Corporate Misconduct Prediction with Support Vector Machine in the Construction Industry

Ran Wang[1], Chia-Jung Lee[2], Shu-Chien Hsu[3*], Cheng-Yu Lee[4]

Abstract

Corporate misconduct may lead to severe economic loss and even fatal injuries to workers and residents in the construction industry. Previous studies have proven that board composition in organizations can be related to illegal business behaviors. By analyzing board composition data from 45 publicly listed construction companies in Taiwan, this paper provides a tool for predicting corporate misconduct. A Support Vector Machine (SVM) was used to construct such a prediction model, and a logistic regression model was employed as a benchmark to assess the performance of the established SVM model. The established SVM model achieved an accuracy rate of 72.22% for predicting the occurrence of corporate misconduct when applied to all observations in the sample, and a rate of 90% accuracy in predicting misconduct by companies found guilty of doing so in the sample, thus performing better than the logistic regression model. The developed model yields new insights on previous research and can guide stakeholders to reduce the risk of illegal business acts occurring in the construction industry.

Keywords: Corporate misconduct; Support vector machine; Board composition

[1] Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, 181 Chatham Road South, Hung Hom, Kowloon, Hong Kong SAR.
[2] Assistant Professor, Institute of Human Resource Management, National Sun Yat-Sen University, Kaohsiung, 70 Lienhai Rd., Kaohsiung 80424, Taiwan
[3] * Assistant Professor, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University. (Corresponding Author) Address: ZS944, 9/F, Block Z, 181 Chatham Road South, Hung Hom, Kowloon, Hong Kong SAR. E-mail: mark.hsu@polyu.edu.hk.
4 Associate Professor, Department of Industrial Management and Information, Southern Taiwan University of Science and Technology, 1 Nan-Tai Street, Yongkang Dist., Tainan 71005, Taiwan.

Introduction

Hundreds of firms are fined because of corporate misconduct each year (Karpoff et al., 1999). These violations have been ingrained in the construction industry (Transparency International, 2005). Construction firms face a long list of ethical challenges (Ho, 2011) at any stage of a project. Any illegal behavior taken by construction companies can lead to severe economic loss and even fatal injuries to workers and residents (Transparency International, 2005). In 2016, a construction platform of Fengcheng Coal Power Plant toppled over in Jiangxi Province of China, which killed 74 people and injured two others. The chairman, chief engineer and project leaders of the construction company and relevant others were arrested on the suspicion of collusion bidding, dereliction of duty, bribery, and embezzlement (CNN, 2016). In the same year, 27 people were confirmed dead, and at least 80 were injured due to the collapse of a flyover in India. Several top executives of the construction company were arrested, including its deputy general manager (The Indian Express, 2016). In Nigeria, a church that was under construction and that had no building permit collapsed and killed 29 persons due to multiple structural faults (BBC News, 2016). Cases like these are an ongoing problem, and most of them are attributed to human factors (e.g., managerial misconduct), which may be avoided by effective monitoring within organizations.

There is a merging consensus that corporate misconduct often results from the actions or inactions, deliberate or inadvertent, by the top managers of the organization (e.g., Daboub et al., 1995; Harris and Bromiley, 2007; Donoher et al., 2007). Some agency theorists (Fama and Jensen, 1983; Mace, 1971; Vance, 1964) advocate that the managerial misconduct may be avoided by the effective monitoring of board of directors. Thus, corporate boards play an important role in deterring illegal business practices. Some research has demonstrated that the likelihood of illegal corporate behavior is associated with board composition (Beasley, 1996; Kassinis and Vafeas, 2002;

Chen et al., 2006; Lee et al., 2018). However, just identifying the determinants of corporate misconduct is still far from effectively preventing its occurrence. It is essential to be able to accurately predict the probability of bad events in advance, and then a valid countermeasure can be initiated. The purpose of this study is to provide such a prediction tool based on historical information on board composition and corporate misconduct. Among the various prediction methods, support vector machine (SVM) will be adopted due to its ability to solve nonlinear and high dimensional classification problems (Cortes and Vapnik, 1995). This characteristic makes SVM quite suited for the corporate misconduct prediction problem because it has been established that at least one variable of board composition has a U-shaped relation with illegal behaviors and there are nine variables considered determinants of corporate misconduct in the construction industry (Lee et al., 2018). To evaluate the prediction performance of SVM, a logistic regression model, which is commonly used for prediction, will be employed to serve as a benchmark. This study would extend the research about corporate misconduct in terms of the construction industry and facilitate shareholders to take some precautions to reduce exposure of corporate scandals, investors or lenders to reconsider a company and regulators to target violating companies.

The structure of this paper is as follows: the next section reviews the existing studies about corporate misconduct, the role of the corporate board in the construction industry, and their relationship in order to demonstrate the feasibility of predicting the probability of corporate misconduct based on information about board composition. Next, the methodology is presented, including the two prediction methods, SVM, and logistic regression, and the data collection and processing is described. The respective implementation and prediction results of the two methods before the comparison are then presented, followed by a discussion and concluding summary.

