FINANCIAL PREDICTION MODELS FROM INTERNAL AND EXTERNAL FIRM FACTORS BASED ON COMPANIES LISTED ON THE STOCK EXCHANGE OF THAILAND

Janthorn Sinthupundaja[1], Navee Chiadamrong[2*], and Nathridee Suppakitjarak[3]

Received: December 19, 2016; Revised: February 07, 2017; Accepted: February 15, 2017

Abstract

In this study, 3 prediction methods using statistical and machine learning techniques, namely logistic regression, artificial neural network (ANN), and support vector machine are compared to classify a company's financial performance in relation to the average return on assets of all the companies listed on the Stock Exchange of Thailand in each year. In total, there are 1968 firm-year observations for the period from 2005 to 2014. Our estimated models use a combination of internal firm factors including firm characteristics and financial indicators, and external firm factors from political, economic, social, and technological aspects as indicators for changes in the macro-economic environment. The results suggest that the ANN outperforms the other techniques by achieving 71.85% accuracy rates. With these prediction models, managers and decision makers can predict a firm's financial performance more accurately and can keep track of the performance 1 year in advance which helps to identify the firm's future business trends; these are very important factors for decision makers as to whether or not they should take necessary action to improve their firm's performance.

Keywords: **Thai public companies, financial performance, prediction models, internal firm factors, external firm factors**

[1] *School of Manufacturing Systems and Mechanical Engineering, Sirindhorn International Institute of Technology (SIIT), Thammasat University, Khlong Luang, Pathum Thani 12120, Thailand. E-mail: janthorns@gmail.com*
[2] *School of Manufacturing Systems and Mechanical Engineering, Sirindhorn International Institute of Technology (SIIT), Thammasat University, Khlong Luang, Pathum Thani 12120, Thailand. E-mail: navee@siit.tu.ac.th.*
[3] *Faculty of Commerce and Accountancy, Chulalongkorn University, Phayathai Road, Pathum Wan, Bangkok 10330. E-mail:* nathridee@cbs.chula.ac.th
[*] *Corresponding author*

Introduction

Thai public companies or firms have faced challenges in recent years due to internal and external pressures. Internal challenges have arisen from inefficient management and lack of useful information for making decisions while external challenges have come from the uncertainty of the world economy as well as the Thai economy, fluctuation of currencies that can affect import and export activities, and more challenging competitors. Development of the Thai economy and the financial markets has led to an increasing number of public companies on the Stock Exchange in Thailand (SET). According to the record provided by the SET (2017), the number of public companies increased from 404 to 537 from 2005 to 2014. Market capitalization of the SET dramatically increased from 5.11million baht in 2005 to 13.86 million baht in 2014. These numbers emphasize that public companies play an important role in the Thai economy. Failure of public companies has serious negative consequences from both economic and social impacts.

Numerous past studies have focused on failure predictions of firms (Zhang *et al*., 1999; Bose and Pal, 2006; Boyacioglu *et al*., 2009; Delen *et al*., 2013; Tinoco and Wilson, 2013). Their common purpose is to classify bankrupt firms from non-bankrupt firms. However, it is also very important to study financial performance among non-bankrupt firms. A firm should know what its financial health level is compared to others in the same market. Evaluating a firm's performance is considered a powerful tool for helping decision-makers to make their decisions in doing business.

In the past, many classification models were based only on statistical tools such as logistic regression and linear discriminant analysis. Explanatory and prediction models are developed using a given number of determinants. However, there are many limitations to statistical models, such as low prediction accuracy rates and the requirements of several assumptions. Lately, classification analysis of machine learning techniques has been developed and applied into financial and business fields. Accuracy rates of machine learning techniques have been reported to be better than the accuracy rates of traditional statistical methods from much empirical research (Tam and Kiang, 1992; Shin *et al*., 2005; Bose and Pal, 2006; Youn and Gu, 2010; Ogut *et al*., 2012; Mahajan *et al*., 2014).

In this study, we select the artificial neural network (ANN) and support vector machine (SVM) machine learning techniques, in comparison with the classical statistical classification technique of logistic regression. In addition, a cross-validation technique is also employed to investigate the robustness, and to access the accuracy and validity of a model. The objective of this study is to develop prediction models for predicting a firm's financial performance, which is the return on assets (ROA). Firms can use the models to predict their financial performance in comparison with other public companies in the market. Managers and related decision makers can benefit from knowing the financial positions of their firms in the coming year. So, important decisions and appropriate strategies can be adapted to improve the current financial performance and importantly prevent failure or bankruptcy. Firms with a good financial status can maintain themselves or find investment opportunities while firms with a bad financial status have to improve their financing, operating, and investing activities to survive in a competitive market.

The paper is structured as follows. Section 2 discusses the research background and literature review that are relevant to the modelling approach. Section 3 provides the methodology of the study including data, classifying outcome variables, and independent variables. Section 4 describes experimental conditions. Section 5 presents empirical results. Finally, Section 6 states the conclusions, limitation, and further studies.

Background

Studies of a firm's financial and business performance predictions in the past dealt with various problem statements such as bankruptcy prediction, performance prediction, business profitability, and credit ratings. One of the most popular objectives is to develop bankruptcy prediction models from statistical methods or machine learning techniques (Zhang *et al.*, 1999; Bose and Pal, 2006; Boyacioglu *et al.*, 2009; Ray, 2011; Malhotra and Mukherjee, 2013; Xu *et al.*, 2014; Ciampi, 2015). For instance, Tinoco and Wilson (2013) studied the financial distress and bankruptcy prediction models for listed companies using accounting, market, and macroeconomic variables. The dataset was drawn from the 2012 London Share Price Database. The study developed risk models for listed companies that predict financial distress and bankruptcy. The models were benchmarked against models from the neural network models in Altman's study (1968). Tam and Kiang (1992) studied the bank bankruptcy prediction model by comparing the statistical analysis with the neural network models. Their results showed that neural network models are generally more accurate than other statistical analyses in evaluating a bank's status.

