Data Mining for Bankruptcy Prediction: An Experiment in Vietnam

Dang Ngoc Hung
Faculty of Accounting and Auditing
Hanoi University of Industry
298, Cau Dien street, Bac Tu Liem, Hanoi, Vietnam
dangngochung@haui.edu.vn

Vu Thi Thanh Binh
Faculty of Accounting and Auditing
Hanoi University of Industry
298, Cau Dien street, Bac Tu Liem, Hanoi, Vietnam
vuthithanhbinh@haui.edu.vn

Abstract—In the history of the world economy, the bankruptcy of some large companies has caused global financial crises. The study aimed to postulate a model of bankruptcy prediction for listed companies on Vietnam's stock market. The research used six popular algorithms in data mining to predict bankruptcy risk with data collected from 4693 observations in the period 2009-2020. The research results showed that Logistic algorithms, Artificial Neural Network, Decision Tree have a high level of predicting bankruptcy with an accuracy of 98%. The study identified the three most important indicators: inventory turnover ratio, debt to equity ratio, and debt ratio that affect the corporate bankruptcy prediction. The study showed the threshold points of 10-indicators to avoid bankruptcy likelihood. These results recommended that the model could be applied in practice to reduce risks for businesses and investors in the Vietnamese market.

Index Terms—Bankruptcy prediction, data mining, Artificial Intelligence, Decision Tree, Z-Altman index.

I. INTRODUCTION

THE REPORT of The World Bank [1] indicates that Vietnam is an active emerging economy with speedy economic growth in the East Asia area. Besides the development of business, the economy has many potential risks. The context is that the global economic growth outlook is somewhat bleak in the face of uncertain potential risks such as the US-China Trade War, Brexit, inflationary trends due to unpredictable price changes. The epidemic of Coronavirus (Covid-19 epidemic) strongly affects people's psychology in general and stock investors in particular. Measuring the health of enterprises in Vietnam is now extremely urgent. There are several models to predict corporate bankruptcy, for example, the model of market approach [2] and the model of accounting approach [3]. Employing models in practice in Vietnam's stock market is essential because of the difficulty of qualitative predictability in the increasingly unpredictable environment. The models support how to measure the bankruptcy prediction from potential business risks. In Vietnam, the quality of accounting information is not too excellent [4] and companies listed or unlisted on the stock market report losses leading to a high risk of bankruptcy. To ensure the rights and benefit of enterprises and creditors, the Law of Vietnam in which the Law on Bankruptcy 2014 and the Law on Securities 2010 (the latest being the Law on Securities 2019 takes effect from January 1, 2021) have issued and concretized these regulations.

Previous studies have given diverse criteria as financial ratios in predicting corporate bankruptcy. Some studies show that the Z-Score model has a strong practical application of financial status to the prediction of bankruptcy as studied by Liang, Lu, Tsai, Shih [5], Barboza, et al. [6], Chou, et al. [7], Antunes, et al. [8], Le, et al. [9], Le, et al. [10], Veganzones and Séverin [11], Mai, et al. [12], Son, et al. [13], Chen, et al. [14]. However, previous studies were mainly used in developed countries to predict bankruptcy and few studies applied data mining in predicting bankruptcy, especially in emerging security markets such as Vietnam.

This study uses several data mining techniques to predict corporate bankruptcy for a Vietnamese case study. The main contributions of this study are as follows: (i) building a framework model for predicting bankruptcy, (ii) Collecting Vietnam's data sets for the past twelve years for the bankruptcy prediction, (iii) testing to compare technical performance for predicting bankruptcy on the Vietnamese dataset; and (iv) Combining Bagging and Boosting methods, the test results show the best overall accuracy of 98% to improve forecasting bankruptcy.

Adopting and combining new techniques to improve the accuracy in forecasting corporate bankruptcy is encouraged by researchers and practitioners. The results help to reinforce and enhance the bankrupting prediction model.

II. LITERATURE REVIEW

Research on predicting the financial downturn of companies through Z-score and Zeta models Altman [15], this is a handbook that presents the quantitative techniques commonly used in research papers. Empirical finance research along with real, modern research examples. By referring to this handbook, the author has understood and applied it to the study of the Z-score model. Konglai and Jingjing [16] compiled a sample of failed managed groups and normally managed groups that contained 130 listed companies from Shanghai and Shenzhen exchanges in 2009. Using the MDA discriminant analysis model and the logistic model, the author chooses 5 financial factors: profitability, debt repayment ability, operating ability, growth ability, and capital structure. Ohlson [17] was the first to apply the logistic regression model in the study to predict the probability of default of enterprises. Some related studies such as Meeampol, et al. [18] in the Thai stock market. Research by Kumar and Rao [19], on a new method to estimate internal credit risk and predict bankruptcy under the Basel II regime. The results of the study showed that the Z-score could predict bankruptcy with 98.6% accuracy compared with 93.5% according to Altman's score.
Researchers use various algorithms of intelligent techniques to solve the problem of corporate bankruptcy [20]. According to Serrano-Cinca [21] and Fletcher and Goss [22], neural networks (NNs) are the most commonly used technique. And the data mining algorithms used to predict bankruptcy risk include decision trees (DT) and support vector machines (SVM) [23]. A decision tree is a structured hierarchical tree used to classify objects based on a series of rules. When given data about objects containing attributes along with their classes, the decision tree will generate rules to predict the class of the unknown objects (unseen data). Support vector machines (SVM) is a supervised machine learning model used to analyze and classify data. SVM takes incoming data and classifies them into two different classes. Many studies have used data mining techniques in predicting bankruptcy. Some studies related to predicting bankruptcy using data mining techniques are listed in Table 1.