Literature Review

According to Vaughan (1999), corporate misconduct is the "acts of omission or commission by individuals or groups of individuals acting in their organizational roles who violate internal rules, laws, or administrative regulations on behalf of organization goals." Construction companies, influenced by the fierce competition and complex policy environment (Brooks, 1992), are likely to commit fraud (Le et al., 2014a, b), bid cutting (May et al., 2001), collusive tendering (Zarkada-Fraser and Skitmore, 2000), and other illegal activities. These recurring illegal behaviors have raised concern in the academic community and scholars have started to focus on the antecedents of corporate misconduct. From a macro perspective, some scholars have attributed these activities to ineffective supervision (Van De Bunt, 2010; Bowen et al., 2012) or lack of mature regulations and standards (Bologna and Del Nord, 2000; Tabish and Jha, 2011). From a micro perspective, many studies have suggested that individual unethical decision-making is responsible for project failure and corporate scandals in the construction industry (Zarkada-Fraser and Skitmore, 2000; Ho, 2011; Abdul-Rahman et al., 2010). Most of them could focus on the employees (Ho, 2010), and professionals (Fan et al., 2001; Fan and Fox, 2009). However, the misconduct by these low-level employees may be engaged under the formal authority of top managers (Kristof-Brown et al., 2005; Jordan et al., 2013; Schaubroeck et al., 2012). How to avoid the direct and indirect illegal behaviors by management seems more significant.

Corporate boards, as the representatives of stockholders, play a crucial role in strategic decision making (Judge and Zeithaml, 1992; Haynes and Hillman, 2010; Bailey and Peck, 2013) as well as monitoring the performance of management (Ferris et al., 2003; Hillman et al., 2008). This has led scholars to examine the impacts of board composition. Agency theorists believe that the board of directors has the responsibility to maximize the wealth of a company (Fama and

Jensen, 1983; Mace, 1971; Vance, 1964). Directors need to determine the development direction, approve company actions, and monitor operations (Business Roundtable, 2016). In the construction sector, the board's role is even more important because project managers control information and possess tacit knowledge (Rebeiz, 2001). Without the commitment of the board and a well-established governance structure, it is easier for managers to adopt or ratify improper conduct that reduces the risk of their loss or maximizes their returns at the expense of shareholder wealth. This perspective indicates the crucial role of corporate boards in preventing or at least potentially deterring misconduct.

Many corporate governance studies span multiple research areas including economy, strategic management, and civil engineering (Fama and Jensen, 1983; Vance, 1964; Herman, 1981; Miozzo and Dewick, 2002), and most of the works focus on boards' impact on firm performance. Rebeiz and Salameh (2006) showed that a company's financial returns are influenced by the independent directors and chairmanship of the board. Combining a questionnaire survey and interviews, Luo (2001) revealed that for Sino-foreign construction joint ventures, company performance is related to the combined control of the board, management, and ownership structure. Only a few studies have considered the relationship between the board of directors and corporate misconduct in the construction industry. From the perspective of directors' capacity, Lee et al. (2018) explored how directors' industrial and education background affect the occurrence of illegal business behaviors. Their empirical results indicate that multiple directorships, board members' experience, and education diversity have a significant impact on corporate misconduct. This finding suggests that a board with insufficient capacity may face problems in properly fulfilling their duties and confirms the close relationship between directors' capacity and corporate misconduct. Though those studies have attempted to reveal the relationship between the board and corporate misconduct and indeed laid the foundation for us to develop a tool to predict corporate misconduct by drawing upon board-related information on construction companies, it is still far from avoiding the occurrence of illegal practices.

A misconduct prediction model could facilitate firms to take precautions as well as regulators, such as securities regulatory commission and other oversight agencies, to identify fraudulent firms for potential violation investigation (Perols et al. 2017). Thus, some studies have been attempting to construct prediction models for corporate misconduct. Mainly based on financial ratios, Pai et al. (2011) integrated sequential forward selection for feature selection, support vector machine (SVM) to determine the likelihood of financial statement fraud, and a classification and regression tree (CART) to enable auditors to increase substantive testing. Fanning and Cogger (1998) used financial ratios and other qualitative variables and constructed an Artificial Neural Network (ANN) model to detect fraudulent financial statements by management. These studies are in terms of financial statements issues and only take advantage of financial ratios. But other important variables are ignored, such as characteristics of corporate governance especially the board composition that is an endogenously determined characteristic of the firm (Brown et al. 2011; Coles et al. 2012; Hermalin and Weisbach 2001) and may dominant the occurrence of corporate misconduct. More importantly, there is a lack of building a specific prediction model in terms the construction industry, which is a high-risk sector (Rebeiz and Salameh, 2006).

Research Design

To develop the tool for corporate misconduct prediction in the construction industry, this study employed a support vector machine (SVM), which is a popular machine learning method in multiple disciplines. To assess the prediction performance of SVM, a logistic regression model was used as a comparison reference.

Support Vector Machine (SVM)

SVM was first introduced in 1995 (Cortes and Vapnik, 1995) and soon became a popular method for analysis (Chen and Lin, 2010) due to its ability to solve nonlinear and high-dimensional recognition problems (Cortes and Vapnik, 1995). This method has been applied to construction engineering studies for pattern classification problems such as cost estimates (An et al., 2007), contractor default prediction (Tserng et al., 2011), and spatial dynamics of the housing market (Chen et al. 2017). This study also applied this method to predict the potential nonlinear effects of board composition on corporate misconduct for construction companies.