However, the study of a firm's financial health while it is still operating under non-bankruptcy conditions is also an interesting issue, as a firm can prepare itself for upcoming situations. Lam (2004) investigated the abilities of neural networks, integrating fundamental and technical analysis, to predict financial performance represented by the rate of return on common shareholders' equity. The analysis was based on a sample of 365 Standard & Poor's companies from 1985 to 1995. In order to compensate for data noise and parameter error, as well as to study predictive models and procedures, an extraction technique was applied to convert the connection weights from trained neural networks to symbolic classification rules. The determinants of financial performance in the study include 16 financial ratios and 11 macro-economic variables. The neural network models were compared with the top one-third returns in the market (maximum benchmark) and overall market average return (minimum benchmark). The results showed that the neural networks outperformed the market returns in some of the cases. Delen *et al.* (2013) developed firm performance models with a 2-step analysis methodology, including exploratory factor analysis to identify the dimensions of the financial ratios, and 4 decision tree algorithms to analyse prediction models, based on the central tendency measure (median) values of the ROA and return on equity (ROE) as a split criterion.

Analysis Techniques

As mentioned, this study uses 3 statistical and machine learning techniques for a firm's financial performance prediction. These techniques can be explained as follows:

Logistic Regression

Logistic regression, or logit regression, or the logit model, is commonly used in business and the financial literature (Youn and Gu, 2010). It overcomes the methodological difficulties associated with discriminants analysis. Logistic regression is an approach used for estimating the probability and group membership of independent variables by making logistic transformation of the linear combination of dependent variables (Ogut *et al.*, 2012). Using the maximum-likelihood method, it attempts to build a regression model that best describes group membership (Lussier, 1995). The advantages of logistic regression are that it takes the form of a non-linear regression equation, and variables in the models can be explained (Youn and Gu, 2010). Equation (1) presents a logistic regression model:

$$Log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n \; ; \qquad (1)$$

where: β_0 = the intercept; β_1 to β_n = estimated coefficients; X_1 to X_n = independent variables; and p is the probability that the event will occur, $1 - p$ is the probability that the event will not occur, and n is the total number of variables, as follows in Equation 2:

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + ... + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + ... + \beta_n X_n}} \qquad (2)$$

Artificial Neural Network (ANN)

An ANN is capable of processing vast amounts of data and making predictions that are sometimes surprisingly accurate. It is a flexible, nonparametric modelling tool that does not require prior assumptions regarding the data distribution or the relationship between the variables (Wu *et al.*, 2006). Neural networks were developed from simulating biological nervous systems with the use of computers. Neural network models can be found extensively in applications in financial services including bankruptcy prediction and performance prediction (Tam and Kiang, 1992; Zhang *et al*., 1999; Bose and Pal, 2006; Pao, 2008; Mokhatab Rafiei *et al*., 2011; Delen *et al*., 2013).

A popular form of ANN is called the multi-layer perception neural network in which all nodes and layers are arranged in a feedforward manner (Zhang *et al*., 1999). The pre-dominance of the supervised feedforward multilayer perceptron trained by back-propagation (BP), which is the most popular algorithm, is basically a gradient steepest descent method with a constant step size (Rumelhart *et al*., 1986). The BP neural network consists of an input layer, an output layer, and 1 or many intervening layers (or hidden layers), as shown in Figure 1 (Pao, 2008). The aim of the network training is to minimize the differences between the neural networks and the known target values for all training patterns (Rich and Knight, 1991; Zhang *et al*., 1999; Roiger and Geatz, 2003; Russell and Norvig, 2003).

Neural network model and parameters:

N = number of input nodes
M = number of hidden nodes
J = number of output nodes
x_n = inputs at the n^{th} node
s_m = sum of output at the m^{th} node
y_m = outputs from the m^{th} hidden node after the activation function
v_j = sum of output at the j^{th} node
z_j = outputs from the j^{th} output node after the activation function
z_m = outputs from the m^{th} hidden node after the activation function
t_j = target values from output layer at the j^{th} node
w_{nm} = connection weight between the n^{th} input node and m^{th} hidden node
w_{mj} = connection weight between the m^{th} hidden node and j^{th} output node
e_j = error values at the j^{th} output node
e_m = error values at the m^{th} hidden node
b_m = a bias at the m^{th} hidden node
b_j = a bias at the j^{th} node
η = learning rate between 0 and 1

Initially, small values are randomly assigned to all connection weights (w_{mj} and w_{nm}). Then, outputs from the hidden nodes (s_m) are calculated with Equation (3) and adjusted with the activation function. In this study, the sigmoid function, which is the most popular activation function as suggested by Basterretxea *et al.* (2004), is employed as shown in Equation (4) and Figure 2. Next, outputs (y_m) from the m^{th} hidden node are computed by Equation (5). Then, outputs (v_j) from the j^{th} hidden node are calculated with Equation (6) and the value is adjusted with the sigmoid function. Similarly, outputs (z_j) from the j^{th} hidden node are calculated and adjusted by the sigmoid function in Equation (7). The network error at the j^{th} node (e_j) and the sum of the error from the hidden nodes (e_m) are calculated from Equation (8) and Equation (9), respectively. The connection weights of the

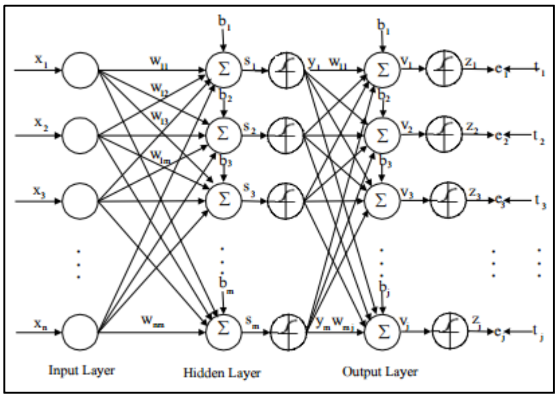

Figure 1. Multilayer perceptron neural network

network (w_{mj} and w_{nm}) are adjusted by Equation (10) and Equation (11), respectively.

$$s_m = \sum_{n=1}^{N} x_n w_{nm} \quad (3)$$
$$f(x) = \frac{1}{1+e^{-x}} \quad (4)$$
$$y_m = f(s_m) \quad (5)$$
$$v_j = \sum_{m=1}^{M} y_m w_{mj} \quad (6)$$
$$z_j = f(v_j) \quad (7)$$
$$e_j = z_j(1-z_j)(t_j - z_j) \quad (8)$$
$$e_m = y_m(1-y_m)\sum_{j=1}^{J} e_j w_{mj} \quad (9)$$
$$w_{mj} = w_{mj} + \eta e_j z_j \quad (10)$$
$$w_{nm} = w_{nm} + \eta e_m z_m \quad (11)$$

Once the network completes its training, its weights and thresholds are determined. Thus, a calculation analysis can be started.