### III. METHODOLOGY

#### A. Measuring Variables

There are many measures for predicting bankruptcy; however, each measure has both advantages and disadvantages. Ghazali, et al. [24] state that the Altman Z-Score is probably the most popular measure of a company’s financial health and has been used to determine bankruptcy prediction in numerous studies. This study will determine the bankruptcy prediction based on the Z-score approach of Altman [3]. Altman's Z-score gives a calculation of the Z-score based on the following formula:

\[
Z = 0.717A_1 + 0.847A_2 + 0.107A_3 + 0.420A_4 + 0.998A_5
\]

In which: A1 - Current assets minus current liabilities, then divided by total assets; A2 - Retained profit divided by total assets; A3 - Profit before tax and interest divided by total assets; A4 - Book value of equity divided by total liabilities; A5 - Revenue divided by total assets.

If the Z-index < 1.81, the company is in the bankruptcy prediction zone that the likelihood of bankruptcy will be assigned a value of 1. Otherwise, it will be assigned a value of 0.

This study uses 30 attributes of financial indicators including liquidity ratios, capital budgeting ratios, profitability ratios, efficiency ratios (activities ratios), market ratios, and debt ratios (leverage ratios). The properties are briefly described in Appendix 1.

#### B. Applying Data Mining Algorithms

Data mining has many different expressions. It is the process of automatically extracting valuable information which is predictive information hidden in the huge amount of data in reality. Data mining emphasizes automated and predictive aspects. This study uses Logistic Regression, Bayesian Network, K-nearest neighbor, Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree that is commonly used to predict bankruptcy.

1. **Logistic Regression**: The Logistic regression model introduced by Berkson [25] is a commonly used tool in data analysis with binary variables. Some developments by Altman, et al. [26] and Flitman [27] are used in multivariate regression analysis, discriminant analysis. From this binary dependent variable, a procedure will be used to predict the probability of the event occurring according to the rule if the predicted probability is greater than 0.5 (default cut-off point) then the prediction result will be “yes” occurs, otherwise, the predicted result will be given as "no". The Binary Logistic regression model is as follows:

![Fig. 1 Binary logistic regression model](image)

P is the probability that \( Y = 1 \) when the independent variables take on a particular value. Accordingly, the probability that the event does not occur is:
\[ 1 - P = \text{Prob}(Y = 0) = 1 - \frac{e^z}{1 + e^z} = \frac{1}{1 + e^z} \]

The regression coefficients were estimated by the method of Maximum Likelihood (ML). The logit regression model can be used to estimate the log(likelihood) ratio for each independent variable of the model of Ohlson [17]. The parameters \( \beta_i \) were estimated by the method of ML.

2) Bayesian: Bayesian Network is applied for classification based on a probabilistic graphical model and the probability of the Bayesian Network is a value from 0 to 1. Bayesian Network is a set of variables and their conditional dependencies that are linked together by a probability association. According to Carlin and Louis [28], the Bayesian method is more about statistics than regression. For fraud detection, a Bayesian network will be built with Bayesian rule along with the condition \( P(Y=1) + P(Y=0) = 1 \) written as follows:

\[
P(Y=1|X) = \frac{P(X|Y=1)P(Y=1)}{P(X)}
\]

\[
P(Y=0|X) = \frac{P(X|Y=0)P(Y=0)}{P(X)}
\]

In which:

\[ P(X) = P(Y=1)P(X|Y=1)+P(Y=0)P(X|Y=0) \]

The components are calculated as follows: \( P(Y=1) \) is the error rate of the sample used to run the model, assuming the variables are independent.

3) K-Nearest Neighbors (K-NN): K-Nearest Neighbors algorithm is used in data mining. K-NN is a method to classify objects based on query points and all the objects in the training data. An object is classified based on its K neighbors. K is a positive integer that is determined before performing algorithms. Euclidean distance is often used to calculate the distance between objects.

4) Artificial Neural Network (ANN): Artificial Neural Network is an information processing model that is simulated based on the activity of the nervous system of an organism. A neural network can consist of one or more neurons that each neuron is an information processing unit and the connections between neurons form a network structure. A neural network is a computational model defined by parameters: Neuron type, connection architecture, and learning algorithms. The neurons are connected by a weight matrix. The typical structure of a neural network consists of three layers: input, hidden, and output [29] (see Fig. 2).

