, Jl. Raya ITS Surabaya

When the shareholders will make the investment to a bank, the shareholder must first see the performance of banks is good or not [2]. In some cases accurately predicted the performance of a bank can also through economic and financial ratio, the current assets / total assets, current assets - cash / total assets, current assets / loans, reserve / loans, net income / total assets, net income / total capital share, net income / loans, cost of sales / sales, cash flow / loan.

Some research [3] use neural network approach for performance predictions, neural network considered as an alternative network to predict accuracy that can result in the total value of the error more or less the same Error Type 1 and Error Type 2.

Type of error is determined from the predicted performance of the bank. Error Type 1 as the number of "actually poor performance banks" predicted as "adequate performance banks" expressed as percentage of total poor performance banks and Error Type 2 as the number of "actually adequate performance banks" predicted as "poor performance banks" expressed as a percentage of total adequate performance banks.

Ryu and yue researchers [9] introduce isotonik to predict the spread of a financial company and produce MLFF-BP, logistic regression, and probit methods. By using the data from one of a financial company, predicted the failure of small community banks and regional banks or big banks use MLFF-BP, MDA, and professional assessment. For community banks and regional banks, the researchers observed that the neural network model produce MDA model, especially type I error. The result is predicted in the small community bank is less accurate than the regional banks.

The objective of the this study is primarily to experiment several method of soft computing to analysis Error Type 1 and Error Type 2 of bank performance prediction.

The rest of the paper is organized as follows; Section 2 describes methods of bank performance prediction such as SVM, PNN, and RBFN methods. Section 3 describes experimental result and evaluation performance, and Section 4 describes conclusion of this research.

METHODS OF BANK PERFORMANCE PREDICTION

This section presents methods of bank performance prediction that used in this research. This research use Support Vector Machine (SVM), Probabilistic Neural

Network (PNN) and Radial Basis Function Neural Network (RBFN) methods. To improve accuracy prediction of both neural network methods, this research use Principal Component Analysis (PCA) to reduce the dimension of the input space. Each of the constituent methods is briefly discussed below.

Support Vector Machine

Support Vector Machine (SVM) [1] is a method for obtaining the optimal boundary of two sets in a vector space independently on the probabilistic distributions of training vectors. Its fundamental idea is locating the boundary that is most distant from the vectors nearest to the boundary in both of the sets. Note that the optimal boundary should classify not only the training vectors, but also unknown vectors in each set. Although the distribution of each set is unknown, this boundary is expected to be the optimal classification of the sets, since this boundary is the most isolated one from both of the sets. The training vectors closest to the boundary are called support vectors. Figure 1 illustrates the optimal boundary by SVM method.
Figure 1. Optimal Boundary by SV M method

The optimal boundary is computed as decision surface of the form:

$$f(x) = \text{sgn}(g(x)) \qquad (1)$$

where,

$$g(x) = \left(\sum_{i=1}^{l^*} y_i \alpha_i K(x, x_i^*) + b \right) \qquad (2)$$

In Equation 2, K is one of many possible kernel functions, $y_i \in \{-1,1\}$ is the class label of the data point x_i^*, and $\{x_i^*\}_{i=1}^{l^*}$ is subset of the training data set. x_i^* are called *support vectors* and are the points from the data set that fall closest to the separating hyper plane. Finally, the coefficients α_i and b are determined by solving a large-scale quadratic programming problem. The kernel function K that is used in the component classifier is a quadratic polynomial and has the form shown below:

$$K(x, x_i^*) = (x.x_i^* + 1)^2 \qquad (3)$$

$f(x) \in \{-1,1\}$ in equation (1) is referred to as the binary class of the data point x which is being classified by the SVM. Values of 1 and -1 refer to the classes of positive and the negative training examples respectively. As Equation (1) shows, the binary class of a data point is the sign of the raw output *g(x)* of the SVM classifier.

The raw output of a SVM classifier is the distance of a data point from the decision hyper plane. In general, the greater the magnitude of the raw output, the more likely a classified data point belongs to the binary class it is grouped into by the SVM classifier.

