Learning the Relationship Between Corporate Governance and Company Performance Using Data Mining

Darie Moldovan(✉) and Simona Mutu

Faculty of Economics and Business Administration, Babeş-Bolyai University, Theodor Mihali 58-60, 400591 Cluj-Napoca, Romania
{darie.moldovan,simona.mutu}@econ.ubbcluj.ro

Abstract. The objective of this paper is to identify the relationship between corporate governance variables and firm performance by employing data mining methods. We choose two dependent variables, Tobin's Q ratio and Altman Z-score, as measures for the companies' performances and apply machine learning techniques on the data collected from the components companies of three major stock indexes: S&P 500, STOXX Europe 600 and STOXX Eastern Europe 300. We use decision trees and logistic regressions as learning algorithms, and then we compare their performances. For the US components, we found a positive connection between the presence of women in the board and the company performance, while in Western Europe that it is better to employ a larger audit committee in order to lower the bankruptcy risk. An independent chairperson is a positive factor related to Altman Z-score, for the companies from Eastern Europe.

Keywords: Machine learning · Corporate governance · Firm performance · Classification · Logistic regression · Decision Trees

1 Introduction

The study of corporate governance has increased continuously in the last twenty years, benefiting from time to time boosts such as the large insolvencies of the early 2000s (Enron, WorldCom, Parmalat) or the effects of the recent financial crisis, which shown multiple weaknesses in the corporate governance systems of the companies.

The corporate governance rules are different from country to country, but major differences can be found between developed markets and emerging markets. The emerging markets countries implemented a series of reforms to their corporate governance systems in terms of legal measures, corporate governance codes or cross listings [13]. The integration of emerging markets within the global financial systems brought new investment opportunities, but also challenges in capital allocation. The financial crisis made portfolio managers to look more closely to the quality of corporate governance before taking their investment decisions.

P. Perner (Ed.): MLDM 2015, LNAI 9166, pp. 368–381, 2015.
DOI: 10.1007/978-3-319-21024-7_25

Our objective is to demonstrate the utility of data mining methods in identifying the relationships between corporate governance and performance. As a proxy for corporate performance we use two dependent variables: the company's performance evaluation (represented by the Tobin's Q ratio [45]) and the financial distress level (according to Altman Z-score [2]). Our testbed is represented by the data collected from three corporate governance systems: United States, Western Europe and Eastern Europe. We used the data from the companies composing the following three stock indexes: S&P500, STOXX Europe 600 and STOXX Eastern Europe 300. We selected 50 attributes, both financial indicators and corporate governance variables, in order to create a "competition" between them to explain the Tobin's Q ratio and Altman Z-score.

We used classification algorithms and regressions in order to demonstrate the utility of data mining techniques in creating portraits of different typologies for the companies analyzed. We employed the software Weka 3.7 to make our experiments, a powerful open source tool developed by the University of Waikato, New Zealand [25].

Our findings reveal that for each dataset the performances of algorithms are close to each other, with Adaboost M1 being the most consistent in terms of classification accuracy in most of the cases. In case of American companies the Tobin's Q ratio is positively influenced by the percentage women in the board, while an independent lead director and a financial leverage higher than 2.5 generates a higher risk of bankruptcy. For the Western European companies, the presence of an independent CEO and large percentage of women in the board could lead to weaker performances, being negatively correlated with Tobin's Q. As for the Eastern European companies, a smaller age range for the board members and a financial leverage less than 4 enhances companies' performance, while an independent chairperson or a woman as CEO reduces bankruptcy probability.

The obtained results can be useful either for investment decision or as a guide for board companies looking to improve their corporate governance policies, both in developed, but also in emerging markets.

The remaining of this paper is structured as follows. Section 2 is dedicated to describe the related work in this field. First we described the research approaches in using the Tobin's Q ratio and the Altman Z-score in relation with the corporate governance. Second, we review the literature concerning the application of data mining methods in the field of corporate governance, demonstrating the usefulness of the algorithms in solving research questions in this area. Section 3 describes the data used in this research. Section 4 describes in detail the experiments conducted. Section 5 presents the results. Finally, the conclusions of the paper are described in Sect. 6.