5) Support Vector Machine (SVM): A support vector machine (SVM) is a classical algorithm that solves problems of big data classification [30]. SVM takes input and classifies them into two different classes. With a given set of training examples belonging to two given categories, SVM builds an SVM model to classify other examples into those two categories. SVM learns a hyperplane to classify the data set into two separate classes by constructing a hyperplane or a set of hyperplanes in a multi-dimensional or infinite-dimensional space. For the best classification, it is necessary to determine the optimal hyperplane located as far away from the data points of all classes as possible.

Fig. 3 depicts the SVM algorithm: Given a training set represented in a vector space where each document is a point, this method finds a decision hyperplane \( h \) that can best divide the points on this space into two separate layers, respectively, the layer containing the data containing the feature simulated by the black dot and the layer containing the data containing the feature simulated by the white dot. The quality of this hyperplane is determined by the boundary of the nearest data point of each layer to this plane. The purpose of the SVM algorithm is to find the maximum boundary distance.

6) Decision Tree: A Decision Tree is a classification model introduced by Belson [31], widely used in many different fields. After the introduction of the machine learning method system, the Decision Tree was further developed with the C4.5 algorithm by Quinlan [32] and the ID3 algorithm by Quinlan [33]. A Decision Tree is a structured classification tree that classifies objects based on sequences of rules. To determine which variable to use classification first, which variable to use later, the information weight (entropy) for each variable is calculated, the higher entropy, the more categorical information the variable carries.

C. Combining Techniques for Data Mining

For improving the accuracy of the method of hybridization of models in the classification problem, this research employed Boosting and Bagging to improve the accuracy of the classification algorithms.

Bagging comes from two abbreviations, Bootstrap and Aggregation [34]. Bagging is a combination of independent base models that leads to a significant reduction in errors. Therefore, the goal is to get as many base models as independent as possible. Bagging generates classifiers from subsets that revert to the Bootstrap samples and a machine learning algorithm, each of which generates a basic classifier. The classifiers will be combined by the majority voting method. That is, when there is an example that needs to be classified, each classifier will produce a result. And the result that appears the most will be taken as the result of the
combiner. The Bagging generates N-selected training sets with iterations from the original training data set.

Boosting is a method of building a set of weak classifiers to improve the efficiency of these classifiers. After each iteration, the weak classifier will focus on learning on elements that were misclassified in previous iterations. To classify newly arrived data, people use the majority voting rule from the classification results of each weak classification model [35].

D. Evaluating The Model

Confusion Matrix is commonly used in model evaluation. This study employs a calculation of indices of Confusion Matrix as shown in Table 2.

<table>
<thead>
<tr>
<th>Reality</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td>Negative</td>
<td>False Positive (FP)</td>
</tr>
</tbody>
</table>

The effectiveness of the opinion classification model is evaluated based on 4 indexes: Accuracy, Precision, Recall, and Harmonized Mean (F1-score). In which:

\[
\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
F1-\text{Score} = 2 \frac{1}{\text{Precision} + 1/\text{Recall}}
\]

E. Collecting Data

This study uses data collected from the Vietnamese stock exchange in the period 2009 - 2020. Data is collected from audited financial statements of listed companies after excluding companies in the field of listed companies, banking, securities, and insurance sectors. After determining the indicators, the data used to perform analysis and forecasting is 4693 observations, presented in Table 3 by year and by field.

The study objectives are to use data mining algorithms including Logistic Regression, Bayesian Network, K-nearest neighbor, Artificial Neural Network, Support Vector Machine, and Decision Tree for predicting bankruptcy and to determine the accuracy of these data mining algorithms. The data are randomly divided into 2 parts to build and test the model: Training data is used for building the research model and testing data is used to test the predictive likelihood of the model. The description of indicator characteristics in the research model is presented in Appendix 1. Out of 4693 observations, 2395 observations are at risk of bankruptcy, accounting for 51.03% and vice versa 48.97% is normal. Thus, the data on the number of normal enterprises and the bankruptcy likelihood is quite balanced.

Appendix 1 reveals a testing result of the difference in the mean value of 30 indicators in the research model between the normal enterprise group and the bankruptcy likelihood group. 27/30 indicators that have a difference between the two groups and are statistically significant, only 3 of the indicators of growth have no difference between the two normal groups and the bankruptcy likelihood group including X18-Operating profit growth, X19-Net profit growth, and X20-Equity growth.

IV. RESULTS AND DISCUSSIONS

To achieve the objective of the study on the question of commonly used classification algorithms in Data mining, which algorithm gives the best predictive results, Weka software is applied to research data to conduct experiments. Logistic regression and ANN algorithms give a high probability of bankruptcy prediction (accuracy over 97%).

