Predicting the Survival or Failure of Click-and-Mortar Corporations

Indranil Bose
The University of Hong Kong
bose@business.hku.hk

Raktim Pal
James Madison University
palrx@jmu.edu

Abstract

With the boom in e-business, several corporations have emerged in the late nineties that have primarily conducted their business through the Internet and the Web. They have come to be known as the dotcoms or click-and-mortar corporations. The success of these companies has been short lived and many of these companies have failed rapidly in a short span of 4-5 years. This research is an investigation of the burst of the dotcom bubble from a financial perspective. Data from the financial statements of several survived and failed dotcom companies is used to compute financial ratios, which are analyzed using two data mining techniques - discriminant analysis (DA) and neural networks (NN) to find out whether they can predict the financial fate of companies. Neural networks perform better than discriminant analysis in predicting survival or failure of click-and-mortar corporations. The key financial ratios that play a major role in the process of prediction are identified.

1. Introduction

The late nineties have witnessed a phenomenal revolution – the growth of the Internet based companies. Companies have started selling merchandise on the Web ranging from pet supplies to garden tools to cosmetics. Some of these companies have a physical presence in retailing like Barnes and Nobles, Wal-Mart etc. These are the brick-and-mortar corporations. Many of these companies also have online subsidiaries, which produce their own financial reports separate from the parent company. There are others who are essentially pure play Internet and have only an online existence. These are known as the dotcoms or the click-and-mortar corporations. As the number of Internet users across the world increases at a steady rate, retailing over the Internet continues to gain popularity. Forrester Research has predicted that e-commerce sales, growing at a steady rate of nineteen percent per year, will increase to $229.0 billion in 2008 from $95.7 billion in 2003 [6].

However, the phenomenal growth in Internet related e-tailing has suffered a major setback in the early years of the new millennium. In 2001, Forrester Research had indicated that weak financial strength, increased competition, and investor flight would drive most click-and-mortar companies out of business by 2001. At the same time, the Gartner Group had predicted that as many as ninety five per cent of all click-and-mortar companies will fail by 2002 [3]. Sure enough, the prediction of the financial pundits has come true and the dotcom bubble has burst. Within a very short time, several corporations that have seen phenomenal growth in their stock prices in the late nineties have gone out of business.

Several factors are thought to be responsible for this demise and as indicated by Sharma [9], these have included among others, the dotcoms' inability to improve revenues and earnings, failure to post profits, attempt to capture a major market share in the smallest possible time, and tendency to operate in limited geographical areas. The objective of this research is to conduct a rigorous study of the financial statements of dotcom companies and to discover the factors responsible for the survival or attrition of these companies. Data mining techniques are used to find out how good financial ratios are for predicting performance of click-and-mortar corporations and to identify those financial ratios that are most predictive of financial health of click-and-mortar companies based on historical financial data.

The results of this research may offer directives to emerging companies that plan to do business on the Internet in the coming years. It will be able to provide guidelines about which financial factors to monitor closely for long-term competitive advantage in the market. It will be extremely useful for new as well as seasoned investors who plan to invest in similar companies. It can act as a guide to investors on what factors to look for to ascertain the financial health of a click-and-mortar company before they make the decision to purchase stocks of such a corporation.

2. Literature Review

Due to the importance of predicting the financial health of an organization, this is a widely researched area. Different approaches and techniques have been used for forecasting the likelihood of failures as indicated by Wong et al. [10] and it is quite common to use financial ratios. The interest is this area has been spurred by Altman [1] with the use of multiple discriminant analysis with five financial ratios, to predict the risk of failure. In addition to discriminant analysis, several other techniques have been used.