2 Related Work

Our paper connects several strands of literature. We relate to the studies that focus on the study of executives compensations influence on companies' performance and on the importance of the board members' independence [1,7,43]. The separation of the executive management from the stakeholders can lead sometimes to a tendency of the managers to accumulate wealth. The participation of the shareholders in the board can have a positive effect on the company's performance.

In other studies, Guest [24] found a strong negative impact of the board size on the profitability of the company and the Tobin's Q ratio; Li [34] determined that independent directors' presence is positively related with the performance of the company while Lückerath-Rovers [36] associated positively the presence of women on boards with the companies' performances.

To evaluate the firm performance, a widely accepted indicator as reference by many researchers in this field is the Tobin's Q. Several studies were made around the opportunity to use this ratio as a measure of intangible assets of the company [5,6,11,33], but also negative arguments were brought by Dybvig [17].

Regarding the nexus between governance and performance in emerging and developed markets, Gibson [23] suggests that corporate governance is ineffective in emerging markets. Claessens and Djankov [12] found that firms with concentrated and foreign ownership are more profitable, registering a higher labor productivity.

Further, our research is connected to the studies on companies' bankruptcy. The probability of default could be enhanced by the "dominant" leadership style [27], the proportion of outside directors [16], the average number of directorships in other firms held by outside directors [18]and by a small and younger board [39]. During crisis periods, [40] found that companies which change their CEO and with a low percentage of shares held by managers are more likely to fail than the others.

The literature proposes several methods to measure the companies' default like the Z-score of Altman [2], the O-score of Ohlson [38] and the structural model default probabilities of Merton [37], among others. However, a large number of studies advocate for the Altman Z-score [4,5,8]. Also, we link to the research on machine learning applications in this area. Related studies are represented by the works of Creamer [14,15] who uses boosting in order to examine the relationship between the Tobin's Q and variables representative for south American banks, but also for the S&P 500 component companies. Tsai [46] focuses on applying data mining in order to observe the intangible assets value on companies from South Korea. Lu [35] is using association rules to address the same problem.

Other authors are focusing on detecting credit risk, creating credit scoring systems with the help of data mining, and using the Altman Z-score, an indicator accepted on a large scale by researchers and professionals in the field [10,29,30].

We contribute to the existing literature in several ways. First, we analyze a unique sample of listed companies which are included in three major stock indexes: S&P 500, STOXX Europe 600 and STOXX Eastern Europe 300, representative for developed and emerging markets. Second, we explore a large and unique set of variables that characterize different corporate governance and accounting aspects. Third, we add to the literature on data mining and corporate performance, by applying machine learning techniques to investigate the impact of companies' governance and financial characteristics on their performance.

3 Data Preprocessing

We collected data for the companies composing three stock indexes: S&P500 (SPX), STOXX Europe 600 (SXXP) and STOXX Eastern Europe 300 (EEBP). The data source was Bloomberg, from where we extracted 52 variables for 1400 companies. The chosen dependent variables were Tobin's Q ratio, as a measure for the performance of the companies and Altman Z-score, as a measure of the financial distress.

Tobin's Q represents the ratio between the market value of a company and the replacement cost of the firm's assets and is computed using the following formula:

$$\begin{aligned} Tobin's\,Q = (&Market\,Cap + Total\,Liabilities + Preferred\,Equity \\ &+ Minority\,Interest)/Total\,Assets \end{aligned} \tag{1}$$

In order to find the relationship between these variables and the corporate governance, we chose 50 variables, both from the corporate governance and accounting area. The accounting indexes are well known for their predictive power, being used by researchers in detecting patterns inside the financial statements of the companies [2,30,42].

Among the corporate governance variables we can mention the board size, the number of independent directors, the presence of women on board, the percent of female executives, the average age of the board members or the percent of independent directors inside the board. Our interest is to find if there is a connection between these variables and the performance of the company, expressed by Tobin's Q ratio and, respectively the financial distress, expressed by the Altman Z-score.