To improve the accuracy of the method of hybridization of models in the classification, Boosting and Bagging methods are employed. The results presented in Figure 5, Figure 6, and Figure 7 show the accuracy. Bankruptcy prediction results of Bagging and Boosting methods have improved over the original basic methods.
With 30 ratios used for forecasting in the model, which ratio is the most important and has the most predictive significance? When using Weka software to identify significant variables in bankruptcy prediction, Figure 8 shows 10 ratios that have the greatest impact on corporate bankruptcy prediction. In which X13 - Total asset turnover ratio is the most important indicator, followed by X5 - Debt to equity ratio, and X24 - Debt ratio.

We select the 10 most significant indicators as a set of important indicators in predicting bankruptcy including X13 - Total assets turnover ratio, X5 - Debt to equity ratio, X24 - Debt ratio, X3 - Receivables turnover ratio, X26 - Receivables conversion period, X27 - Payables conversion period, X7 - Operating cash flows ratio, X1 - Current ratio, X14 - Inventory conversion period, and X2 - Quick ratio. To test again whether the financial ratios are the most important in predicting bankruptcy, we use the dataset with a set of 10-ratio replaces the set of 30-ratio.

**V. CONCLUSIONS AND RECOMMENDATIONS**

This study uses data mining to predict corporate bankruptcy. The sample is companies that have been listed in Vietnamese enterprises. Appendix 7 depicts the Decision Tree results of the 10 most important indicators which lead to the prediction is far superior to that of the dataset with full variables. From these results, this research suggests choosing a set of the most important indicators for predicting bankruptcy that saves resources in forecasting at high accuracy.

The research continues to study using the Decision Tree algorithm (J48) after removing the ratio that has no influence or little importance to perform the analysis. The results show that the Decision Tree algorithms predict bankruptcy with an accuracy of 97.9%, implying that it is appropriate to use the Decision Tree model to predict bankruptcy for Vietnamese enterprises. Even in the Bayesian and KNN algorithms, the accuracy of prediction accuracy rates such as Logistics, ANN, and DT. Even in the Bayesian and KNN algorithms, the accuracy of the prediction is far superior to that of the dataset with full indicators. From these results, this research suggests choosing a set of the most important indicators for predicting bankruptcy that saves resources in forecasting at high accuracy.
Vietnam in the period 2009-2020. This study is to evaluate whether data mining algorithms can be used to predict the bankruptcy of companies in Vietnam accurately or not, which financial indicators are the most effective ratios to predict. To achieve the research objectives, the research has in turn used algorithms including Logistic Regression, Bayesian Network, K-nearest neighbor, Artificial Neural Network (ANN), Support Vector Machine, and Decision Tree. Based on the research results, it can be seen that all 6 methods are accurate in predicting the status, normal, or risk of bankruptcy, of the companies in the sample. In addition, we recommend the use of Decision Tree, ANN that will give the highest prediction accuracy. It can be concluded that these models are suitable for predicting bankruptcy for Vietnamese enterprises in the current period. Moreover, the research has shown 10 financial ratios that are most important in predicting bankruptcy risk. From the above research results, the research has some recommendations for businesses and investors as well as practical suggestions for listed companies to minimize bankruptcy risk.

Total assets turnover ratios, debt to equity ratio, and debt ratio are the three most important indicators in predicting the bankruptcy risk of a business. The results also show that Vietnamese enterprises during the study period are at risk of bankruptcy due to improper implementation of investment decisions stemming from the use of excessive financial leverage and inefficient business activities. The evidence of this research is an important scientific basis for financial managers when planning strategies.

It is necessary to carry out the process to improve the health of the business on the existing foundation, the process of making fundamental changes in the business to increase the ability to operate more efficiently, and create a better "new normal" environment for the business to achieve the strategies and goals. The research postulates some recommendations: Prepare financial statements under the current regulations of the Ministry of Finance; Financial statements must be audited by reputable auditing agencies; In addition to cultivating knowledge about management and law, the listed companies need to regularly improve their knowledge in corporate finance, especially financial ratios to measure business health.

The results show that financial managers need to be careful with regulations on mobilizing funding sources, fully exploiting internal capital sources, especially from retained earnings to reduce the cost of using corporate capital and to limit the use of debt, especially short-term debt. Moreover, the financial managers need to increase the exploitation of highly liquid assets to improve investment efficiency. Furthermore, the financial managers need to regularly re-check the investment regulations so that the business plan can be adjusted in time.

However, the study still has certain limitations. The factors which impact corporate bankruptcy are not only in financial ratios but also come from human behavior. This study has not mentioned the intervening factors such as human behavior, crowd psychology, and speculation affecting the increase or decrease in bankruptcy risk of listed companies in Vietnam.