It has been often assumed that the problem of predicting survival or failure of companies based on financial ratios is similar to detecting patterns in large volumes of data. So it is quite natural to use neural networks, which are known for their success in pattern recognition in a variety of disciplines, for predicting corporate failures. Examples of the use of neural networks for prediction of failures of companies have been included in [5], [7], and [8]. Researchers have often wondered about the combination of financial ratios to be used as an input to the classification system. Many researchers have used the five financial ratios (working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value equity/book value of total debt, and sales/total assets) whether using neural networks or other traditional approaches. However, there are existing research papers where more than five financial ratios have been used. In addition to choosing different number of financial ratios, researchers have often differed in their choice of nature of firms in their studies. Some of the criteria that have been used for the choice of firms have included asset or capital size and sales.

From the above discussion it can be meaningfully concluded that several competitive techniques have been used for studying the problem of prediction of failures of corporations using financial ratio analysis. But to the best of the authors' knowledge, there has been no study on prediction of survival or failure of click-and-mortar companies. Click-and-mortar companies are a recent phenomenon, and it is conjectured that since click-and-mortar corporations are formed differently and also function quite differently from traditional corporations, a different set of financial factors may be responsible for their demise than what has been reported in the literature. It is also not known whether it is possible to use statistical techniques like discriminant analysis or artificial intelligence techniques like neural networks for prediction of financial health of such companies.

3. Research Methodology

In this paper the click-and-mortar corporations and online subsidiaries of brick-and-mortar corporations are studied to find out whether they will survive or cease to exist. Data mining techniques are used to conduct this study. Data mining is defined as identifying patterns from large volumes of data. Several techniques have been suggested in the literature for data mining in different domains of application [2]. In this paper two different data mining techniques are used for exploring the data on click-and-mortar corporations. The path of research followed in this paper includes some key steps which are described below.

3.1. Step 1: Data Collection

For this study, financial information for two hundred and forty click-and-mortar publicly traded corporations for the last decade (1993-2003) is collected from the WRDS (Wharton Research Data Services) database. Financial data for these companies have included such details as price to earnings ratio, dividends per share, current assets and liabilities, net revenues etc. Based on the collected financial data, twenty four financial ratios are calculated for the click-and-mortar corporations. The list of the twenty four financial ratios that is used in this paper is given in Table 1. The choice of ratios is governed by the ratios that have been popularly used in the literature and have been identified by Dimitras et al. [4]. In Table 1, variables 1-15 are obtained from Dimitras et al. [4] and variables 16-24 are created by authors in order to identify new financial ratios that might be important, specifically for click-and-mortar corporations.

3.2. Step 2: Data Cleaning and Preprocessing

The collected data is categorized into two classes – firms that have survived and firms that have failed. A firm is said to have survived if it is publicly traded (as of August 2003) and if the stock price is more than 1 cent and these types of firms are indicated by '1'. If a firm previously has traded publicly but currently is trading at a price below one cent then it is considered to have failed and is indicated by '0'. The data set consists of information on financial ratios for two hundred and forty click-and-mortar corporations of which half (one hundred and twenty firms) have failed and the other half (one hundred and twenty firms) have survived. The next step involves formation of random samples from the data. Each random sample is divided into two parts – the 'training' random sample and the 'testing' random sample. The ratio of training to testing

random sample size is taken as 80:20 and care is taken to include all two hundred and forty click-and-mortar corporations in either the training sample or the testing sample, so that no data is lost in the process. In this way, ten random training samples and ten random testing samples are created.

3.3. Step 3: Data Analysis

This is the most important step in data mining. In this paper, three different techniques are used. The problem under study is a classification problem where the inputs consist of the twenty four financial ratios of the company and the output is a binary output. The goal is to classify the firms with unknown outputs (i.e., the testing sample) as '0' or '1' and to identify the variables that play the most important role in the process.

The main goal of DA is to construct a discriminant function for the object based on a weighted linear combination of the inputs, and to classify the object into one of the output classes based on the computed discriminant 'score'. In this paper linear DA is used.