The accounting variables chosen are among the most popular used in fundamental analysis of the companies. Indicators such as Operating Profit Margin, Return on Equity, Dividend Payout, Return on Capital or Price to Book ratio are all indicators frequently used to analyze the financial statements of a company. A list of some of the variables used in our study can be found in Appendix A.

During the preprocessing stage of the knowledge discovery process [20] we had to deal with the missing values inside the data set and removing outliers. We removed all the instances that had missing data on the attribute of the dependent variables or did not find enough information to calculate their values. We also removed the outliers for all the variables in order to eliminate any possible data errors.

The class variables Tobin's Q ratio and Altman Z-score were processed in order to obtain nominal variables. As suggested by Creamer [14], we discretized Tobin's Q ratio in order to obtain two classes, dividing each dataset according to its median value. In this way, a company that lies in the upper side of the median will be looked positively by the machine learning algorithms.

The Altman Z-score variable was divided into three classes, according to the definitions provided by Altman [2,3]. The indicator is computed considering the following formula:

$$\begin{aligned} Altman\,Z\,Score = 1.2 * (Working\,Capital/Tangible\,Assets) \\ + 1.4 * (Retained\,Earnings/Tangible\,Assets) \\ + 3.3 * (EBIT/Tangible\,Assets) \\ + 0.6 * (Market\,Value\,of\,Equity/Total\,Liabilities) \\ + (Sales/TangibleAssets) \end{aligned} \tag{2}$$

For the SPX and SXXP companies we used the classification values specific to the developed markets. In this case, a company that has a Z-score above 2.99 is considered to be in the "safe" zone, if the score is bellow 2.99 but above 1.81 it is considered to be in the "gray" zone and if it lies below 1.81 then it is in financial distress. The companies from the emerging markets are evaluated on a slightly different scale, where the "safe" zone begins at a score of 2.6, the "grey" zone is located between 1.1 and 2.6, while the "distress" zone is situated bellow the Z-score of 1.1.

After the preprocessing of the three datasets, the SPX dataset remained with 496 records, the SXXP dataset with 595 instances and the EEBP dataset with 297 instances.

4 Experiments

For each dataset, we performed an attribute evaluation in relation to the class, in order to eliminate all the attributes that do not offer an information gain with respect to the class. The attribute selection operation was performed only if the classification precision improved. If not, the dataset was maintained unaltered. Considering the benchmark realized by Hall [26] we used two alternative feature selection methods: CFS (Correlation based Feature Selection) and Information Gain Attribute Ranking. The first is a method that evaluates subsets of attributes rather than individual attributes. The algorithm takes into account the usefulness of the features to predict the class, considering the level of intercorrelation between them. The second method calculates the entropy before and after observing an attribute. The amount by which the entropy decreases represents the information gain provided by the attribute [41]. We used for our experiments four different algorithms: Alternating Decision Trees (ADTree), Adaboost M1 with ADTree, J48 and Simple Logistic. The Alternating Decision Tree [21,28] consists of decision nodes and prediction nodes; an instance will be classified by following all the paths for which the decision nodes are true and summing the values of the prediction nodes from the path.

The Boosting is a method introduced by Freund [22], that applies repeatedly an algorithm (named "weak learner") to different weightings of the same training set. It was first used for binary problems, but has new implementations that are capable to deal with multinomial classes. In our experiments we used the Adaboost M1 [22] implementation with ADTree as the weak learner algorithm.

Simple logistic is an algorithm designed by Sumner [44]and it works as a classifier for building linear logistic regression models. It uses LogitBoost [32] as base learners for fitting the logistic models.