References


<table>
<thead>
<tr>
<th>Indicators</th>
<th>Measure</th>
<th>Mean</th>
<th>Std.</th>
<th>1%</th>
<th>5%</th>
<th>95%</th>
<th>99%</th>
<th>Normal group (2298; 48.97%)</th>
<th>Bankruptcy risk (2095; 51.03%)</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current ratio</td>
<td>X1: Current assets/Current liabilities</td>
<td>2.0</td>
<td>1.7</td>
<td>0.5</td>
<td>1.1</td>
<td>1.5</td>
<td>2.2</td>
<td>9.7</td>
<td>2.571</td>
<td>1.493</td>
<td>0.0000</td>
</tr>
<tr>
<td>Quick ratio</td>
<td>X2: (Current assets - Inventories)/Current liabilities</td>
<td>1.4</td>
<td>1.5</td>
<td>0.2</td>
<td>0.6</td>
<td>1.6</td>
<td>2.2</td>
<td>8.7</td>
<td>1.885</td>
<td>0.983</td>
<td>0.0000</td>
</tr>
<tr>
<td>Receivables turnover ratio</td>
<td>X3: Net sales/Receivables</td>
<td>7.7</td>
<td>10.3</td>
<td>0.4</td>
<td>2.2</td>
<td>4.5</td>
<td>8.6</td>
<td>54.0</td>
<td>11.048</td>
<td>4.435</td>
<td>0.0000</td>
</tr>
<tr>
<td>Current operating cash flows ratio</td>
<td>X4: Operating cash flows/Current liabilities</td>
<td>0.3</td>
<td>0.6</td>
<td>-0.9</td>
<td>0.0</td>
<td>0.1</td>
<td>0.4</td>
<td>2.6</td>
<td>0.396</td>
<td>0.128</td>
<td>0.0000</td>
</tr>
<tr>
<td>Debt to equity ratio (D/E)</td>
<td>X5: Total liabilities/Shareholder’s equity</td>
<td>1.5</td>
<td>1.4</td>
<td>0.1</td>
<td>0.5</td>
<td>1.1</td>
<td>2.0</td>
<td>7.0</td>
<td>0.876</td>
<td>2.139</td>
<td>-33.1772</td>
</tr>
<tr>
<td>Fixed assets to long-term capital ratio</td>
<td>X6: Fixed assets/Total liabilities and owners equity – Current liabilities</td>
<td>0.3</td>
<td>0.3</td>
<td>0.0</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>1.1</td>
<td>0.320</td>
<td>0.362</td>
<td>-5.079</td>
</tr>
<tr>
<td>Operating cash flows ratio</td>
<td>X7: Operating cash flows/Total debts</td>
<td>0.2</td>
<td>0.5</td>
<td>-0.7</td>
<td>0.0</td>
<td>0.1</td>
<td>0.3</td>
<td>2.2</td>
<td>0.336</td>
<td>0.052</td>
<td>22.1178</td>
</tr>
<tr>
<td>Operating ratio</td>
<td>X8: Net profit/Net sales</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.079</td>
<td>0.071</td>
<td>2.6419</td>
</tr>
<tr>
<td>Operating profit margin ratio</td>
<td>X9: Operating profit/Net sales</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.5</td>
<td>0.966</td>
<td>0.087</td>
<td>2.4434</td>
<td>0.0146</td>
</tr>
<tr>
<td>Net profit to book value of equity ratio</td>
<td>X10: Net profit/Book value of equity</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.151</td>
<td>0.094</td>
<td>19.8108</td>
</tr>
<tr>
<td>Return on equity ratio</td>
<td>X11: Net profit/Equity</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.149</td>
<td>0.090</td>
<td>20.7952</td>
</tr>
<tr>
<td>Operating income to book value of shareholder’s equity</td>
<td>X12: Operating income/Book value of shareholder’s equity</td>
<td>0.2</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>0.187</td>
<td>0.118</td>
<td>19.9933</td>
</tr>
<tr>
<td>Total asset turnover ratio</td>
<td>X13: Net sales/Total assets</td>
<td>1.2</td>
<td>0.9</td>
<td>0.1</td>
<td>0.5</td>
<td>0.9</td>
<td>1.5</td>
<td>4.6</td>
<td>1.673</td>
<td>0.683</td>
<td>45.1438</td>
</tr>
<tr>
<td>Inventory conversion period</td>
<td>X14: Inventories x365 days/Cost of goods sold</td>
<td>-167.6</td>
<td>377.2</td>
<td>-2001.1</td>
<td>-147.9</td>
<td>-78.1</td>
<td>-34.2</td>
<td>0.0</td>
<td>-73.660</td>
<td>257.680</td>
<td>17.2326</td>
</tr>
<tr>
<td>Fixed asset turnover ratio</td>
<td>X15: Net sales/Fixed assets</td>
<td>21.5</td>
<td>50.4</td>
<td>0.3</td>