A NN is trained by using examples of known classification. The network learns the classification function by iteratively adjusting the weights associated with the network connections. A popular algorithm for supervised learning in NN is the back-propagation algorithm. The strength of a NN is that it can learn any classification function and it gives good classification results in the presence of noise in the data.

3.4. Step 4: Interpretation and Evaluation

In the last step of the data mining process the numerical results that are obtained from experimenting with the two methods are compared to find the method that gives the better classification accuracy in most cases. Next, the variables (financial ratios) that play a key role in the prediction process are identified and intuitive explanations of the observed phenomena are provided using basic knowledge of accounting and finance.

4. Numerical Experimentation

The first set of experiments involves the use of the DA technique for the classification problem. DA is conducted using SAS System for Windows Version 8.2. Initially, all twenty four financial ratios are considered as input variables. Initial experimentation indicates that better results are likely using a subset of the twenty four input variables. Hence, stepwise DA is used to identify the subset of input variables. Both forward and backward elimination methods are used.

Forward elimination starts with one variable and keeps on adding other variables till no more variables can be added to the model. Using forward elimination three variables are found to be most predictive. These variables are CF/S, S/TA, and OI/MC. The backward elimination procedure starts with all twenty four variables and keeps on eliminating variables till no variables can be eliminated. Backward elimination has given the following list of predictive variables - TD/TA, EBIT/TA, CF/S, RE/TA, S/TA, GP/TA, CA/TA, OI/S, OI/MC, CASH/S, NI/(TA-TL). Stepwise DA is followed by systematically modifying the best models that are obtained using forward and backward elimination. The top twenty models that are obtained using DA are listed in Table 2. In Table 2, the type I accuracy is calculated by identifying the percentage of cases where a '0' (failed) is correctly identified as a '0' (failed) and the type II accuracy is calculated by identifying the percentage of cases where a '1' (survived) is correctly identified as a '1' (survived). For any particular model, the type I and type II accuracies are obtained by taking the average over all samples. The total accuracy is the average of the type I and type II accuracies for a model. The total training accuracies are higher than testing accuracies in general and it is observed that model 2 consisting of variables TD/TA, OI/TA, CF/S, RE/TA, S/TA, GP/TA, and OI/MC provides the best total accuracy (i.e., 69.17%) for testing samples. It is also observed that type I accuracies are lower than type II accuracies in general for training though the same doesn't always hold for testing. In Figure 1, the frequency of occurrence of the twenty four variables in the top twenty models is shown. From the graph it can be seen that variables RE/TA, OI/MC, and S/TA are the three most frequently occurring variables.

The next set of experiments involves the use of NN. The experimentation is conducted using the software called Predict by Neuralworks. Predict uses the back-propagation algorithm. Using all possible choices of parameters that Predict allows, it is possible to identify the top twenty models. Like DA, the accuracy figures are obtained using averages over all random samples. The top twenty models (in terms of total testing accuracy) that have been obtained using NN are listed in Table 3. The best testing accuracy (i.e., 75.83%) is obtained for model 8 using eleven variables: TD/TA, OI/TA, NI/TA, CF/TD, RE/TA, QA/TA, P/E, S/MC, LTD/TA, OI/S, and OI/MC. It is observed that, in comparison to DA, NN has produced results that are more accurate with respect to the testing samples. An interesting observation is that, unlike DA, in case of NN the total training accuracies vary more and higher training accuracy doesn't always lead to higher testing accuracy. This is probably due to

overtraining in some models. A graph showing the frequency of occurrence of the different variables for the top twenty models using NN is shown in Figure 2. It is observed that the most frequently occurring variable is RE/TA. It must be noted here that RE/TA is also one of the most frequently occurring variables for DA as well. This indicates the importance of this factor for predicting the financial health of a company. This finding also corroborates to the research findings of previous researchers [4] who have identified RE/TA as an important variable for predicting financial health of traditional (i.e., not click-and-mortar) corporations.