Let $x = \{x_i\}$ as a multi-dimension input vector, $y = \{y_i\}$ as the corresponding label of x, and the set $D = \{(x_i, y_i)|x_i \in R, y_i \in [-1,1]\}_{i=1}^{n}$ as the training data. Then the hyperplane H could be established as

$$\begin{matrix} W^T x_i + b \geq 1 \\ W^T x_i + b \leq -1 \end{matrix}, if \begin{cases} y_i = 1 \\ y_i = -1 \end{cases} \quad (1)$$

Where W^T is a normal vector perpendicular to the above hyperplane. When the margin is the maximum by the separating data, the optimal hyperplanes H can be demonstrated as the following equation:

$$W^T x + b = 0 \quad (2)$$

If the training data (x_i, y_i) satisfy the two conditions: ① linearly separable and ② $y_i - [W^T x_i + b] - 1 \geq 0$, the classification interval, namely the distance between the two hyperplanes described by equation (2), could be presented as $\frac{2}{||W||}$. When it is maximum, the optimal solution could be achieved, which is equivalent to the following formula:

$$\min \frac{\|W\|^2}{2}$$
$$s.t. \quad y_i(W^T x + b) \geq 1 \tag{3}$$
$$i = 1,2, \ldots, n$$

Equation (3) could be changed to the following to prevent over-fitting:

$$\min_{W,W^T,b,\xi} \frac{\|W\|^2}{2} + c \sum_{i=1}^{n} \xi_i$$
$$s.t. \quad y_i(W^T x + b) \geq 1 - \xi_i \tag{4}$$
$$\xi_i \geq 0, i = 1,2, \ldots, n$$

where ξ represents the slack and should be greater than 1 if any dataset is not classified correctly, or equal 0. Penalty factor c is a constant and could become very large when attempting to minimize the training errors. However, to prevent over-fitting, it is usually set to a small value.

By solving the corresponding Lagrange formula of Equation (4), the following final decision function can be obtained:

$$f(x) = sgn(\sum_{i=1}^{n} y_i \alpha_i K(x, x_i) + b) \tag{5}$$

where α_i is the Lagrange multiplier and when $\alpha_i \neq 0$, the corresponding training sample is support vector; $K(x, x_i)$ is the kernel function; b is the bias.

According to Equations (4) and (5), the selection of the kernel function for computation is as important as that of parameters. There are four kinds of kernel functions, such as Linear Kernel, Polynomial Kernel, Sigmoid Kernel, and Radial Basis Function (RBF) Kernel. The last one was chosen for this study because it has been shown to be more effective than others (Min et al., 2006; Sun et al., 2014). More importantly, it is capable of coping with the nonlinear relationship and high dimensional problem (Keerthi and Lin, 2003). The formula of this function is as follows:

$$\mathrm{K}(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{6}$$

where the parameter $\gamma > 0$.

Logistic Regression Model

Following Li et al. (2017), Ang and Goh (2013) and Tserng et al. (2011), a logistic regression model was employed as a comparison method to evaluate the prediction performance of SVM. Logistic regression models have been widely applied in many fields, including medicine (Ostir et al., 2000; Gu et al., 2016), computer science (Zhao et al., 2016), and atmospheric science (Melcón et al., 2017). It also has been used in some studies on predicting the performance of construction contractors (Tserng et al., 2011; Russell and Jaselskis, 1992). Its popularity stems from its ability to better interpret the rule of data for dichotomous outcomes in contrast to common continuous outcomes (Lee, 1986).

Like generalized linear regression models, logistic regression models are also able to determine the relationship between dependent variables and independent variables. The main differences between logistic regression models and other regression models are the value distribution of dependent variables and the model formulation (Harrell, 2015). In a logistic regression model, the independent variable, also known as the label y_i, can be only 0 or 1. And its model function expression is

$$C(Y|X) = Prob\{Y = 1|X\} = (1 + exp(-X\beta - a))^{-1} \tag{7}$$

where β and a are estimated parameters. The former is the change in log odds that $Y = 1$ per unit change in X. The latter is a constant. The parameter estimation method used in the present study was maximum likelihood estimation, in order to determine the parameter value when the following log-likelihood function is maximum:

$$L(x_1, \cdots, x_n|\beta, a) = \sum_{i=1}^{n} y_i log p_i + (1 - y_i) log(1 - p_i) \tag{8}$$

Where $p_i = Prob\{Y = 1|x_i\}$.

Sample and Data

The variables that have been proved to associate with corporate misconduct would be given priority to be used. Following Beasley (1996), Wang and Hsu (2013) and Lee et al. (2018), board-related variables and the corporate violation data were collected. Our sample includes 45 construction companies in Taiwan and relevant information on them spanning from 2003 to 2014 was collected from the Corporate Social Responsibility (CSR) database provided by the Taiwan Economic Journal (TEJ) database. All the data and materials in the TEJ database are drawn from companies' annual reports, press releases, and other public records. To reduce the impacts of some unknown factors, the input data for SVM and the logistic regression model was obtained by calculating the average of each item of the original data over three consecutive years. Then the data on the 45 companies from 2005 to 2014 were included in this paper, and 450 firm-year observations were yielded. Of these observations, 200 were considered identify corporate misconduct because the firms were found guilty in litigated cases in the focal year.

Measures

According to prior literature (Beasley, 1996; Rebeiz and Salameh, 2006; Wang and Hsu, 2013; Lee et al., 2018), this study operationalized the input variables as follow.

Corporate misconduct (CM) is the label y_i mentioned in Equation (1) -(8). It is operationalized by whether the construction company was found guilty in court during the period under study of an illegal action including collusion and professional negligence. Namely, when corporate misconduct is detected by competent authorities and the corporate is proved to be a violator, the corresponding label $y_i = 1$, otherwise $y_i = -1$ (e.g., Schnatterly, 2003; Mishina et al., 2010). The data is obtained according to the Corporate Social Responsibility (CSR) database provided by Taiwan Economic Journal (TEJ) database, of which the data source is the public announcements about punishments and violations issued by Taiwan Stock Exchange Corporation

(TWSE) and Taipei Exchange as well as press releases by Economic Daily News, Commercial Times and other public information.

This study considered nine aspects to measure board composition, which were used as input vectors. The nine vectors include two variables about directors' experience (directorships and industry experience), two variables about directors' knowledge (education diversity and average education level) and five other variables about board composition.