Table 1. Tobin's Q as class – Algorithms performance – SPX dataset

Algorithm	Correctly classified instances	Coverage of cases (0.95 level)	Precision Class 0	Precision Class 1	ROC area
Adaboost M1	89.7177 %	91.3306 %	0.89	0.905	0.957
J48	85.2823 %	92.9435 %	0.841	0.866	0.854
Simple log	90.3226 %	98.5887 %	0.9	0.906	0.952
ADTree	88.1048 %	99.5968 %	0.865	0.898	0.941

Table 2. Tobin's Q as class – Algorithms performance – SXXP dataset

Algorithm	Correctly classified instances	Coverage of cases (0.95 level)	Precision Class 0	Precision Class 1	ROC area
Adaboost M1	88.2353 %	91.4286 %	0.891	0.874	0.946
J48	87.395 %	97.3109 %	0.871	0.877	0.874
Simple log	85.042 %	98.6555 %	0.845	0.856	0.927
ADTree	87.563 %	99.3277 %	0.881	0.87	0.948

Table 3. Tobin's Q as class – Algorithms performance – SXXP dataset

Algorithm	Correctly classified instances	Coverage of cases (0.95 level)	Precision Class 0	Precision Class 1	ROC area
Adaboost M1	81.8182 %	87.8788 %	0.823	0.813	0.889
J48	77.1044 %	100 %	0.783	0.76	0.836
Simple log	76.431 %	98.9899 %	0.78	0.75	0.861
ADTree	80.8081 %	99.6633 %	0.811	0.805	0.893

J48 is the implementation of Quinlan's C4.5 [41], and using the concept of information gain, or entropy splits the data into subsets. The greater the information gain brought by a certain attribute, the higher will be in the decision tree.

4.1 Tobin's Q Ratio as Class

The first round of experiments was conducted having the Tobin's Q ratio as class variable. Observing the robust results obtained by Creamer [15] in employing Adaboost implemented with Alternating Decision Trees, we chose the Weka algorithm Adaboost M1 with ADTree as a benchmark for our experiments related to Tobin's Q ratio. The other algorithms used were the J.48 decision tree with a 0.05 confidence factor, the Simple Logistic Regression and the ADTRee. The validation method chosen was 10-fold cross validation, as a useful way for validation when the dataset is not very large [31].

Table 4. Altman Z-score as class – Algorithms performance – SPX dataset

Algorithm	Correctly classified instances	Coverage of cases (0.95 level)	Precision Class 0	Precision Class 1	Precision Class 2	ROC area
Adaboost M1	82.9443 %	92.8187 %	0.843	0.725	0.865	0.942
J48	73.6086 %	91.2029 %	0.722	0.569	0.816	0.843
Simple log	79.1741 %	96.0503 %	0.793	0.655	0.843	0.907
LADTree	75.9425 %	94.3084 %	0.768	0.574	0.856	0.9

Table 5. Altman Z-score as class – Algorithms performance – SXXP dataset

Algorithm	Correctly classified instances	Coverage of cases (0.95 level)	Precision Class 0	Precision Class 1	Precision Class 2	ROC area
Adaboost M1	76.2681 %	93.1159 %	0.804	0.522	0.864	0.769
J48	74.2754 %	96.1957 %	0.788	0.531	0.803	0.856
Simple log	77.5362 %	98.3696 %	0.792	0.61	0.811	0.906
LADTree	73.7319 %	98.0072 %	0.793	0.53	0.807	0.893

Table 6. Altman Z-score as class – Algorithms performance – EEBP dataset

Algorithm	Correctly classified instances	Coverage of cases (0.95 level)	Precision Class 0	Precision Class 1	Precision Class 2	ROC area
Adaboost M1	63.7602 %	99.455 %	0.746	0.496	0.675	0.771
J48	65.6676 %	98.3651 %	0.79	0.615	0.598	0.8
Simple log	69.7548 %	99.1826 %	0.795	0.543	0.731	0.849
LADTree	64.5777 %	97.8202 %	0.796	0.485	0.683	0.823

To measure the performance of the algorithms we looked for the correctly classified instances and for the classification precision for every class. By observing the precision in every class (correctly classified instances/class) we monitor the consistency of the classification in detail, excluding the possibility that the global computed accuracy could hide the lack of performance in classifying a certain class. Supplementary, we considered the Coverage of cases (0.95 level) and the ROC area [19]. In Tables 1, 2 and 3 we present the compared performance of the algorithms after a tenfold cross-validation for each of the datasets.