5. Conclusion

In this paper DA and NN are used to find out whether it is possible to predict the survival or failure of click-and-mortar corporations based on financial ratios. On the whole, NN seems to perform better than DA. It is observed that in most cases the average type II accuracy is higher than the average type I accuracy. Both methods suffer from overtraining to an extent and the testing accuracies are mostly lower than the training accuracies. An important contribution of this paper is identification of the predictive financial ratios. The financial ratios RE/TA, OI/MC, OI/TA, and CF/S are commonly identified as highly predictive factors by both DA and NN. This clearly indicates that cash flow and earnings are very important factors and can be used for distinguishing between survived and failed click-and-mortar corporations. It is reported in [9], that John Hagel of McKinsey and Company Inc. have stated in a dot-com bootcamp that "revenues in E-commerce are a distraction". Many dotcoms fail due to their inability to generate revenues and cash flows in the short run. The findings of this research also point in the same direction.

There are some limitations to this research. It explores the problem of classification using two techniques only. Also, the results obtained are dependent on the definition used for failure of dot-coms. Use of a new definition may lead to different results. Also, the choice of the financial ratios is only based on the authors' domain knowledge, and available literature. However, there is a need to validate the use of these ratios by expert financial analysts. Again, the results obtained in this research are dependent on the available data. WRDS lists only dot-coms with significant market capitalization and that list may not be exhaustive. There is a need to conduct a more in-depth study in future with a more exhaustive sample of all available financially healthy and unhealthy dot-coms.

In future more extensive studies can be conducted using other methods of data mining such as Bayesian learning, decision trees, rough sets, support vector machines etc. so that their performance can be compared to those of DA and NN. In particular the method of rough sets appears promising because it can lead to the discovery of rules which can relate the independent and dependent variables. In this paper we have used overall accuracy of prediction as a measure for deciding which models are good or not. In future other benchmarks such as cost of misclassification may be used for evaluating the models. Another future research direction may be to use ideas from the field of organizational ecology to study the survival and mortality of corporations.

References

[1] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *The Journal of Finance, 23,* 1968, pp. 589-609.

[2] I. Bose and R. K. Mahapatra, "Business data mining – a machine learning perspective", *Journal of Information & Management, 39(3),* 2001, pp. 211-225.

[3] E. Cane, "Online laughing stocks: betting on dotcom failures takes off", Available online at: http://www.financeasia.com/articles/428EB6C0-4BC2-11D4-8C0F0008C72B383C.cfm, June 28, 2000.

[4] A. I. Dimitras, S. H. Zanakis, and C. Zopounidis, "A survey of business failures with an emphasis on prediction methods and industrial applications", *European Journal of Operational Research, 90,* 1996, pp. 487-513.

[5] R. C. Lacher, P. K. Coats, S. C. Sharma, and L. F. Fant, "A neural network for classifying the financial health of a firm", *European Journal of Operational Research, 85,* 1995, pp. 53-65.

[6] L. Rush, "US e-commerce to see significant growth by 2008", Available online at: http://www.ecommerce-guide.com/news/research/article.php/10371_2245631, 2003.

[7] J. R. Shah and M. B. Murtaza, "A neural network based clustering procedure for bankruptcy prediction", *American Business Review, 18(2),* 2000, pp. 80-86.

[8] R. Sharda and R. L. Wilson, "Neural network experiments in business failure forecasting: predictive performance measurement issues", *International Journal of Computational Intelligence and Organizations, 1(2),* 1996, pp. 107-117.

[9] A. Sharma, "Dot-coms' coma", *The Journal of Systems and Software, 26,* 2001, pp. 101-104.

[10] B. K. Wong, V. S. Lai, and J. Lam, "A bibliography of neural network business applications research: 1994-1998", *Computers and Operations Research, 27,* 2000, pp. 1045-1076.