Support Vector Machine (SVM)

The SVM has been a popular tool for financial decision making and predictive modelling (Shin *et al*., 2005; Ogut *et al*., 2012). It is a classification, recognition, regression, and time series technique (Boyacioglu *et al*., 2009). The SVM uses a linear model to implement non-linear class boundaries by mapping input vectors nonlinearly into a high-dimensional feature space (Vapnik, 1998; Kumar and Ravis, 2007). The SVM produces a binary classifier which optimally separates hyperplanes through non-linear mapping of the input vectors into a high-dimensional feature space. The SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors. If the data is linearly separated, the SVM trains linear machines for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points (Vapnik, 1998; Shin *et al*., 2005). The linear classifier has the simple form of $f(x) = W^T x + b$, where W is the weight vector, T is the transpose notation of weight, x is the training data, and b is the bias. The training examples that are closest to the maximum margin hyperplane are called support vectors (Lee, 2007). If the data are not linearly separated, the SVM uses non-linear machines to find a hyperplane that minimizes the number of errors for the training set.

Materials and Methods

Data

Data were collected from all public companies listed on the SET during the period from 2005 to 2014. Both internal firm factors and external firm factors were taken into the consideration. Internal firm factors were drawn from income statements and balance sheets, which were obtained from Chulalongkorn University (Datastream system), Maruey Library, the SET, and the Securities and Exchange Commission, while the data of external firm factors were gathered from the Bank of Thailand (2017). In total, there are 8 industries listed in the SET, comprising agro and food, consumer products, industrials, property and construction, resources, technology, finance, and services. The finance industry was excluded from the analyses, since the financial industry structure and its nature are different from the other industries, because the main

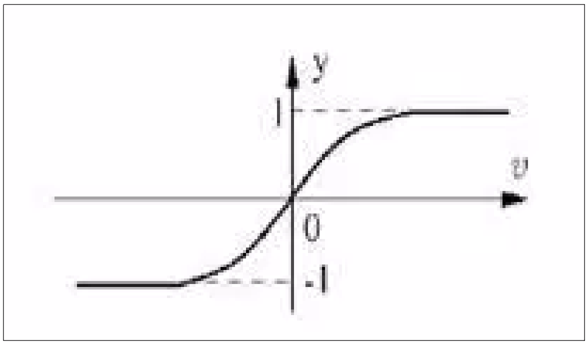

Figure 2. Log-sigmoid transfer function

role of the financial industry is to provide financial advice and services to customers in the form of deposits, investments, loans, and insurance. There is no manufacturing or inventory holding of typical assets as appears in the other industries. Therefore, different financial performance indicators are required. The entire dataset contains 5,874 observations (in 11 years). However, many firms are excluded because essential data are missing due to mergers, suspensions, and bankruptcies. As a result, 1968 observations (in 11 years) remain in the analyses. Table 1 shows the list of industries and their number of observations.

Classifying a Dependent Variable

Performance is the function of an ability of an organization to gain and manage resources in several different ways to develop competitive advantages (Iswatia and Anshoria, 2007). It also refers to the results of undertaking activities. High performance reflects efficiency of management and resource usage (Naser and Mokhtar, 2004). Financial performance is used as a reflection of financial records and reports, in order to determine the efficiency and operation of economic unity management (Khalifa and Shafii, 2013). Firm performance is commonly represented by the ROA (Alzharani *et al.*, 2011; Doan and Nguyen, 2011; Almajali *et al.*, 2012; Pantea *et al.*, 2014), which measures the effectiveness of economic unity to utilize the assets and efficiency in generating profits, and is calculated by net income divided by total assets.

In this study, the ROA is also employed as a split criterion to reflect the performance level of the firms in the market. A firm is classified as having a financially above average performance when its ROA is above the average ROA of listed companies while a firm is classified as having a financially below average performance when its ROA is below the average ROA of the listed companies. As a result, 565 firm-years are classified as financially above average performance firms and 1403 firm-years are classified as financially below average.

Independent Variables

Internal Firm Factors

There are a final 21 internal firm variables, which are classified into 8 factors, comprising size, age, leverage, liquidity, equity, cash flow, sales, and expenses in the analyses. These independent factors involve the firms' characteristics and financial indicators. In fact, greater numbers of variables in each factor were initially included in the analyses, but some of them were excluded during the preliminary experiment as it was found that there were high correlations among these variables with some of the remaining variables, meaning that they could represent the same construct and direction. The excluded variables include current assets, shareholders' equity, total liabilities, and retained earnings. Table 2 presents the list of independent variables from the firms' internal factors. The firms' characteristics and financial indicators remain in the analyses.

Firm characteristic factors describe the status of firms from their physical appearances. The factors of age and size are considered to be important factors to represent firm characteristics and are often used to be the determinants of firm performance (Hovakimian *et al.*, 2001; Agustinus and Rachmadi, 2008; Almajali *et al.*, 2012; Naidu and Chand, 2013; Sharma and Raina, 2013). Age implies the

Table 1. List of industries and their number of observations

Industries	Number of observations
Agro & food	280
Consumer products	214
Industrials	378
Property & construction	320
Resources	161
Services	451
Technology	164
Total	**1968**

basic aspects of a firm. The impact of age on the firm's performance is still very much in doubt, in which conflicting results have often been reported. Age is considered as an important factor towards a firm's financial performance. Age can positively affect the financial performance since older firms have more experience and enjoy superior performance (Almajali *et al*., 2012). However, age can affect a firm's financial performance negatively. For example, Evans (1987) and Loderer and Waelchli (2010) showed that a firm's financial performance declines with age since old firms may become inefficient. In this study, the factor of age is represented by both the established age and listed age in the market. Firms of different sizes can apply different financial strategies, which can affect their firms' performances. Large firms are more likely to have more access to formal credits and internalize many of the capital allocation functions carried out by financial markets and financial intermediaries. In contrast, small firms are likely to face tougher obstacles in obtaining financing, accessing legal systems, or dealing with corruption (Biesebroeck, 2005). In this study, the firm's size is represented by its market value over total assets. Market value is used to refer to the market capitalization of a firm. It is commonly used to represent the size of a firm (Yang *et al*., 2010). We divided market value with total assets to eliminate the bias due to the different sizes of a firm's assets.