Directorships (DS) have been discussed in corporate governance literature for years. Some scholars argue that multiple directorships could be beneficial for companies because interlocking directorates facilitate information exchange, knowledge integration, and resources leverage (Johnson et al., 2013; Field et al., 2013). However, as the directorships increase, the directors may become too busy to devote themselves to their duties (Grossman and Hart, 1980). As for the threshold of the busy directors, the consensus of most scholars is that a director may be too busy to adequately carry out duties if the person sits on three or more boards (Fich and Shivdasani, 2006). Moreover, the Council of Institutional Investors also recommends that directors should serve on no more than three boards. Therefore, to reflect not only the interlocking activities but also the busy status of directors, this paper applies the average number of directorships to measure the directorships of directors (Ferris et al., 2003).

Industry experience (IE), regarded as "borrowed experience" by Huff (1982), refers to the personal work experience that is directly related to the construction industry. Many empirical studies have shown that different personal work experience may lead to different behaviors and beliefs (Bantel and Jackson, 1989; Huff, 1990). A director with industry experience is expected to have a better understanding of the industry context than others without such experience (Porac et al., 1989; Kroll et al., 2008). Different industry experience can help decision makers see things with a wider perspective (Eisenhardt and Bourgeois, 1988), which may lead to innovation and better decision quality (Jackson et al., 1995), though conflicts of the board members (Pelled, 1996). The present study measure industry experience using the Blau Index (Blau, 1977), also known as the Diversity Index. This index was proposed to operationalize the diversity of some nominal features, such as ethnicity (Chen et al., 2016) and culture (Richard et al., 2004). Its calculation expression is as follows:

$$Diversity\ Index = 1 - \sum {p_i}^2 \tag{9}$$

where p_i is the percentage of the i-th kind of directors for the whole board. For industry experience, the directors are classified into three categories. The first two categories, following Sundaramurthy and Lewis (2003), include the insiders who have the advantage of firm-specific knowledge and the directors who don't have a position in the company or on other boards in the construction industry. The last category of directors, in line with Barroso et al. (2011), refers to the one who holds a seat on another board in the construction industry and thus has knowledge related to the industry context.

Education background is considered to represent the professional knowledge and skills of a director (Hambrick and Mason, 1984). The directors with a higher education level are expected to have more knowledge (Chiang and He, 2010), and better capacity to process information (Dollinger, 1984). Therefore, education level could be an indicator to evaluate the monitoring effectiveness of the board (Daily and Dalton, 1994). Beyond the boards' education level, the degree of diversity across the board members' educational backgrounds is also important. Dissimilar educational backgrounds could be a source of different ideas and perspectives, which eventually improve board members' decision-making quality (Westphal and Zajac, 1995). To reflect these considerations, this study uses two variables, *education diversity index* (ED) and the *average level* *of education* (ALE). Similar to the industry experience diversity index, the ED is also calculated by the diversity index, and the categories of directors are based on their education degrees, including doctor's degree, master's degree, bachelor's degree, high school diploma, and others. Consistent with Graham et al. (2012) and Baran and Forst (2015), the average level of education was obtained using the following equation:

$$Education\ Level = \sum_{i=1}^{5}(N_i \times W_i)/\sum_{i=1}^{5} N_i \qquad (10)$$

where N_i is the number of directors that get the i-th education degree, and W_i is the weight of the corresponding education degree. For doctor's degree, W_5=5; for master's degree W_4=4; bachelor's degree W_3=3; high school diploma, W_2=2; and for others, W_1=1.

This paper also takes into consideration the board size, directors' gender, directors' tenure and board independence. Large *board size* (BS) is expected to have better monitoring capability (Baran and Forst, 2015), but a smaller board can, in certain cases, work more effectively than the larger board of directors. Some empirical studies indicate that cognitive and behavioral patterns are different between female directors and male ones (Huang and Kisgen, 2013). Thus, the directors' gender is operationalized as the *number of female directors* (NFD). Directors with different tenure may establish various ties and relationships, and this may influence the other board members (Hillman et al., 2008; Johnson et al., 2013). This paper measure tenure by *average tenure* (AT) and *tenure variation* (TV) following Kor and Sundaramurthy (2009). *Board independence* (BI), is operationalized by the number of outside directors. The dominant perspective on manager-board relationships suggests that structural board independence increases the board's overall power in its relationship with top managers (Ryan and Wiggins, 2004). To prevent managers from using their leadership position on the board to dictate the agenda of board meetings (Lorsch and Maclver, 1989), structurally independent boards can constrain managers' power and limit the concealment of negative outcomes to shareholders (Abrahamson and Park, 1994).

Results

Implementation and Results of SVM

There are two steps in this SVM experiment: first, establishing the prediction model based on the training data, and then testing the performance of the established model based on testing data. Thus it is necessary to prepare the training data and testing data before conducting the SVM experiment. To avoid the over-fitting that results from using a set of data as training data and testing data simultaneously (Hu et al., 2015), the collected data from 2005 to 2014 was divided unequally according to the year. One part is training data and consists of the data of the 45 companies between 2005 and 2012, and the rest is used for testing. Both the training and testing data were inputted into MatLab loaded with LibSVM and normalized to the range of [0, 1] except the label y_i (the data about corporate misconduct), to obtain a stable model.