Table 1. List of financial ratios used in data analysis

Variables	Symbols	Description
1	WC/TA	Working Capital/Total Assets
2	TD/TA	Total Debt/Total Assets
3	CA/CL	Current Assets/Current Liabilities
4	OI/TA	Operating Income/Total Assets
5	NI/TA	Net Income/Total Assets
6	CF/TD	Cash Flow/Total Debt
7	QA/CL	Quick Assets/Current Liabilities
8	CF/S	Cash Flow/Sales
9	RE/TA	Retained Earnings/Total Assets
10	S/TA	Sales/Total Assets
11	GP/TA	Gross Profit/Total Assets
12	NI/SE	Net Income/Shareholders' Equity
13	C/TA	Cash/Total Assets
14	I/S	Inventory/Sales
15	QA/TA	Quick Assets/Total Assets
16	P/E	Price Per Share/Earnings Per Share
17	S/MC	Sales/Market Capitalization
18	CA/TA	Current Assets/Total Assets
19	LTD/TA	Long Term Debt/Total Assets
20	OI/S	Operating Income/Sales
21	OI/MC	Operating Income/Market Capitalization
22	C/S	Cash/Sales
23	CA/S	Current Assets/Sales
24	NI/(TA-TL)	Net Income/(Total Assets – Total Liabilities)

Table 2. Top twenty models obtained by DA

Model No.	Variables in Model	Training			Testing		
		Type I	Type II	Total	Type I	Type II	Total
1	2,4,8,9,10,11,21,24	69.73	73.11	71.46	69.19	67.45	68.33
2	2,4,8,9,10,11,21	69.32	73.11	71.25	69.59	68.69	69.17
3	2,4,8,9,10,11,18,20,21	64.54	72.63	68.54	69.48	62.78	66.04
4	2,4,8,,10,11,21,22	69.32	73.11	71.25	67.74	67.45	67.71
5	2,4,8,9,10,11,18,20,21,24	65.77	73.05	69.38	69.44	62.37	65.83
6	2,4,8,9,10,11,21,22,24	67.05	75.02	71.04	65.11	69.52	67.50
7	2,4,8,9,10,11,18,21	63.27	71.86	67.55	66.56	64.67	65.63
8	4,8,9,10,11,21,24	70.89	67.88	69.43	67.18	63.65	65.21
9	2,4,8,9,10,11,18,20,21,22	61.96	78.51	70.16	64.12	67.58	66.04
10	2,4,8,9,10,11,18,21,22,24	62.14	74.70	68.39	63.06	67.52	65.42
11	2,4,8,9,10,11,18,21,24	63.67	72.37	68.02	65.69	65.18	65.42
12	4,8,9,10,11,21	70.15	67.67	68.96	67.51	63.56	65.42
13	4,8,9,10,11,18,20,21,22,24	61.95	75.26	68.54	63.96	67.35	65.83
14	2,4,8,9,10,11,18,21,22	62.14	74.39	68.23	64.74	66.68	65.83
15	2,4,8,9,10,21,24	71.99	66.01	69.11	70.86	63.15	66.88
16	2,4,8,9,10,11,20,21,22	62.88	79.37	71.04	61.38	72.87	67.29
17	2,4,8,9,10,21	71.29	66.01	68.75	70.29	63.90	67.08
18	9,21,10,2,5	67.14	66.02	66.67	66.43	63.92	65.21
19	2,4,8,9,10,11,20,21,22,24	62.87	79.58	71.15	59.50	73.35	66.67
20	4,8,9,10,11,18,20,21,22	61.75	75.37	68.49	64.00	65.21	64.79