Financial indicators are the current financial status of firms, showing how firms manage their resources for achieving their goals. In this study, financial indicators include the factors of liquidity, equity, cash flow, sales, level of leverage, and expenses. Liquidity is considered as one of the important determinants of firms' financial performance (Bagchi, 2015), since, good liquidity management helps firms to maintain a surplus, reduce risk, and improve company survival rate (Pratheepkanth, 2011). In addition, liquidity would allow a firm to deal with unexpected contingencies and to cope with its obligations during periods of low earnings (Khalifa and Shafii, 2013). The liquidity is represented by

Table 2. List of firm characteristics and financial indicators factors

Groups	Factors	Variables	
Firm characteristics	Age	Established age	X_1
		Listed age	X_2
	Size	Market value/total assets	X_3
Firm financial indicators	Liquidity	Quick ratio: (current assets - total inventories)/current liabilities	X_4
		Inventory turnover ratio: Sales/inventories	X_5
		Current ratio: Current assets/current liabilities	X_6
		Income/current liabilities	X_7
		Cash/current liabilities	X_8
	Equity	Earnings per share	X_9
	Cash flow	Cash flow from investing activities	X_{10}
		Cash flow from operating activities	X_{11}
		Cash flow from financing activities	X_{12}
		Cash	X_{13}
	Sales	Sales/total assets	X_{14}
		Sales/current liabilities	X_{15}
		Sales/equity	X_{16}
		Sales/expenses	X_{17}
	Leverage	Debt/total assets	X_{18}
		Debt/equity	X_{19}
	Expenses	Cost of goods sold	X_{20}
		Depreciation	X_{21}

the quick ratio (current assets minus total inventories over current liabilities), inventory turnover ratio (sales over inventories), current ratio (current assets over current liabilities), income over current liabilities, and cash over current liabilities. Equity affects investor decisions and financial funds which affect a firm's activities. This causes a relationship between equity and the firm's performance. The indicator of equity is earnings per share, which is the firm's profits that are allocated to the shares of common stock.

Cash flow is also important since it affects a firm's management strategies. Holding too much cash prevents a firm from maximizing the profit by investing cash on activities. However, too little cash holding is also risky, in that a firm will not be able to meet its short-term obligations. The factor of cash flow consists of the amount of cash holding and cash flow from investing, operating, and financing activities. The factor of sales represents the sales activities, which is sales over total assets, sales over current liabilities, sales over equity, and sales over expenses. Regarding the level of leverage, there are 2 opposing effects on the level of leverage depending on a firm's financial strategies. A high level of leverage could improve the firm's performance because managers are under pressure and need to make value-maximizing decisions. On the other hand, a high level of debt brings high financial cost. If borrowed debt cannot generate sufficient income or the firm cannot utilize loans well enough, this leads to lower the firm's financial performance (Zeitun and Tian, 2007; Liargovas and Skandalis, 2010). The factor of leverage is represented by debt over total assets and debt over equity. Expenses are also important to firms and directly affect a firm's profit. The factor of expenses is represented by the cost of goods sold and depreciation expenses on fixed assets. These expenses cover the expenses of both producing the products and services as well as acquiring fixed assets for doing business

External Firm Factors (or Macro-Economic Factors)

In addition to the firm characteristics and financial indicators, 4 external firm factors were selected following 4 aspects from the PEST (political, economic, social, and technological) analysis. The first macro-economic variable is corporate tax rate, representing political aspects. Political issues could directly or indirectly affect a firm's activities and performances. An increase or decrease in corporate tax rate affects management decisions and strategies since it links directly to cost and profit. The second macro-economic variable is gross domestic product (GDP), representing the economic aspects. GDP is a primary indicator to measure the strength of a national economy. Firm performance tends to be better in a good economic environment. The third macro-economic variable is the number of births, representing the social aspects. The number of births is an indicator of social conditions, showing the stability of society. There is evidence that the birth rate can be high when the social conditions are suitable. Birth rate also reflects social culture, lifestyle, consumer needs, and economic growth (Weintraub, 1962). In addition, Galbraith and Thomas (1941) reported that the connection between birth rates and business cycles is strong. The fourth macro-economic variable is gross domestic expenditure on research and development as a percentage of GDP, representing the technological aspect. It is the amount the government spends on research and development in Thailand, which can directly reflect the amount of investment in technology (Table 3).

Table 3. List of external factors

Factors	Variables	
Political aspects	Corporate tax rate	X_{22}
Economic aspects	Gross Domestic Product (GDP)	X_{23}
Social aspects	Number of births	X_{24}
Technological aspects	Gross Domestic Expenditure on R&D as a Percentage of GDP	X_{25}

Accuracy Rates of the Results

The prediction performance of the models is evaluated in the form of the percentage of accuracy by looking at the rates to which the output of a model matches the actual value for the corresponding input values, in both the training and test data sets. The accuracy rate is the proportional sum of observations that is correctly predicted over the total number of observations. The higher the rates, the better the performance of the model's predictions. For example, if the model predicts that a firm is classified as financially above the average (y = 1) and the firm is actually financially above the average (y = 1), the percentage of accuracy of this firm in a given year is 100% and vice versa. A summation of these percentages of all observations divided by the total number of observations presents the accuracy rate of the model.