<td>3.3</td>
<td>7.5</td>
<td>17.1</td>
<td>280.1</td>
<td>26.764</td>
<td>16.441</td>
<td>7.0549</td>
</tr>
<tr>
<td>Revenue growth</td>
<td>X16: Net sales year t/Net sales year t-1</td>
<td>1.2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.9</td>
<td>1.1</td>
<td>1.2</td>
<td>3.7</td>
<td>1.139</td>
<td>1.177</td>
<td>-2.2745</td>
</tr>
<tr>
<td>Assets growth</td>
<td>X17: Total assets year t/Total assets year t-1</td>
<td>1.1</td>
<td>0.3</td>
<td>0.7</td>
<td>1.0</td>
<td>1.1</td>
<td>1.2</td>
<td>2.4</td>
<td>1.095</td>
<td>1.156</td>
<td>-6.381</td>
</tr>
<tr>
<td>Operating profit growth</td>
<td>X18: Operating profit year t/Operating profit year t-1</td>
<td>1.2</td>
<td>2.0</td>
<td>-2.7</td>
<td>0.7</td>
<td>1.0</td>
<td>1.3</td>
<td>8.9</td>
<td>1.218</td>
<td>1.155</td>
<td>1.0887</td>
</tr>
<tr>
<td>Net profit growth</td>
<td>X19: Net profit year t/Net profit year t-1</td>
<td>1.2</td>
<td>2.6</td>
<td>-3.4</td>
<td>0.7</td>
<td>1.0</td>
<td>1.3</td>
<td>10.8</td>
<td>1.243</td>
<td>1.203</td>
<td>0.5261</td>
</tr>
<tr>
<td>Equity growth</td>
<td>X20: Equity year t/Equity year t-1</td>
<td>1.1</td>
<td>0.3</td>
<td>0.8</td>
<td>1.0</td>
<td>1.1</td>
<td>2.3</td>
<td>1.212</td>
<td>1.133</td>
<td>-1.252</td>
<td>0.2206</td>
</tr>
<tr>
<td>Stock price trend</td>
<td>X21: Share price year t/Share price year t-1</td>
<td>1.2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.8</td>
<td>1.0</td>
<td>1.4</td>
<td>3.3</td>
<td>1.204</td>
<td>1.149</td>
<td>3.6644</td>
</tr>
<tr>
<td>Earnings per share (EPS)</td>
<td>X22: Net profit/Number of shares</td>
<td>2264.8</td>
<td>2238.3</td>
<td>-1511.0</td>
<td>660.0</td>
<td>1667.0</td>
<td>3257.0</td>
<td>9907.0</td>
<td>2809994.0</td>
<td>1664933</td>
<td>19.115</td>
</tr>
<tr>
<td>General solvency</td>
<td>X23: Total assets/Current liabilities</td>
<td>3.7</td>
<td>3.5</td>
<td>1.2</td>
<td>1.8</td>
<td>2.5</td>
<td>4.2</td>
<td>19.1</td>
<td>4.333</td>
<td>3.079</td>
<td>13.0359</td>
</tr>
<tr>
<td>Debt ratio</td>
<td>X24: Total debts/Total assets</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>0.9</td>
<td>0.382</td>
<td>0.067</td>
<td>-42.124</td>
</tr>
<tr>
<td>Firm size</td>
<td>X25: Ln(total assets)</td>
<td>27.2</td>
<td>1.4</td>
<td>24.1</td>
<td>26.2</td>
<td>27.2</td>
<td>28.1</td>
<td>30.7</td>
<td>26.688</td>
<td>27.623</td>
<td>-24.0555</td>
</tr>
<tr>
<td>Receivables conversion period</td>
<td>X26: Receivables x365 days/Net sales</td>
<td>136.2</td>
<td>178.6</td>
<td>8.8</td>
<td>42.4</td>
<td>81.8</td>
<td>165.5</td>
<td>847.3</td>
<td>76.747</td>
<td>193.238</td>
<td>-23.9024</td>
</tr>
<tr>
<td>Payables conversion period</td>
<td>X27: Payables x365 days/Net sales</td>
<td>45.4</td>
<td>52.1</td>
<td>1.0</td>
<td>16.2</td>
<td>31.0</td>
<td>55.2</td>
<td>264.7</td>
<td>26.677</td>
<td>63.598</td>
<td>-25.6931</td>
</tr>
<tr>
<td>Book value</td>
<td>X28: Shareholder’s equity/Number of shares</td>
<td>18253</td>
<td>8237.2</td>
<td>7206</td>
<td>12721</td>
<td>16010</td>
<td>21274</td>
<td>90973</td>
<td>1920050.0</td>
<td>17344450</td>
<td>7.7654</td>
</tr>
<tr>
<td>State ownership ratio</td>
<td>X29: Proportion of State holding shares</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.5</td>
<td>0.8</td>
<td>0.264</td>
<td>0.220</td>
<td>6.219</td>
</tr>
<tr>
<td>Foreign ownership ratio</td>
<td>X30: Proportion of foreign holdings of shares</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
<td>0.628</td>
<td>0.698</td>
<td>-9.4866</td>
</tr>
</tbody>
</table>
### APPENDIX 2