4.2 Altman Z-Score as Class

The second part of our experiments was dedicated to the relation between the corporate governance variables and accounting indexes on one side, and the

Altman Z-score on the other side. The Altman Z-score being an indicator for bankruptcy, our objective was to find which of the variables could be considered red flags for distress, but also to identify the financial safety signals. The Altman Z-score variable had three classes: 0 for "distress", 1 for "gray", 2 for "safe". Because, generally, the number of instances in the "distress" and "gray" classes was much more smaller than those in the "safe" class, we used a re-sampling technique, named SMOTE, or Synthetic Minority Oversampling Technique [9]. This way, we re-balance the datasets by over-sampling the minority class, in order to achieve better classifier performance. This is achieved by creating synthetic minority class examples.

As the ADTree algorithm is not suitable for two class target variables, we used Adaboost M1 with LADTree [28], an adaptation of the original algorithm, in order to support multi-class. The other algorithms used are the logistic regression, implemented to support more than two values for the class variable, the J48 and the simple LADTree. In Tables 4, 5 and 6 we present the performances obtained after running the algorithms with the Altman Z-score variable as class.

5 Results Obtained

The results of the learning algorithms in terms of classification accuracy are different for each dataset used. While for the SPX and SXXP datasets the results do not differ too much, when the algorithms were applied to the EEBP set, the results were not so strong. The explanation for this behavior resides in the significant percent of missing data for the corporate governance variables collected from the Eastern Europe companies. The principles of corporate governance are not so well established for these companies and only a minority of them adopted the same good practices as the western ones.

In terms of algorithms comparison, we can note that for each dataset the performances are close to each other, with Adaboost M1 being the most consistent in terms of classification accuracy in most of the cases.

We can also note the Simple Logistic Regression algorithm that performed better than the others on the EEBP dataset with Altman Z-score as class variable (Table 6).

From a financial perspective we found interesting connections between the corporate governance variables and the two dependent variables, which was the main objective of our research.

For the American companies inside the S&P 500 index, we found a positive correlation between the percentage higher than 20 % of women in the board and the Tobin's Q ratio, but also the presence of an independent lead director in the company along with a financial leverage higher than 2.5 incur a higher risk of bankruptcy.

For the Western European companies, the presence of an independent lead director or a former CEO in the board could be a sign of weaker performances, being negatively correlated with Tobin's Q. A large percentage of women in the board could also affect negatively the performance. For the companies with large

```
J48 pruned tree
------------------

P/B <= 1.879789
|   Indep Lead Dir = Y
|   |   P/B <= 1.366616: 0 (62.37/4.78)
|   |   P/B > 1.366616
|   |   |   Fincl  l <= 2.888611: 1 (15.34/4.0)
|   |   |   Fincl  l > 2.888611: 0 (10.34/0.39)
|   Indep Lead Dir = N: 0 (161.23/8.09)
P/B > 1.879789
|   Asset <= 0.182254: 0 (23.79/2.21)
|   Asset > 0.182254
|   |   Fincl  l <= 3.874863: 1 (254.57/18.0)
|   |   Fincl  l > 3.874863
|   |   |   Oper ROE <= 29.051374
|   |   |   |   Board Size <= 13
|   |   |   |   |   Indep Directors <= 5: 0 (13.04/1.99)
|   |   |   |   |   Indep Directors > 5
|   |   |   |   |   |   Asset <= 0.98851: 0 (11.95/2.72)
|   |   |   |   |   |   Asset > 0.98851: 1 (8.5)
|   |   |   |   Board Size > 13: 0 (14.29)
|   |   |   Oper ROE > 29.051374: 1 (19.57/0.17)

Number of Leaves  :     11

Size of the tree :      21
```

Fig. 1. J48 Results for the SXXP dataset

financial leverage in order to be in the "safe" zone of the Altman Z-score it could be a good idea to adopt an Auditing Committee with more than four people.