Experimental Conditions

Our prediction models aim to predict the financial health performance of companies listed on the SET 1 year in advance. Companies identified as having a financially above average performance were given a value of 1 and those identified as having a financially below average performance were given a value of 0. Firm financial performance in year t^{th} is predicted by the independent variables in year $t\text{-}1^{th}$. The 10-fold cross validation is employed to help minimize the bias associated with the selecting of the training and test data (Zhang *et al.*, 1999). As a result, a data set is divided into 10 subsets followed by the year. One of the 10 subsets is for training and the other 9 subsets are for testing. Training and test processes are repeated 10 times until all subsets of data have been trained and tested. The cross validation estimate of the overall accuracy of a model is calculated by averaging all 10 individual accuracy measures. Prior to the analysis, all data are standardised by subtracting the average value, and then divided by the standard deviation in order to eliminate the bias of scales regarding the effects of coefficient factors. Therefore, the mean and standardized deviation of the standardized data are 0 and 1, respectively. Classification methods, including a statistical technique (logistic regression) and machine learning techniques (the ANN and SVM), are applied and compared to each other, using the prediction accuracy rates of both the training and test data sets.

Results and Discussion

Logistic Regression

Using SPSS 17.0, the models' testing suggests eliminating variables, including established age (X_1), market value/total assets (X_3), sales/inventories (X_5), current assets/ current liabilities (X_6), income/current liabilities (X_7), earnings per share (X_9), cash flow from investing activities (X_{10}), cash flow from financing activities (X_{12}), cash (X_{13}), sales/total assets (X_{14}), sales/current liabilities (X_{15}), sales/equity (X_{16}), debt/equity (X_{19}), corporate tax rate (X_{22}), and number of births (X_{24}), in order to improve the models' fitness. After

Table 4. Hosmer and Lemeshow Test of logistic regression model

Test year	Chi-square	Degree of freedom	*p-value*
2005	15.04	8	0.058
2006	10.101	8	0.258
2007	15.045	8	0.058
2008	15.205	8	0.055
2009	7.180	8	0.517
2010	9.646	8	0.291
2011	13.554	8	0.094
2012	14.888	8	0.061
2013	11.936	8	0.154
2014	10.755	8	0.216

elimination of the above variables, the following results can be presented. The Hosmer and Lemeshow goodness of fit test is used to indicate a good fitting model when its value is greater than 0.05 since it fails to reject the null hypothesis (there is no difference between the observed and model-predicted values), implying that the model's estimates fit the data at an acceptable level. Our Hosmer and Lemeshow goodness of fit tests for all testing years are greater than 0.05, as shown in Table 4, which indicates the acceptance of the hypothesis and therefore our models are considered good fits.

Equations 12 to 21 present the logistic regression models from 2005 to 2014, respectively, where p is the probability that firm performance is financially above average in the market. The results show that independent variables are significantly different in each testing year. For example, in 2005, there are 6 significant variables - listed age (X_2), cash flow from operating activities (X_{11}), sales/expenses (X_{17}), debt/total assets (X_{18}), depreciation (X_{21}), and gross domestic expenditure on research and development as a percentage of GDP (X_{25}), while there are only 4 significant variables in 2006, - listed age (X_2), cash flow from operating activities (X_{11}), sales/expenses (X_{17}), and debt/total assets (X_{18}).

$$p = \frac{e^{1.51 + 4.048\,X2 + 0.274\,X11 + 0.221\,X17 + 0.461\,X18 - 0.087\,X21 - 0.05\,X25}}{1+e^{1.51 + 4.048\,X2 + 0.274\,X11 + 0.221\,X17 + 0.461\,X18 - 0.087\,X21 - 0.05\,X25}}$$

for year 2005 (12)

$$p = \frac{e^{1.4481 - 4.022\,X2 - 0752\,X11 - 0.184\,X17 - 0.402\,X18}}{1+e^{1.4481 - 4.022\,X2 - 0752\,X11 - 0.184\,X17 - 0.402\,X18}}$$

for year 2006 (13)

$$p = \frac{e^{1.525 - 4.121\,X2 - 0.785\,X11 + 0.187\,X17 + 0.413\,X18}}{1+e^{1.525 - 4.121\,X2 - 0.785\,X11 + 0.187\,X17 + 0.413\,X18}}$$

for year 2007 (14)

$$p = \frac{e^{1.664 - 4.981\,X2 + 0.763\,X11 + 0.182\,X17 + 0.421\,X18}}{1+e^{1.664 - 4.981\,X2 + 0.763\,X11 + 0.182\,X17 + 0.421\,X18}}$$

for year 2008 (15)

$$p = \frac{e^{1.543 - 3.885\,X5 + 0.823\,X14 + 0.125\,X20 + 0.358\,X21}}{1+e^{1.543 - 3.885\,X5 + 0.823\,X14 + 0.125\,X20 + 0.358\,X21}}$$

for year 2009 (16)

$$p = \frac{e^{1.563 - 3.911\,X2 + 0.854\,X11 - 0.346\,X18}}{1+e^{1.563 - 3.911\,X2 + 0.854\,X11 - 0.346\,X18}}$$

for year 2010 (17)

$$p = \frac{e^{4.314 - 0.551\,X11 + 0.395\,X18}}{1+e^{4.314 - 0.551\,X11 + 0.395\,X18}}$$

for year 2011 (18)

$$p = \frac{e^{1.419 + 3.903\,X2 - 1.151\,X4 + 0.584\,X11 + 0.240\,X17 + 0.397\,X18}}{1+e^{1.419 + 3.903\,X2 - 1.151\,X4 + 0.584\,X11 + 0.240\,X17 + 0.397\,X18}}$$

for year 2012 (19)

$$p = \frac{e^{1.496 + 3.763\,X2 + 0.764\,X11 + 0.174\,X17 + 0.375\,X18 + 3.155\,X11 + 0.171\,X17 + 0.527\,X18}}{1+e^{1.496 + 3.763\,X2 + 0.764\,X11 + 0.174\,X17 + 0.375\,X18 + 3.155\,X11 + 0.171\,X17 + 0.527\,X18}}$$

for year 2013 (20)

$$p = \frac{e^{1.547 + 4.947\, X2 + 0.199\, X11 + 0.257\, X17 + 0.707\, X25}}{1 + e^{1.547 + 4.947\, X2 + 0.199\, X11 + 0.257\, X17 + 0.707\, X25}}$$
for year 2014 (21)

The prediction accuracy rates of the logistic regression for all testing years from 2005 to 2014 are shown in Table 5. The training data set obtains the highest accuracy rate of 74.46% in 2010 while the test data set is at 79.71% in 2011. The accuracy rates of the training data set are higher than the test data set for almost every testing year, except 2011 and 2012.