**STATISTICAL RESULTS OF BANKRUPTCY PREDICTION**

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistics</td>
<td>97.42%</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
<td>0.998</td>
<td>0.998</td>
<td>97.34%</td>
<td>0.973</td>
<td>0.973</td>
<td>0.973</td>
<td>0.997</td>
<td>0.998</td>
</tr>
<tr>
<td>Bayes</td>
<td>87.00%</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.917</td>
<td>0.918</td>
<td>85.30%</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.917</td>
<td>0.918</td>
</tr>
<tr>
<td>KNN</td>
<td>89.42%</td>
<td>0.898</td>
<td>0.898</td>
<td>0.898</td>
<td>0.958</td>
<td>0.949</td>
<td>84.66%</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.914</td>
<td>0.918</td>
</tr>
<tr>
<td>ANN</td>
<td>96.70%</td>
<td>0.967</td>
<td>0.967</td>
<td>0.967</td>
<td>0.996</td>
<td>0.997</td>
<td>97.12%</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SVM</td>
<td>96.00%</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>93.53%</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>95.98%</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>95.53%</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### APPENDIX 3

**STATISTICAL RESULTS OF BANKRUPTCY PREDICTION BY BAGGING METHOD**

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistics</td>
<td>97.23%</td>
<td>0.973</td>
<td>0.973</td>
<td>0.973</td>
<td>0.998</td>
<td>0.998</td>
<td>97.23%</td>
<td>0.972</td>
<td>0.972</td>
<td>0.972</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>Bayes</td>
<td>87.05%</td>
<td>0.871</td>
<td>0.871</td>
<td>0.871</td>
<td>0.937</td>
<td>0.928</td>
<td>85.94%</td>
<td>0.862</td>
<td>0.862</td>
<td>0.862</td>
<td>0.917</td>
<td>0.917</td>
</tr>
<tr>
<td>KNN</td>
<td>88.79%</td>
<td>0.892</td>
<td>0.892</td>
<td>0.892</td>
<td>0.947</td>
<td>0.94</td>
<td>84.13%</td>
<td>0.841</td>
<td>0.841</td>
<td>0.841</td>
<td>0.884</td>
<td>0.884</td>
</tr>
<tr>
<td>ANN</td>
<td>97.95%</td>
<td>0.979</td>
<td>0.979</td>
<td>0.979</td>
<td>0.999</td>
<td>0.999</td>
<td>97.55%</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>SVM</td>
<td>95.87%</td>
<td>0.959</td>
<td>0.959</td>
<td>0.959</td>
<td>0.977</td>
<td>0.967</td>
<td>95.31%</td>
<td>0.953</td>
<td>0.953</td>
<td>0.953</td>
<td>0.973</td>
<td>0.973</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>97.20%</td>
<td>0.972</td>
<td>0.972</td>
<td>0.972</td>
<td>0.995</td>
<td>0.994</td>
<td>97.98%</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### APPENDIX 4

**STATISTICAL RESULTS OF BANKRUPTCY PREDICTION BY BOOSTING METHOD**

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistics</td>
<td>97.42%</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
<td>0.998</td>
<td>0.998</td>
<td>97.34%</td>
<td>0.973</td>
<td>0.973</td>
<td>0.973</td>
<td>0.997</td>
<td>0.998</td>
</tr>
<tr>
<td>Bayes</td>
<td>93.03%</td>
<td>0.931</td>
<td>0.931</td>
<td>0.931</td>
<td>0.984</td>
<td>0.984</td>
<td>92.33%</td>
<td>0.924</td>
<td>0.924</td>
<td>0.924</td>
<td>0.983</td>
<td>0.984</td>
</tr>
<tr>
<td>KNN</td>
<td>89.50%</td>
<td>0.895</td>
<td>0.895</td>
<td>0.895</td>
<td>0.927</td>
<td>0.911</td>
<td>84.66%</td>
<td>0.847</td>
<td>0.847</td>
<td>0.847</td>
<td>0.885</td>
<td>0.884</td>
</tr>
<tr>
<td>ANN</td>
<td>96.70%</td>
<td>0.967</td>
<td>0.967</td>
<td>0.967</td>
<td>0.979</td>
<td>0.977</td>
<td>97.12%</td>
<td>0.971</td>
<td>0.971</td>
<td>0.971</td>
<td>0.971</td>
<td>0.971</td>
</tr>
<tr>
<td>SVM</td>
<td>96.54%</td>
<td>0.965</td>
<td>0.965</td>
<td>0.965</td>
<td>0.995</td>
<td>0.995</td>
<td>96.27%</td>
<td>0.963</td>
<td>0.963</td>
<td>0.963</td>
<td>0.995</td>
<td>0.994</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>97.15%</td>
<td>0.971</td>
<td>0.971</td>
<td>0.971</td>
<td>0.996</td>
<td>0.995</td>
<td>96.70%</td>
<td>0.967</td>
<td>0.967</td>
<td>0.967</td>
<td>0.997</td>
<td>0.997</td>
</tr>
</tbody>
</table>
APPENDIX 5
STATISTIC RESULTS OF BANKRUPTCY PREDICTION BY DECISION TREE