Principal Component Analysis

Principal component analysis (PCA) [5] has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis - from neuroscience to computer graphics - because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it.

Probabilistic Neural Network

The PNN [4] employs Bayesian decision-making theory based on an estimate of the probability density in the data space and Parzen estimates to make predictions. PNN requires onepass training and hence learning is very fast. However, the PNN works for problems with integer outputs, hence, can be used for classification problems. PNN is not stuck in some local minima of the error surface. The PNN as implemented here has 54 neurons in the input layer corresponding to the 54 input variables in the dataset. The pattern layer stores all the training patterns one in each pattern neuron. The summation layer has two neurons with one neuron catering to the numerator and another to the denominator of the non-parametric regression estimate of Parzen. Finally, the output layer has one neuron indicating the class code of the pattern.

Radial Basis Function

RBFN [4], another member of the feed-forward neural networks, has both unsupervised and supervised training phases. In the unsupervised phase, the input data are clustered and cluster details are sent to the hidden neurons, where radial basis functions of the inputs are computed by making use of the center and the standard deviation of the clusters. The radial basis functions are similar to kernel functions in kernel regression. The activation function or the kernel function can assume a variety of functions, though Gaussian radial basis functions are the most commonly used. The learning between hidden layer and output layer is of supervised learning type where ordinary least squares technique is used. As a consequence, the weights of the connections between the kernel layer (also called hidden layer) and the output layer are determined. Thus, it comprises a hybrid of unsupervised an

supervised learning. RBFN can solve many complex prediction problems with quite satisfying performance.

EXPERIMENTAL RESULT AND ANALYSIS PERFORMANCE

The first process of this system is collect financial report of banks that registered in Bank Indonesia about 110 banks. Data are taken as the period of 1 year, so that the amount of data used for 1320 data. From the amount of data, divided into 660 data for training data, and 660 data for testing data. There are six bank categories that used to evaluate bank performance prediction i.e. Bank Pembangunan Daerah, Bank Persero, Bank Devisa, Bank Non Devisa and Bank Asing.

Data are taken based on the variables from the financial report below:

1. *Earning asset*
2. *Total loans*
3. *Core deposit*
4. *Non-interest*
5. *Interest income*
6. *Gain(losses)*
7. *Non-interest expense-wages and salary*
8. *Total interest expense*
9. *Provision expense*
10. *Off balance sheet commitment*
11. *Obligation and letter of credit*

The second process of this system identifies bank financial variables that would be used to classify data. There are two bank financial variables such good and poor variables.

The third process of this system is bank performance prediction using SVM, PCA-PNN, and PCA-RBFN methods to produce Error Type 1, Error Type 2 and accuracy. To evaluate Error Type 1, Error Type 2 and accuracy for each method used testing data of each bank category. PCA method is used to get best feature of dataset before is classified by PNN or RBFN. Base on several references that PCA is one of the best methods for reducing attribute of dataset but not lost important information of data.

The experiment results of one bank sample of Bank Pembagungan Daerah categories on 1 year (Table 1) show that Error Type 1 of PCA-RBFN method is lowest and the accuracy of SVM method is highest.

The experiment results of one bank sample of Bank Persero categories on 1 year (Table 2) show that Error Type 1 of SVM method is highest and the accuracy of SVM method is lowest.

The experiment results of one bank sample of Bank Devisa categories on 1 year (Table 3) show that the accuracy of SVM method is highest.

The experiment results of one bank sample of Bank Non Devisa categories on 1 year (Table 4) show that Error Type 1 of SVM method is highest and the accuracy of SVM method is lowest.

The experiment results of one bank sample of Bank Asing categories on 1 year (Table 5) show that the accuracy of SVM method is lowest.

The experimental results show that the accuracy rate of bank performance prediction of PCA-PNN or PCA-RBFN methods are higher than SVM method for Bank Persero, Bank Non Devisa and Bank Asing categories. But, the accuracy rate of SVM method is higher than PCA-PNN or PCA-RBFN methods for Bank Pembangunan Daerah and Bank Devisa categories. The accuracy rate of PCA-PNN method for all bank categories is comparable to that PCA-RBFN method.