Overall, the average accuracy rates of the prediction models for the training and test data sets are 73.56% and 68.42%, respectively.

Artificial Neural Network (ANN)

The ANN models were built and analysed using MATLAB R2014a with back-propagation of the Levenberg-Marquardt learning process. The analysis is based on a 3-layer ANN: 1 input, 1 hidden, and 1 output layer. The number of hidden nodes was varied from 1 to 13 hidden nodes. Table 6 presents the accuracy rates of the ANNs during 2005 to 2014 for all testing years. The average values show that neural classifiers with 3 hidden nodes produce the highest average classification accuracy rates for the test data, with average accuracy rates of 71.36% and 71.85% for the training and test data sets, respectively.

Support Vector Machine (SVM)

The models of the SVM in this study were analysed via LibSVM, which is an open source machine learning library, developed by the National Taiwan University (Chang and Lin, 2011). In the models, the radial basis function with kernel is employed. Table 7 presents the accuracy rates of the SVM for all testing years. The result shows that the model in the 2014 test year predicts with the highest accuracy for both the training and test data sets, which are 72.50% and 86.67%, respectively. The average accuracy rates for all testing years are 64.80% and 71.28% for the training and test data sets, respectively.

Classification Summary

Table 8 summarizes the average accuracy rates of the 3 prediction models from all testing years. The highest average accuracy rate was predicted by the ANN at a rate of 71.85%, followed by the SVM and logistic regression at 71.27% and 68.42%, respectively, in the test set. It can be seen that machine learning techniques (both the ANN and SVM) outperform the logistic regression. Although the results are slightly different, this would affect the firms' decisions. However, logistic regression predicts with the highest accuracy

Table 5. Result of logistic regression

Test year	Training	Test
2005	73.60%	70.91%
2006	72.68%	59.74%
2007	73.85%	68.70%
2008	74.11%	65.92%
2009	74.44%	65.13%
2010	**74.46%**	**59.24%**
2011	**72.90%**	**79.71%**
2012	72.78%	73.81%
2013	73.90%	70.37%
2014	72.92%	70.67%
Average	**73.56%**	**68.42%**

Table 6. Results of artificial neural networks

Hidden node(s)	Test year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
1	Training	71.74%	70.93%	70.31%	71.21%	72.42%	72.78%	71.66%	70.59%	71.50%	72.77%	71.59%
	Test	70.91%	71.51%	75.00%	71.51%	61.03%	60.51%	68.12%	78.57%	68.25%	66.00%	69.14%
2	Training	72.14%	71.27%	71.32%	70.26%	72.42%	72.28%	72.06%	70.25%	71.67%	72.83%	71.65%
	Test	70.91%	71.51%	72.50%	72.07%	61.03%	57.96%	71.01%	80.00%	68.25%	66.67%	69.19%
3	**Training**	**71.62%**	**71.27%**	**70.36%**	**71.27%**	**73.04%**	**72.45%**	**71.78%**	**70.25%**	**71.61%**	**69.97%**	**71.36%**
	Test	**71.36%**	**71.51%**	**75.00%**	**71.51%**	**64.10%**	**57.96%**	**71.50%**	**80.00%**	**68.25%**	**87.33%**	**71.85%**
4	Training	71.40%	70.82%	70.87%	70.99%	72.42%	72.50%	71.49%	70.19%	71.61%	69.91%	71.22%
	Test	70.45%	68.72%	74.50%	70.39%	61.54%	57.96%	69.57%	78.57%	68.25%	85.33%	70.53%
5	Training	72.94%	70.32%	69.97%	71.27%	72.76%	72.50%	74.45%	69.40%	71.84%	74.15%	71.96%
	Test	70.00%	65.92%	74.50%	71.51%	60.51%	57.96%	74.88%	78.57%	70.90%	76.00%	70.08%
6	Training	71.40%	71.27%	56.73%	71.27%	72.87%	72.72%	71.55%	64.16%	72.06%	70.52%	69.46%
	Test	70.45%	71.51%	67.50%	71.51%	61.54%	59.24%	68.60%	75.71%	69.84%	83.33%	69.92%
7	Training	72.88%	71.05%	70.02%	71.27%	72.36%	72.56%	73.25%	70.14%	71.78%	71.12%	71.64%
	Test	71.36%	72.07%	75.00%	71.51%	61.03%	57.96%	71.50%	80.00%	67.72%	79.33%	70.75%
8	Training	71.40%	68.70%	69.97%	71.27%	73.10%	73.22%	71.66%	71.05%	71.61%	70.24%	71.22%
	Test	70.45%	48.60%	75.00%	71.51%	61.03%	59.87%	69.57%	78.10%	68.25%	83.33%	68.57%
9	Training	71.22%	69.37%	70.36%	71.21%	73.04%	72.50%	71.66%	70.19%	71.61%	69.91%	71.11%
	Test	70.45%	63.13%	75.50%	71.51%	61.54%	57.96%	68.60%	80.00%	68.25%	87.33%	70.43%
10	Training	74.14%	69.37%	70.25%	70.99%	72.65%	72.83%	71.61%	72.01%	71.73%	70.68%	71.63%
	Test	70.91%	63.13%	75.00%	71.51%	61.03%	59.24%	68.60%	75.24%	68.78%	82.67%	69.61%
11	Training	74.31%	68.31%	70.02%	71.32%	72.70%	72.56%	73.08%	64.28%	71.22%	72.88%	71.07%
	Test	75.00%	57.54%	74.50%	71.51%	61.03%	57.96%	67.63%	76.67%	67.72%	76.67%	68.62%
12	Training	71.51%	69.98%	69.51%	71.32%	72.59%	72.17%	71.27%	70.93%	65.60%	69.58%	70.45%
	Test	70.45%	65.36%	73.50%	71.51%	60.00%	59.87%	68.12%	73.33%	68.25%	84.67%	69.51%
13	Training	72.25%	69.14%	70.53%	71.27%	72.42%	72.83%	72.12%	70.36%	71.61%	70.68%	71.32%
	Test	70.91%	61.45%	75.50%	71.51%	61.54%	59.24%	68.60%	77.62%	68.25%	80.00%	69.46%

rate (73.56%) for the training data set as compared with the other techniques, but it shows the worst result in the test data set.