X13 <= 1.4654
|   X5 <= 0.511
|     X5 <= 0.3402
|       X13 <= 0.3798
|       |   X24 <= 0.2247: Normal (53.0)
|       |   X24 > 0.2247
|       |     X1 <= 3.1641: Bankruptcy (6.0)
|       |     X1 > 3.1641: Normal (8.0)
|     X13 > 0.3798: Normal (305.0)
|   X5 > 0.511
|     X13 <= 1.6915: Bankruptcy (6.0/1.0)
|     X5 <= 1.0544
|      X1 <= 1.2168: Bankruptcy (277.0/1.0)
|      X1 > 1.2168
|   X13 <= 1.2471
|     X1 <= 1.807: Bankruptcy (109.0/1.0)
|     |   X5 <= 1.617: Normal (6.0/1.0)
|     |   X5 > 1.617: Bankruptcy (3.0)
|     X13 > 1.2471
|       X1 <= 1.5
|       |   X13 <= 1.427
|       |     X1 <= 1.1012: Bankruptcy (13.0)
|       |     X1 > 1.1012
|       |       X5 <= 2.8411: Normal (11.0/1.0)
|       |       X5 > 2.8411: Bankruptcy (3.0)
|       X13 > 1.6915: Normal (121.0)
|     |   X13 <= 1.0733
|     |     X5 <= 0.9134
|     |     |   X14 <= -4.27: Normal (26.0/1.0)
|     |     |   X14 > -4.27: Bankruptcy (3.0/1.0)
|     |     X5 > 0.9134: Bankruptcy (6.0/1.0)
|     |   X13 > 1.0733: Normal (65.0)
|     |   X5 > 1.0544
|     |     X1 <= 1.2168: Bankruptcy (277.0/1.0)
|     |     X1 > 1.2168
|     X13 <= 1.2471
|       X1 <= 1.807: Bankruptcy (109.0/1.0)
|       |   X5 <= 1.617: Normal (6.0/1.0)
|       |   X5 > 1.617: Bankruptcy (3.0)
|     |   X13 > 1.2471
|       |   X1 <= 1.5
|       |     X13 <= 1.427
|       |     |   X1 <= 1.1012: Bankruptcy (13.0)
|       |     |   X1 > 1.1012
|       |     |       X5 <= 2.8411: Normal (11.0/1.0)
|       |     |       X5 > 2.8411: Bankruptcy (3.0)
|     |     X13 > 1.6915: Normal (121.0/2.0)
|     |   X13 <= 1.0733
|     |     X5 <= 0.9134
|     |     |   X14 <= -4.27: Normal (26.0/1.0)
|     |     |   X14 > -4.27: Bankruptcy (3.0/1.0)
|     |     X5 > 0.9134: Bankruptcy (6.0/1.0)
|     |   X13 > 1.0733: Normal (65.0)
|     |   X5 > 1.0544
|     |     X1 <= 1.2168: Bankruptcy (277.0/1.0)
|     |     X1 > 1.2168
|     X13 <= 1.2471
|       |   X1 <= 1.807: Bankruptcy (109.0/1.0)
|       |     X5 <= 1.617: Normal (6.0/1.0)
|       |     X5 > 1.617: Bankruptcy (3.0)
|       |     X13 > 1.2471
|       |       |   X1 <= 1.5
|       |       |     X13 <= 1.427
|       |       |     |   X1 <= 1.1012: Bankruptcy (13.0)
|       |       |     |   X1 > 1.1012
|       |       |     |       X5 <= 2.8411: Normal (11.0/1.0)
|       |       |     |       X5 > 2.8411: Bankruptcy (3.0)
|       |     |     X13 > 1.6915: Normal (121.0/2.0)
|       |     |   X13 <= 1.0733
|       |     |     X5 <= 0.9134
|       |     |     |   X14 <= -4.27: Normal (26.0/1.0)
|       |     |     |   X14 > -4.27: Bankruptcy (3.0/1.0)
|       |     |     X5 > 0.9134: Bankruptcy (6.0/1.0)
|       |     |   X13 > 1.0733: Normal (65.0)
|     |     X5 > 1.0544
|     |       X1 <= 1.2168: Bankruptcy (277.0/1.0)
|     |       X1 > 1.2168
|     X13 <= 1.2471
|       |   X1 <= 1.807: Bankruptcy (109.0/1.0)
|       |     X5 <= 1.617: Normal (6.0/1.0)
|       |     X5 > 1.617: Bankruptcy (3.0)
|       |     X13 > 1.2471
|       |       |   X1 <= 1.5
|       |       |     X13 <= 1.427
|       |       |     |   X1 <= 1.1012: Bankruptcy (13.0)
|       |       |     |   X1 > 1.1012
|       |       |     |       X5 <= 2.8411: Normal (11.0/1.0)
|       |       |     |       X5 > 2.8411: Bankruptcy (3.0)
|       |     |     X13 > 1.6915: Normal (121.0/2.0)