When analyzing the Eastern European companies' data, we found that a smaller age range for the board members is positively related with the companies' performance and that a financial leverage less than 4 is needed in order to be on the upper side of the Tobin's Q ratio. To be on the "safe" zone of the Altman Z-score it is important to have an independent chairperson or even a woman as CEO.

We present in Fig. 1 an example of output from Weka, consisting in a decision tree for the SPX dataset with Tobin's Q dependent variable.

6 Conclusions

This paper develops a framework that applies four different machine learning algorithms (Adaboost, ADTree, J48 and Simple Logistic) to three data sets obtained from the component companies of the following stock market indexes: S&P500, STOXX Europe 600 and STOXX Eastern Europe 300. Our findings present several important features which have practical applicability both for the boards of the listed companies, as well as for investors. They can observe which corporate governance variables are closely connected to the performance indicator (Tobin's Q). The board members can learn what changes could be made in the corporate governance in order to align it with the best practices from the most successful companies. The investors can assess the changes in the corporate governance of a certain company and decide if it is heading the good direction. The relation between the corporate governance variables and the distress indicator Altman Z-score can serve to those interested in evaluating the bankruptcy risk of a company. While the indicators are computed using data from the past, a change in the corporate governance could be a flag in which direction the company is heading. By having the corporate governance variables categorized, one could take immediate action. In this respect, we obtained interesting relations between the corporate governance variables and the companies' performances, while observing different particularities of the three different world areas.

The performance of the algorithms is more solid for the S&P500 and STOXX Europe 600 indexes components because of the high availability of the data. For the Eastern Europe index the data related to corporate governance register missing values for many companies, mainly because they did not adopt all the good practices already established with the American and West European companies. The Adaboost M1 algorithm with ADTree and the Simple Logistic Regression obtained the best results, which is consistent with other research in this area [14].

The research could be improved by collecting more consistent data related to Eastern Europe companies, which could be achieved by questioning the targeted companies. Another direction for improvement is to collect historical data for the companies for several years in order to increase the datasets and to observe the variables evolution. In terms of learning algorithms, deep learning methods can be added to detect useful patterns.

Acknowledgments. This work was co financed from the European Social Fund through Sectoral Operational Program Human Resources Development 2007–2013, project number POSDRU/159/1.5/S/134197 and POSDRU/159/1.5/S/142115, Performance and excellence in doctoral and postdoctoral research in Romanian economics science domain" and from UEFISCDI under project JustASR - PN-II-PT-PCCA-2013-4-1644.

Appendix A

Variable	Description
Tax	Tax burden for the last 12 months
Interest	Interest burden for the last 12 months
Asset	Amount of sales or revenues generated per dollar of assets. The ratio is an indicator of the efficiency with which a company is deploying its assets
Fincl l	Financial leverage. Measures the average assets to average equity
Oper ROE	Normalized ROE. Returns on Common Equity based on net income excluding one-time charges
Dvd P/O	Dividend Payout Ratio. Fraction of net income a firm pays to its shareholders in dividends, in percentage
Board Size	Number of Directors on the company's board
% Non Exec Dir on Bd	Percentage of the board of directors that is comprised of non-executive directors
% Indep Directors	Independent directors as a percentage of total board membership
CEO Duality	Indicates whether the company's Chief Executive Officer is also Chairman of the Board, as reported by the company
Indep Chrprsn	Indicates whether the company chairperson was independent as of the fiscal year end
Indep Lead Dir	Indicates whether the company has an independent lead director within the board of directors
Frmr CEO or its Equiv on Bd	Indicates whether a former company chief executive officer (CEO) or person with equivalent role has been a director on the board
% Women on Board	Percentage of Women on the Board of Directors
Ind Dir Bd Mtg Att	Percentage of board meetings attended by independent directors
Unit or 2 Tier Bd Sys	Indicates whether the company's board has a Unitary (1) or Two Tier (2) system. Marked 2 when board system has separate boards for Supervisory/Commissioner board and Management board
Prsdg Dir	Indicates whether the company has a presiding director in its board of directors
% Feml Execs	Number of female executives, as a percentage of total executives.
Feml CEO or Equiv	Indicates whether the company Chief Executive Officer (CEO) or equivalent is female
Age Young Dir	Age of the youngest director on the company board in years
BOD Age Rng	Age range of the members of the company board in years, calculated by subtracting the age of the youngest director on the company board from the age of the oldest director on the company board
Age Old Dir	Age of the oldest director on the company board in years
Bd Avg Age	Average age of the members of the board
Board Duration	Length of a board member's term, in years
Board Mtgs #	Total number of corporate board meetings held in the past year
Exec Dir Bd Dur	Length of an executive director board member's term, in years