From these results, our prediction models can give an early warning to managers of firms of their potential vulnerability when the firms are predicted to be financially below average. As decision makers, it would help them not only to identify critical factors in doing business, but also to adjust their allocated resources towards above average financial performance firms. In addition, the outcomes can also recommend important factors, which contribute to good financial performance in each scenario, as these factors should be taken more into consideration than other factors, given that firms always have limited resources and budgets.

Conclusions

Performance prediction is one of the important factors for decision making in business and financial studies. A good understanding of situations leads to good financial and managerial decisions. This study used 3 prediction methods, i.e., logistic regression, artificial neural etwork, and support vector machine to classify firm performance into 2 groups: financially above or below average performance based on the ROA. The developed models considered independent variables from both internal and external firm factors. The prediction models used independent variables from 1 year prior to the dependent variables. This is to predict whether a firm's performance is above or below average performance in the next coming year, for the companies listed on the SET. The results of this study include contributions from 2 aspects: academic and managerial aspects. For the academic aspect, the results confirm previous research that machine learning attains better accuracy than logistic regression for classification prediction models. For the managerial aspect, these prediction models help managers and decision makers to keep track of the performance 1 year in advance and help identify important business trends. With 10-fold cross validation, both the ANN and SVM attained better performances than a classical statistical technique (logistic regression). The ANN attained the highest

Table 7. Results of support vector machine

Test year	Training	Test
2005	62.76%	70.45%
2006	61.14%	76.62%
2007	62.60%	70.00%
2008	64.39%	72.07%
2009	67.01%	61.54%
2010	65.54%	57.96%
2011	63.26%	68.12%
2012	63.14%	80.00%
2013	65.65%	69.31%
2014	**72.50%**	**86.67%**
Average	**64.80%**	**71.27%**

Table 8. Classification summary

Techniques	Training	Test
Logistic regression	73.56%	68.42%
Artificial neural network (ANN)	71.36%	71.85%
Support vector machine (SVM)	64.80%	71.27%

accuracy rates with an average rate of 71.85% followed by the SVM (71.27%) and logistic regression (68.42%). Our results show that for Thai public companies' performance predictions, the machine learning techniques have an advantage over logistic regression in prediction accuracy. However, they fail to explain the model variables. Conversely, logistic regression can be easily understood, with its predicting equations that are not provided with machine leaning techniques. This helps to identify important factors which contribute to a firm's financial success in a given year.

The models in the study can only predict a firm's financial performance just 1 year in advance. The findings presented in this paper should be considered with this limitation. Further studies should consider extending the prediction time frame to more than 1 year, as the consequences of an action could have effects on a firm's financial performance further into the future. In addition, an evolutionary approach such as genetic algorithms can also be used in the optimization functions in the ANN and SVM for improving their prediction accuracy and helping to identify important factors which contribute to a firm's financial success. Thus, in the future, new models with an evolutionary process for optimizing classifiers can be further investigated.

Acknowledgments

This research is financially supported by Sirindhorn International Institute of Technology (SIIT), Thammasat University, Japan Advanced Institute of Science and Technology, and The Thailand National Electronics and Computer Technology Centre.

References

Agustinus, P. and Rachmadi, P. (2008). Determinants of corporate performance of listed companies in Indonesia. Munich, Germany: Munich University Library. Available from: mpra.ub.uni-muenchen.de/6777/1/MPRA paper 6777.pdf. Accessed date: 1 December 2016.

Almajali, A.Y., Alamro, S.A., and Al Soub, Y.Z. (2012). Factors affecting the financial performance of Jordanian insurance companies listed at Amman Stock Exchange. J. Manag. Research., 4(2):266-289.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Financ., 23(4):589-609.

Alzharani, A.M., Ahmad, A.B., and Aljaaidi, K.S. (2011). An empirical investigation of factors associated with firm performance evidence from Kingdom of Saudi Arabia. International Proceedings of Economics Development and Research, 25(1):30-36.

Bagchi, B. (2015). A long-run and short-run cointegration model explaining relationship between liquidity management and profitability. Int. J. Busi. Excel., 8(2):123-145.

Bank of Thailand. (2017). Statistics. Available from: www2.bot.or.th. Accessed date: 2 January 2017

Basterretxea, K., Tarela, J.M., and Del Campo, I. (2004). Approximation of sigmoid function and the derivative for hardware implementation of artificial neurons. IEE P-Circ. Dev. Syst., 151(1):18-24.

Biesebroeck, J.V. (2005). Firm size matters: Growth and productivity growth in African manufacturing. Econ. Dev. Cult. Change, 53(3):545-583.

Bose, I. and Pal, R. (2006). Predicting the survival or failure of click-and-mortar corporations: A knowledge discovery approach. Eur. J. Oper. Res., 174(2):959-982.

Boyacioglu, M.A., Kara, Y., and Baykan, O.K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of saving deposit insurance fund (SDIF) transferred banks in Turkey. Expert Syst. Appl., 36(2):3355-3366.

Chang, C.C. and Lin, C.J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3):27: 1-27:27.

Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. J. Bus. Res., 68(5):1012-1025.

Delen, D., Kuzey, C., and Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. Expert Syst. Appl., 40(10):3970-3983.

Doan, N.P. and Nguyen, J.D. (2011). Firm characteristics, capital structure and operational performance: A Vietnamese study. Proceedings of the 7th Asia-Pacific Economic Association Conference; June 24-25, 2011; Busan, South Korea.

Evans, D.S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. J. Ind. Econ., 35(4):567-581.

Galbraith, V.L. and Thomas, D.S. (1941). Birth rates and the interwar business cycles. J. Am. Stat. Assoc., 36(216):465-476.