To construct the SVM prediction model, two parameters, penalty factor c and RBF kernel parameter γ (kernel function parameter g in LibSVM) needed to be determined. The method this study took advantage of is grid search and 10-fold cross-validation based on training data, following Sun et al. (2014). The former refers to choosing the value of the two parameters in a grid range at a certain value step. For example, the grid range of c and g is $\{2^{-10}, 2^{-9}, \cdots, 2^{9}, 2^{10}\}$ if the step of parameters value is 1. Those pairs of parameters value could lead to 441 prediction models. To optimize the parameters and obtain the best model, 10-fold cross-validation was employed. The training data was divided randomly into 10 subsets, nine of which were used to construct a preliminary model. Then the accuracy performance of the model was evaluated based on the remaining subset when the two parameters are assigned values to the grid range. This process was repeated 10 times by changing the remaining subset and, 10 prediction accuracies were generated. The average of the 10 cross-validation accuracies is calculated as the final accuracy. By changing the pairs of parameters value, the above procedure was repeated hundreds of times and the highest validation accuracy was obtained. In this study, after grid search and 10-fold cross-validation, the best cross-validation accuracy was found to be 79.7222%, with the corresponding values of the penalty factor c and the RBF kernel parameter g being 2.8284 and 26.9087 respectively, (see Fig. 1). Even though there may be more than one pair of c and g resulting in the best cross validation accuracy, the minimum (c, g), in particular the minimum c, was selected (Zhou et al., 2016).

Using the parameters corresponding to the highest validation accuracy, the final SVM prediction model was established. To assess its performance, this paper inputted the testing data except for the original labels into the model, and then some predicted labels were generated. By comparing the predicted labels with the original labels, this study was able to obtain the classification accuracies of the established SVM model, in particular, the accuracy of the company-year observations on corporate misconduct.

Based on the overall data used on corporate misconduct and board composition, the prediction results of SVM model are shown in Table 1. To present the performance of this model, three kinds of performance metrics, overall testing accuracy, sensitivity, and specificity, would be used following Chaudhuri and De (2011). According to Su and Chen (2012), the overall testing accuracy is the percentage of correct classifications among the total classifications. In Table 1, the overall testing accuracy was 72.22%, that is, the constructed SVM model could accurately predict whether corporate misconduct would occur for 65 observations among the all 90 testing observations. The other two performance metrics, sensitivity, and specificity measures the quality of the established rules, that is, whether the rule that could distinguish the event (here refers to corporate misconduct) do or do not happen truly is right. Sensitivity gauges the percentage of events (those observations with corporate misconduct in this study) that are classified correctly among the observations do have the event truly, while specificity is the percentage of non-events (those observations without corporate misconduct in this study) that are classified correctly among the observations do not have the event truly. These two metrics are necessary for a skewed data set having the issue of class imbalance (the number of observations without corporate misconduct is over three times that of observations having corporate misconduct) and unequal misclassification costs (Su and Chen, 2012). In this study, sensitivity is 90%, and specificity is 67.14%, meaning that the observations of some illegal actions taken could be identified with greater accuracy than those which do not. These results are acceptable because the cost of misclassifying the observations with misconduct is substantially higher than that of mistaking the observations without misconduct in the real world, even though they are considered equally in this paper.

Implementation and Results of Logistic Regression

To evaluate the predictive performance of the established SVM model further, logistic regressions were performed to provide a benchmark. The data is the same as that of SVM, but the square of directorships is added because the multiple directorships have been shown to have a U-shaped impact on the likelihood of misconduct occurrence (Lee et al., 2018) and there is a linear systematic component in the log-odds (Hosmer et al., 2008). Using SPSS 22 as the implementation platform, the results of the logistic regression analysis were obtained, shown in Table 2 and Table 3.

According to the coefficients and their significance in Table 2, the coefficients of industry experience ($B= -1.302$, $p= 0.034<0.1$) and average level of education ($B= -0.852$, $p= 0.009<0.1$) are significantly negative, so enhancing the experience diversity and education level may reduce the odds of corporate misconduct. The impacts of directorships (p=0.917>0.1 for its quadratic term, p=0.728>0.1 for its linear term) and education diversity (p=0.960>0.1), however, are insignificant. Average tenure (B=-0.067, p=0.096<0.1), tenure variation (B= 0.010, p= 0.027<0.1) and the number of female directors (B=-0.138, p=0.082<0.1) also have significant effects on the probability of illegal company behaviors. Higher tenure, more similar tenure and more female directors may contribute to decreasing the occurrence of illegal actions. Therefore, the logistic regression model could be established as follows.

$$Predicted\ logit\ of\ (CM) = 3.150 - 0.067 \times AT + 0.010 \times TV - 0.138 \times NFD$$
$$-1.302 \times IE - 0.852 \times ALE \qquad (11)$$

The goodness-of-fit test was performed using the Hosmer & Lemeshow test to assess whether the constructed logistic regression model is fit to the practical results. $\chi^2(8) = 5.773$ and $p > 0.05$, according to Peng et al. (2002), showing the model fits the actual data well.

Table 3 presents the classification accuracy of the above logistic regression model. In line with SVM, the results could also be stated by two measures, sensitivity and specificity. Though this model is not sensitive to the corporate misconduct (42.0%), it exhibits great specificity (78.8%), which means it is more appropriate for finding the law-abiding observations.

Comparison of the Results of SVM and Logistic Regression

Based on the above results, both SVM and logistic regression have their advantages and disadvantages. When considering the classification accuracies, SVM can effectively identify the companies who violate the law with higher accuracy. However, while the accuracy performance of logistic regression is marginally verified, it provides more information about each variable. Logistic regression can help us to specifically identify which variables have significant effects on corporate misconduct. In short, SVM may be a more appropriate tool to predict the effects of board composition on corporate misconduct because of the higher accuracy for distinguishing misconduct observations from other observations.

Discussion

To clarify the significance of directors' capacity, indicated by experience and knowledge, in predicting the corporate misconduct, this study conducted a comparative analysis by repeating the implementation of SVM by including (Model 1) or excluding (Model 2) the four variables about industry experience and educational background of the board. The corresponding results are presented in Table 4.