Table 3. Top twenty models obtained by NN

Model No.	Variables in Model	Training			Testing		
		Type I	Type II	Total	Type I	Type II	Total

|							
1	4,8,9,20	72.60	83.96	78.28	66.93	79.06	73.33
2	all	82.03	88.13	85.08	72.15	77.38	75.42
3	2,5,9,17,21	77.02	78.03	77.52	72.02	75.22	73.96
4	3,4,8,9,10,11,12,17,20,21	98.68	99.40	99.04	75.62	68.21	71.88
5	1,4,5,7,8,9,10,11,12,15,20,21,23,24	84.34	88.24	86.29	74.88	69.56	72.29
6	2,5,8,9,10,11,12,19,20,21	82.78	84.25	83.51	73.27	72.80	73.33
7	4,9,15,17	74.63	77.09	75.86	67.61	73.49	71.04
8	2,4,5,6,9,15,16,17,19,20,21	76.07	83.83	79.97	72.65	78.00	75.83
9	4,7,8,9,17,20	81.50	83.72	82.61	72.07	71.60	72.08
10	4,7,9,12,14,17,18,20,23	88.03	90.39	89.21	71.66	73.95	73.33
11	3,4,5,7,8,9,12,15,16,17,20,21,22	93.65	93.15	93.40	72.16	73.52	73.33
12	1,4,8,9,11,12,13,16,17,20,21,22	93.77	94.64	94.21	73.79	70.97	72.71
13	1,4,5,8,9,11,16,17,18,20,21	99.70	99.57	99.63	72.93	69.94	71.88
14	2,4,5,7,8,9,10,11,12,16,20,21,22,23,24	95.56	96.23	95.90	73.08	70.00	71.67
15	2,7,8,9,10,11,17,20,21	85.16	89.89	87.53	73.35	72.95	73.54
16	4,9	72.33	80.96	76.65	67.19	76.15	72.29
17	5,7,9,10,11,16,17,18,20,21,22	76.08	84.43	80.26	72.13	77.06	74.79
18	4,8,9,14,15,16,19,20,21,23	86.71	89.93	88.32	69.13	72.30	70.83
19	2,4,6,7,8,9,16,17,24	84.85	85.27	85.06	69.48	72.44	71.25
20	1,2,4,7,8,9,11,12,17,20	94.55	95.50	95.03	72.27	70.82	71.88
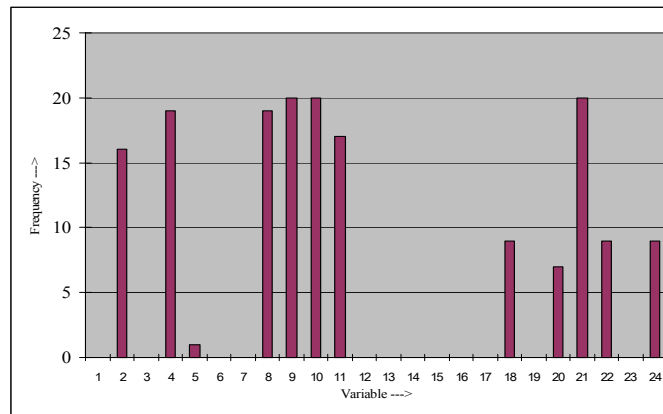
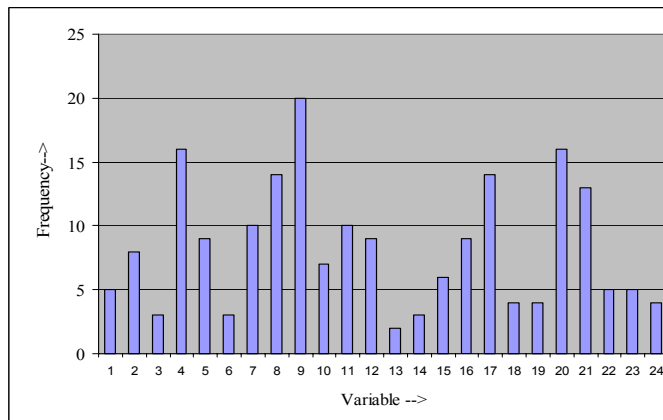**Figure 1. Frequency of variables in top twenty models obtained by DA**

Figure 2. Frequency of variables in top twenty models obtained by NN