References

1. Acharya, V.V., John, K., Sundaram, R.K.: On the optimality of resetting executive stock options. J. Financ. Econ. **57**(1), 65–101 (2000)
2. Altman, E.I.: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Financ. **23**(4), 589–609 (1968)
3. Altman, E.I., Saunders, A.: Credit risk measurement: developments over the last 20 years. J. Banking Financ. **21**(11), 1721–1742 (1997)
4. Becker, B., Strömberg, P.: Fiduciary duties and equity-debtholder conflicts. Rev. Financ. Stud. **25**(6), 1931–1969 (2012)
5. Bhagat, S., Bolton, B.: Corporate governance and firm performance. J. Corp. Financ. **14**(3), 257–273 (2008)
6. Bolton, P., Chen, H., Wang, N.: A unified theory of tobin's q, corporate investment, financing, and risk management. J. Financ. **66**(5), 1545–1578 (2011)
7. Brenner, M., Sundaram, R.K., Yermack, D.: Altering the terms of executive stock options. J. Financ. Econ. **57**(1), 103–128 (2000)
8. Bryan, D., Fernando, G.D., Tripathy, A.: Bankruptcy risk, productivity and firm strategy. Rev. Account. Financ. **12**(4), 309–326 (2013)
9. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**(1), 321–357 (2002)
10. Chen, W., Xiang, G., Liu, Y., Wang, K.: Credit risk evaluation by hybrid data mining technique. Syst. Eng. Procedia **3**, 194–200 (2012)
11. Chung, K.H., Pruitt, S.W.: A simple approximation of Tobin's q. Financ. Manage. **23**, 70–74 (1994)
12. Claessens, S., Djankov, S.: Ownership concentration and corporate performance in the Czech Republic. J. Comp. Econ. **27**(3), 498–513 (1999)
13. Claessens, S., Yurtoglu, B.B.: Corporate governance in emerging markets: a survey. Emerg. Mark. Rev. **15**, 1–33 (2013)
14. Creamer, G., Freund, Y.: Learning a board balanced scorecard to improve corporate performance. Decis. Support Syst. **49**(4), 365–385 (2010)
15. Creamer, G.G., Freund, Y.: Predicting performance and quantifying corporate governance risk for Latin American adrs and banks. In: Financial Engineering and Applications. MIT, Cambridge (2004)
16. Daily, C.M., Dalton, D.R.: Bankruptcy and corporate governance: the impact of board composition and structure. Acad. Manage. J. **37**(6), 1603–1617 (1994)
17. Dybvig, P.H., Warachka, M.: Tobin's q does not measure performance: theory, empirics, and alternative measures. Unpublished Working paper, Washington University, Saint Louis, United Sates (2010)
18. Elloumi, F., Gueyie, J.P.: Financial distress and corporate governance: an empirical analysis. Corp. Governance Int. J. Bus. Soc. **1**(1), 15–23 (2001)
19. Fawcett, T.: An introduction to roc analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
20. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AL Mag. **17**(3), 37 (1996)
21. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: ICML. vol. 99, pp. 124–133 (1999)
22. Freund, Y., Schapire, R.E., et al.: Experiments with a new boosting algorithm. In: ICML, vol. 96, pp. 148–156 (1996)
23. Gibson, M.S.: Is corporate governance ineffective in emerging markets? J. Financ. Quant. Anal. **38**(01), 231–250 (2003)