Compared to Model 1, each of the accuracies in Model 2 is about 10% lower than the corresponding accuracy in Model 1. Thus it can be inferred that the information about industry experience and educational background is quite helpful in predicting corporate misconduct. Moreover, Model 2 could detect a majority of misconduct observations with relatively high precision. It seems that to some extent the SVM model is more sensitive in identifying observations of illegal actions than identifying observations affirming legal compliance.

In Lee et al. (2018), only the average level of education has an insignificant impact on corporate misconduct among the four variables about directors' experience and knowledge, while only board size has a significant effect among the five control variables about the board. To explore whether the insignificant variables are ineffectual in predicting corporate misconduct, a series of comparative analyses were also implemented, and six models (Model 3-8) were constructed and compared to Model 1, with the results shown in Table 5. For Model 3-7, one insignificant variable in Lee et al. (2018) was deleted in each process of SVM operation. For Model 8, only the four significant variables are included, namely, multiple directorships, experience diversity, education diversity, and board size.

Comparing Model 3, Model 4, Model 6 and Model 7 with Model 1, if the average education level, board independence, tenure variation and an average of tenure respectively are neglected, a majority of observations could still be sorted properly. Nevertheless, the accuracy of distinguishing those with illegal practices from others decreases by 10%-20%, which means these above insignificant variables are still important to predicting corporate misconduct. For Model 5, when leaving out the gender distribution of the board, the accuracy for predicting the corporate misconduct remains 90% when the corporate misconduct occurred in a company, but the rest of two testing accuracies decreased by about 5%. From the comparison of Model 1 and Model 8, when only considering the significant variables, the value of three testing accuracies decrease by about 15%. In particular, the accuracy of identifying observations of illegal behaviors fall 20%. The constructed model could group only about half of the observations correctly. According to the above analyses, it seems that apparently insignificant variables still should not be overlooked in predicting illegal actions committed by companies.

Conclusion

Construction firms experience many ethical challenges. The risk of misconduct may in part be explained by insufficient capacity of the boards, which may cause ineffective monitoring and lead to poor strategic decisions. To facilitate firms themselves to take some precautions and regulators to target the fraudulent firms, this study based on the findings of Beasley (1996), Kassinis and Vafeas (2002), Wang and Hsu (2013) and Lee et al. (2018), and tried to predict the occurrence of corporate misconduct using board information as inputs. Drawing upon data for 45 publicly listed construction companies in Taiwan from 2005 to 2014, SVM and logistic regression models were implemented and compared. It is found that the constructed SVM model performs better in predicting illegal actions. In terms of the construction industry, this established SVM may be beneficial for shareholders to take precautions (e.g., changing the board composition) to avoid the managerial misconduct and protect their wealth as well as for investors and lenders (e.g., banks) to estimate the status of firms and make decisions. Moreover, two other noteworthy insights were obtained.

First, information on directors' experience and knowledge is indispensable when predicting the occurrence of corporate misconduct. The results of our study imply that directors' capability has an unneglectable impact on corporate misconduct, which is in line with Beasley (1996), Wang and Hsu (2013), and Lee et al. (2018). Although the role of knowledge and experience of board members has come to the forefront in recent discussions, this study provides a deeper understanding of how industry/education background may influence a director's ability in monitoring and governing companies. This SVM model can be useful to corporations, investors, and other stakeholders as the model's aid in revealing a relationship between board characteristics and the likelihood that a corporation will face legal consequences in violation of the law. This paper indeed revisited and confirmed the importance of board composition in previous literature (Beasley, 1996; Rebeiz and Salameh, 2006; Hillman et al., 2008; Barroso et al., 2011).

Second, the variables that Lee et al. (2018) found to be insignificant also could not be ignored when predicting the illegal behaviors of construction companies. Previous research on corporate misconduct has focused on how a board of directors prevents or deters illegal actions, addressing whether specific variables such board size, tenure, female or independent directors have a direct effect on firms' illegal behavior. This study, however, approached the issue with unconventional machine learning algorithms, asking if statistically insignificant variables are truly not important in predicting corporate misconduct. Taken together, the results of these SVM models lead us to question whether conceptual arguments based on traditional statistical analysis are legitimate. A variable from an empirical analysis found to be statistically insignificant might still be important.

While this paper provides a practical tool (SVM) to foresee the chance of illegal actions taken by construction firms and confirm the impact of director characteristics upon company misconduct, there are still some research limitations. First, although the prediction accuracy in terms of the observations with corporate misconduct is quite high, 90%, the accuracy for all the observations is 72.22%, which suggests that there is still some room for improvement. The rate may increase if more information about board characteristics are considered, such as the relationship among board members and directors' social-emotional wellness. Thus, other aspects of board composition should be included in future research. Second, the data used is only from the construction industry in Taiwan. Although this allowed us to develop a deeper understanding of a specific context, employing more data about the construction industry from different regions is recommended for future studies to confirm the universality of this SVM model considering the different cultural or institutional environment may lead to different behaviors of boards and organizations (Judge and Zeithaml, 1992; Bell et al., 2014).

Acknowledgments

The authors are grateful to acknowledge financial support from the Ministry of Science and Technology of Taiwan (MOST 106-2410-H-110-079) and from The Hong Kong Polytechnic University (1-ZVCF).

References

Abdul-Rahman, H., Wang, C., and Yap, X. W. (2010). “How professional ethics impact construction quality: Perception and evidence in a fast developing economy.” *Scientific Research and Essays*, *5*(23), 3742-3749.

Abrahamson, E., and Park, C. (1994). “Concealment of negative organizational outcomes: An agency theory perspective.” *Acad. Manage. J.*, 37(5), 1302-1334.