Hovakimian, A., Opler, T., and Titman, S. (2001). The debt-equity choice. J. Financ. Quant. Anal., 36(1):1-24.

Iswatia, S. and Anshoria, M. (2007). The influence of intellectual capital to financial performance at insurance companies in Jakata Stock Exchange

(JSE). Proceedings of the 13th Asia Pacific Management Conference; November 18-20, 2007; Melbourne, Australia, p. 1393-1399.

Khalifa, K.M. and Shafii, Z. (2013). Financial performance and identify affecting factors in this performance of non-oil manufacturing companies listed on Libyan Stock Market (LSM). European J. Bus. Manag., 5(12):82-99.

Kumar, P.R. and Ravis, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. Eur. J. Oper. Res., 180(1): 1-28.

Lam, M. (2004). Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. Decis. Support Syst., 37(4):567-581.

Lee, Y.C. (2007). Application of support vector machines to corporate credit rating prediction. Expert Syst. Appl., 33(1):67-74.

Liargovas, P. and Skandalis, K. (2010). Factor affecting firms financial performance: The case of Greece. Global Business and Management Research, 2(2&3):184-197.

Loderer, C. and Waelchli, U. (2010). Firm age and performance. European Corporate Governance Institute, Finance Working Paper No. 230/2009. Rochester, NY, USA: SSRN. Available from: ssrn.com/abstract=1342248. Accessed date: 17 November 2016

Lussier, R.N. (1995). A nonfinancial business success versus failure prediction model for young firms. J. Small Bus. Manage., 33(1):8-20.

Mahajan, R., Garg, S., and Sharma, P. (2014). An illustration of logistic regression technique: a case of processed food sector. Int. J. Bus. Excel., 7(5):659-676.

Malhotra, G. and Mukherjee, A. (2013). Comparative study of forced distribution and absolute rating performance management systems. Int. J. Bus. Excel., 6(2):121-130.

Mokhatab Rafiei, F., Manzari, S.M., and Bostanian, S. (2011). Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. Expert Syst. Appl., 38(8):10210-10217.

Naidu, S. and Chand, A. (2013). Financial problems faced by micro, small and medium enterprises in the small island states: a case study of the manufacturing sector of the Fiji Islands. Int. J. Bus. Excel., 6(1): 1-21.

Naser, K and Mokhtar, M.Z. (2004). Determinants of corporate performance of Malaysian companies. Proceedings of the 4th Asia Pacific Interdisciplinary Research in Accounting Conference; July 4-6, 2004; Singapore, p. 16-25.

Ogut, H., Doganay, M.M., Ceylan, N.B., and Aktas, R. (2012). Prediction of bank financial strength ratings: The case of Turkey. Econ. Model., 29(3):632-640.

Pantea, M., Gligor, D., and Anis, C. (2014). Economic determinants of Romanian firms' financial performance. Procedia - Social and Behavioral Sciences, 124(1):272-281.

Pao, H.T. (2008). A comparison of neural network and multiple regression analysis in modeling capital structure. Expert Syst. Appl., 35(3):720-727.

Pratheepkanth, P. (2011). Capital structure and financial performance: Evidence from selected business companies in Colombo Stock Exchange Sri Lanka. Int. Ref. Res. J., 2(2):171-183.

Ray, S. (2011). Prediction of corporate financial distress in glass and glassware sector of India. Int. J. Bus. Excel., 4(6):678-695.

Rich, E. and Knight, K. (1991). Artificial Intelligence. 2nd ed. Singapore: McGraw-Hill, Singapore, 640p.

Roiger, J.R. and Geatz, M.W. (2003). Data Mining A Tutorial Based Primer. Adison Wesley, Boston, MA, USA, 408p.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning internal representations by error propagation. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Rumelhart, D.E. and Williams, J.L., (eds).:, MIT Press, Cambridge, MA, USA, p. 318-362.

Russell, S. and Norvig, P. (2003). Artificial Intelligence: A Modern Approach. 2nd ed. Prentice Hall, Upper Saddle River, NJ, USA, 960p.

Sharma, S.K. and Raina, D. (2013). Performance analysis of Indian automobile sector using non-parametric approach. Int. J. Bus. Excel., 6(5):505-525.

Shin, K-S., Lee, T.S., and Kim, H-J. (2005). An application of support vector machines in bankruptcy prediction model. Expert Syst. Appl., 28(1):127-135.

Stock Exchange of Thailand. (2017). Companies/ Securities in Focus. Available from: www.set.or.th. Accessed date: 2 January 2017

Tam, K.Y. and Kiang, M.Y. (1992). Managerial applications of neural networks: the case of bank failure prediction. Manage. Sci., 38(7):926-947.

Tinoco, M.H. and Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting market and macroeconomic variables. International Review of Financial Analysis, 30:394-419.

Vapnik, V.N. (1998). Statistical Learning Theory. John Wiley and Sons, Hoboken, NJ, USA, 768p.

Weintraub, R. (1962). The birth rate and economic development: An empirical study. Econometrica, 40(4):812-817.

Wu, W., Lee, V.C., and Tan, T.Y. (2006). Data preprocessing and data parsimony in corporate failure forecast models:Evidence from Australian materials industry. Account. Financ., 46(2):327-345.

Xu, W., Xiao, Z., Dang, X., Yang, D., and Yang, X. (2014). Financial ratio selection for business failure prediction using soft set theory. Knowledge-Based Systems, 63(1):59-67.

Yang, C.C., Lee, C.F., Gu, Y.X., and Lee, Y.W. (2010). Co-determination of capital structure and stock returns - A LISREL approach: An empirical test of Taiwan stock markets. Q. Rev. Econ. Financ., 50(2):222-233.

Youn, H. and Gu, Z. (2010). Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model. Int. J. Hosp. Manag., 29(1):120-127.

Zeitun, R. and Tian, G. (2007). Capital structure and corporate performance: Evidence from Jordan.

Australian Accounting, Business and Finance J., 1(4):40-61.
Zhang, G., Hu, M.Y., Patuwo, B.E., and Indro, D.C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. Eur. J. Oper. Res., 116(1):16-32.