24. Guest, P.M.: The impact of board size on firm performance: evidence from the UK. Eur. J. Financ. **15**(4), 385–404 (2009)
25. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD Explor. Newslett. **11**(1), 10–18 (2009)
26. Hall, M.A., Holmes, G.: Benchmarking attribute selection techniques for discrete class data mining. IEEE Trans. Knowl. Data Eng. **15**(6), 1437–1447 (2003)
27. Hambrick, D.C., D'Aveni, R.A.: Top team deterioration as part of the downward spiral of large corporate bankruptcies. Manage. Sci. **38**(10), 1445–1466 (1992)
28. Holmes, G., Pfahringer, B., Kirkby, R., Frank, E., Hall, M.: Multiclass alternating decision trees. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 161–172. Springer, Heidelberg (2002)
29. Kambal, E., Osman, I., Taha, M., Mohammed, N., Mohammed, S.: Credit scoring using data mining techniques with particular reference to sudanese banks. In: 2013 International Conference on Computing, Electrical and Electronics Engineering (ICCEEE), pp. 378–383. IEEE (2013)
30. Kirkos, E., Spathis, C., Manolopoulos, Y.: Data mining techniques for the detection of fraudulent financial statements. Expert Syst. Appl. **32**(4), 995–1003 (2007)
31. Kohavi, R., et al.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI 2014, pp. 1137–1145 (1995)
32. Landwehr, N., Hall, M., Frank, E.: Logistic Model Trees. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) ECML 2003. LNCS (LNAI), vol. 2837, pp. 241–252. Springer, Heidelberg (2003)
33. Lang, L.H., Stulz, R.M.: Tobin's q, corporate diversification and firm performance. Technical report, National Bureau of Economic Research (1993)
34. Li, H.X., Wang, Z.J., Deng, X.L.: Ownership, independent directors, agency costs and financial distress: evidence from chinese listed companies. Corp. Governance Int. J. Bus. Soc. **8**(5), 622–636 (2008)
35. Lu, Y.H., Tsai, C.F., Yen, D.C., et al.: Discovering important factors of intangible firm value by association rules. Int. J. Digital Account. Res. **10**, 55–85 (2010)
36. Lückerath-Rovers, M.: Women on boards and firm performance. J. Manage. Governance **17**(2), 491–509 (2013)
37. Merton, R.C.: On the pricing of corporate debt: the risk structure of interest rates*. J. Financ. **29**(2), 449–470 (1974)
38. Ohlson, J.A.: Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res. **18**(1), 109–131 (1980)
39. Parker, S., Peters, G.F., Turetsky, H.F.: Corporate governance and corporate failure: a survival analysis. Corp. Governance Int. J. Bus. Soc. **2**(2), 4–12 (2002)
40. Platt, H., Platt, M.: Corporate board attributes and bankruptcy. J. Bus. Res. **65**(8), 1139–1143 (2012)
41. Quinlan, J.R.: C4. 5: Programs for Machine Learning. Elsevier, Burlington (2014)
42. Ravisankar, P., Ravi, V., Rao, G.R., Bose, I.: Detection of financial statement fraud and feature selection using data mining techniques. Decis. Support Syst. **50**(2), 491–500 (2011)
43. Ryan, H.E., Wiggins, R.A.: Who is in whose pocket? Director compensation, board independence, and barriers to effective monitoring. J. Financ. Econ. **73**(3), 497–524 (2004)
44. Sumner, M., Frank, E., Hall, M.: Speeding up logistic model tree induction. In: Jorge, A.M., Torgo, L., Brazdil, P.B., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 675–683. Springer, Heidelberg (2005)

45. Tobin, J.: A general equilibrium approach to monetary theory. J. Money Credit Bank. **1**(1), 15–29 (1969)
46. Tsai, C.F., Lu, Y.H., Yen, D.C.: Determinants of intangible assets value: the data mining approach. Knowl.-Based Syst. **31**, 67–77 (2012)