An, S. H., Park, U. Y., Kang, K. I., Cho, M. Y., and Cho, H. H. (2007). “Application of support vector machines in assessing conceptual cost estimates.” *J. Comput. Civil Eng.*, *21*(4), 259-264.

Ang, R. P., and Goh, D. H. (2013). “Predicting juvenile offending: a comparison of data mining methods.” *Int. J. Offender Ther.*, *57*(2), 191-207.

Bailey, B. C., and Peck, S. I. (2013). “Boardroom Strategic Decision-Making Style: Understanding the Antecedents.” *Corp. Gov-Oxford*, 21(2), 131-146.

Bantel, K. A., and Jackson, S. E. (1989). “Top management and innovations in banking: Does the composition of the top team make a difference?.” *Strategic Manage. J.*, 10(S1), 107-124.

Baran, L., and Forst, A. (2015). “Disproportionate insider control and board of director characteristics.” *J. Corp. Financ.*, 35, 62-80.

Barroso, C., Villegas, M. M., and Pérez-Calero, L. (2011). “Board influence on a firm's internationalization.” *Corp. Gov-Oxford*, 19(4), 351-367.

BBC News. (2016). “Nigeria church which collapsed killing dozens 'had structural faults.'” <http://www.bbc.com/news/world-africa-38347117> (3 January 2018).

Beasley, M. S. (1996). “An empirical analysis of the relation between the board of director composition and financial statement fraud.” *Account. Rev.*, *71*(4), 443-465.

Bell, R. G., Filatotchev, I., and Aguilera, R. V. (2014). "Corporate governance and investors' perceptions of foreign IPO value: An institutional perspective." *Acad. Manage. J.*, 57(1), 301-320.

Blau, P. M. (1977). *Inequality and heterogeneity*, Free Press, New York.

Bologna, R., and Del Nord, R. (2000). "Effects of the law reforming public works contracts on the Italian building process." *Build. Res. Inf.*, *28*(2), 109–118.

Bowen, P. A., Edwards, P. J., and Cattell, K. (2012). "Corruption in the South African construction industry: A thematic analysis of verbatim comments from survey participants." *Constr. Manage. Econ.*, *30*(10), 885–901.

Brooks, G. W. (1992). "Corruption and Racketeering in the New York City Construction Industry: The Final Report of the New York State Organized Crime Task Force." *Industrial and Labor Relations Review*, *45*(3), 606-607.

Brown, P., Beekes, W., and Verhoeven, P. (2011). "Corporate governance, accounting and finance: A review." *Account. Financ.*, 51(1), 96–172.

Business Roundtable. "Principles of Corporate Governance 2016." <https://businessroundtable.org/sites/default/files/Principles-of-Corporate-Governance-2016.pdf> (3 January 2018).

Chaudhuri, A., and De, K. (2011). "Fuzzy support vector machine for bankruptcy prediction." *Appl. Soft. Comput.*, 11(2), 2472-2486.

Chen, G., Crossland, C., and Huang, S. (2016). "Female board representation and corporate acquisition intensity." *Strategic Manage. J.*, *37*(2), 303–313.

Chen, G., Firth, M., Gao, D. N., and Rui, O. M. (2006). "Ownership structure, corporate governance, and fraud: Evidence from China." *J. Corp. Financ.*, *12*(3), 424-448.

Chen, J. H. and Lin, J. Z. (2010). “Developing an SVM based risk hedging prediction model for construction material suppliers.” *Automat. Constr.*, *19*(6), 702-708.

Chen, J. H., Ong, C. F., Zheng, L., and Hsu, S. C. (2017). “Forecasting spatial dynamics of the housing market using support vector machine.” *Int. J. Strateg. Prop. Manag.*, 21(3), 273-283.

Chiang, H. T., and He, L. J. (2010). “Board supervision capability and information transparency.” *Corp. Gov-Oxford*, *18*(1), 18-31.

Coles, J. L., Lemmon, M. L., and Meschke, J. F. (2012). “Structural models and endogeneity in corporate finance: The link between managerial ownership and corporate performance.” *J. Financ. Econ.*, 103(1), 149–168.

Cortes, C., and Vapnik, V. (1995). “Support-vector networks.” *Mach. Learn.*, *20*(3), 273-297.

CNN, C. L. (2016). Chinese power plant bosses arrested over deadly platform collapse. < http://www.cnn.com/2016/11/27/asia/china-construction-accident-arrests/index.html> (3 January 2018).

Daboub, A. J., Rasheed, A. M., Priem, R. L., and Gray, D. (1995). “Top management team characteristics and corporate illegal activity.” *Acad. Manage. Rev.*, 20(1), 138-170.

Daily, C. M., and Dalton, D. R. (1994). “Bankruptcy and corporate governance: The impact of board composition and structure.” *Acad. Manage. J.*, *37*(6), 1603-1617.

Dollinger, M. J. (1984). “Environmental boundary spanning and information processing effects on organizational performance.” *Acad. Manage. J.*, *27*(2), 351-368.

Donoher, W. J., Reed, R., & Storrud-Barnes, S. F. (2007). “Incentive alignment, control, and the issue of misleading financial disclosures.” *J. Manage.*, 33(4), 547-569.

Eisenhardt, K. M., and Bourgeois, L. J. (1988). “Politics of strategic decision making in high-velocity environments: Toward a midrange theory.” *Acad. Manage. J.*, *31*(4), 737-770.

Fama, E. F., and Jensen, M. C. (1983). “Separation of ownership and control.” *The Journal of Law and Economics*, *26*(2), 301-325.

Fan, L. C., and Fox, P. W. (2009). “Exploring factors for ethical decision making: Views from construction professionals.” *J. Prof. Iss. Eng. Ed. Pr.*, *135*(2), 60-69.