Tabel 1. The result of Bank Pembangunan Daerah category

Methods	Error and Accuracy name	Percentages of error and accuracy
SVM	Error Type 1 Error Type 2 Accuracy	16.67% 8.33% **75%**
PCA-RBFN	Error Type 1 Error Type 2 Accuracy	**0%** 33.33% 66.67.%
PCA-PNN	Error Type 1 Error Type 2 Accuracy	25% 8.33% 66.67.%

Table 2. The result of Bank Persero category

Methods	Error and Accuracy name	Percentages of error and accuracy
SVM	Error Type 1 Error Type 2 Accuracy	25% 16.67% 58.33%
PCA-RBFN	Error Type 1 Error Type 2 Accuracy	0% 16.67% **83.33%**
PCA-PNN	Error Type 1 Error Type 2 Accuracy	0% 16.67% **83.33%**

Table 3. The result of two bank sample of Bank Devisa category

Methods	Error and Accuracy name	Percentages of error and accuracy
SVM	Error Type 1 Error Type 2 Accuracy	16.67% 0% **83.33%**
PCA-RBFN	Error Type 1 Error Type 2 Accuracy	0% 25% 75%
PCA-PNN	Error Type 1 Error Type 2 Accuracy	16.67% 8.33% 75%

Table 4. The result of two bank sample of Bank Non Devisa category

Methods	Error and Accuracy name	Percentages of error and accuracy
SVM	Error Type 1 Error Type 2 Accuracy	16.67% 16.67% 66.67%
PCA-RBFN	Error Type 1 Error Type 2 Accuracy	0% 16.67% **83.33%**
PCA-PNN	Error Type 1 Error Type 2 Accuracy	0% 16.67% **83.33%**

Table 5. The result of two bank sample of Bank Asing category

Methods	Error and Accuracy name	Percentages of error and accuracy
SVM	Error Type 1 Error Type 2 Accuracy	8.33% 16.66% 75%
PCA-RBFN	Error Type 1 Error Type 2 Accuracy	16.66% 0% **83.3333%**
PCA-PNN	Error Type 1 Error Type 2 Accuracy	0% 16.66% **83.3333%**

CONCLUSION

This paper presents bank performance prediction using SVM, PCA-PNN, and PCA-RBFN methods. The PCA is used to get best feature of dataset before is classified by PNN or RBFN. The experimental results show that the accuracy rate of bank performance prediction of PCA-PNN or PCA-RBFN methods are higher than SVM method for Bank Persero, Bank Non Devisa and Bank Asing categories. But, the accuracy rate of SVM method is higher than PCA-PNN or PCA-RBFN methods for Bank Pembangunan Daerah and Bank Devisa categories. The accuracy rate of PCA-PNN method for all bank categories is comparable to that PCA-RBFN method. This research can be used to evaluate bank performance in real cases.

REFERENCE

[1]. A. Asano, " Pattern information processing", Session 12 (05. 1. 21), 2004.
[2]. "Bank".(http://en.wikipedia.org/wiki/Bank).
[3]. K.Y. Tam, M. Kiang, Predicting bank failures: a neural network approach, Decis. Sci. 23 (1992) 926–947.
[4]. Simon Haykin, Neural Network: A Comprehensive Foundation.(2005) 240-301.
[5]. Smith, Lindsay I.26 Pebruari 2006."*A Tutorial on Principal Component Analysis*".
[6]. Support Vector Machines, available at http://www.svms.org/introduction.html; libSVM tool can be downloaded from http://www.csie.ntu.edu.tw/-cjlin/libsvm.
[7]. V.Ravi, H.Kurniawan, Peter Nwee Kok Thai, dan P.Ravi Kumar, "Soft computing system for bank performance prediction", IEEE Journal, February 2007.
[8]. Y.U. Ryu, W.T. Yue, Firm bankruptcy prediction: experimental comparison of isotonic separation and other classification approaches, IEEE Trans. Syst., Manage. Cyber.—Part A: Syst. Hum. 35 (5) (2005) 727–737