Fan, L., Ho, C., and Ng, V. (2001). “A study of quantity surveyors’ ethical behavior.” *Constr. Manage. Econ.*, 19(1), 19–36.

Fanning, K. M., & Cogger, K. O. (1998). “Neural network detection of management fraud using published financial data.” *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7(1), 21-41.

Ferris, S. P., Jagannathan, M., and Pritchard, A. C. (2003). “Too busy to mind the business? Monitoring by directors with multiple board appointments.” *J. Financ.*, *58*(3), 1087-1111.

Fich, E. M., and Shivdasani, A. (2006). “Are busy boards effective monitors?.” *J. Financ.*, *61*(2), 689-724.

Field, L., Lowry, M., and Mkrtchyan, A. (2013). “Are busy boards detrimental?.” *J. Financ. Econ.*, 109(1), 63-82.

Graham, J. R., Li, S., and Qiu, J. (2012). “Managerial attributes and executive compensation.” *Rev. Financ. Stud.*, 25(1), 144-186.

Grossman, S. J., and Hart, O. D. (1980). “Takeover bids, the free-rider problem, and the theory of the corporation.” *The Bell Journal of Economics*, *11*(1), 42-64.

Gu, Y., Zhao, K., Luan, X., Liu, Z., Cai, Y., Wang, Q., Zhu, B., and He, J. (2016). “Association between Serum Magnesium Levels and Depression in Stroke Patients.” *Aging Dis.*, *7*(6), 687-690.

Hambrick, D. C., and Mason, P. A. (1984). “Upper echelons: The organization as a reflection of its top managers." *Acad. Manage. Rev.*, *9*(2), 193-206.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, Springer, New York.

Harris, J., and Bromiley, P. (2007). "Incentives to cheat: The influence of executive compensation and firm performance on financial misrepresentation." *Organ Sci.*, 18(3), 350-367.

Haynes, K. T., and Hillman, A. (2010). "The effect of board capital and CEO power on strategic change." *Strategic Manage. J.*, 31(11), 1145-1163.

Hermalin, B. E., and Weisbach, M. S. (2001). *Boards of directors as an endogenously determined institution: A survey of the economic literature.* National Bureau of Economic Research.

Herman, E. S. (1981). *Corporate control, corporate power*, Cambridge University Press, Cambridge.

Hillman, A. J., Nicholson, G., and Shropshire, C. (2008). "Directors' multiple identities, identification, and board monitoring and resource provision." *Organ Sci.*, 19(3), 441-456.

Ho, C. M. F. (2010). "A critique of corporate ethics codes in Hong Kong construction." *Build. Res. Inf.*, *38*(4), 411-427.

Ho, C. M. F. (2011). "Ethics management for the construction industry: A review of ethical decision-making literature." *Engineering, Construction and Architectural Management*, 18(5), 516-537.

Hosmer Jr, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modelling of time to event data (2 ed.)*. John Wiley & Sons, New Jersey.

Hu, J., Qi, J., Peng, Y., and Ren, Q. (2015). "Predicting electrical evoked potential in optic nerve visual prostheses by using support vector regression and case-based prediction." *Inform. Sciences*, *290*, 7-21.

Huang, J., and Kisgen, D. J. (2013). "Gender and corporate finance: Are male executives overconfident relative to female executives?." *J. Financ. Econ.*, *108*(3), 822-839.

Huff, A. S. (1982). "Industry influences on strategy reformulation." *Strategic Manage. J.*, *3*(2), 119-131.

Huff, A. S. (1990). *Mapping strategic thought*, John Wiley & Sons, New Jersey.

Jackson, S. E., May, K. E., and Whitney, K. (1995). "Understanding the dynamics of diversity in decision-making teams." in R. A. Guzzo, E. Salas, and Associates. *Team Effectiveness and Decision Making in Organizations*, Jossey-Bass, San Fancisco.

Johnson, S. G., Schnatterly, K., and Hill, A. D. (2013). "Board composition beyond independence: Social capital, human capital, and demographics." *J. Manage.*, *39*(1), 232-262.

Jordan, J., Brown, M. E., Treviño, L. K., and Finkelstein, S. (2013). "Someone to look up to: Executive–follower ethical reasoning and perceptions of ethical leadership." *J. Manage.*, 39(3), 660-683.

Judge, W. Q., and Zeithaml, C. P. (1992). "Institutional and strategic choice perspectives on board involvement in the strategic decision process." *Acad. Manage. J.*, 35(4), 766-794.

Karpoff, J. M., Lott, J. R., and Rankine, G. (1999). "Environmental violations, legal penalties, and reputation costs." University of Chicago, Law School.

Kassinis, G., and Vafeas, N. (2002). "Corporate boards and outside stakeholders as determinants of environmental litigation." *Strategic Manage. J.*, *23*(5), 399-415.

Keerthi, S. S., and Lin, C. J. (2003). "Asymptotic behaviors of support vector machines with Gaussian kernel." *Neural Comput.*, *15*(7), 1667-1689.

Kor, Y. Y., and Sundaramurthy, C. (2009). "Experience-based human capital and social capital of outside directors." *J. Manage.*, *35*(4), 981-1006.

Kristof-Brown, A. L., Zimmerman, R. D., and Johnson, E. C. (2005). “Consequences of individuals'fit at work: a meta-analysis of person–job, person–organization, person–group, and person–supervisor fit.” *Pers. Psychol.*, 58(2), 281-342.

Kroll, M., Walters, B. A., and Wright, P. (2008). “Board vigilance, director experience, and corporate outcomes.” *Strategic Manage. J.*, *29*(4), 363-382.

Le, Y., Shan, M., Chan, A. P., and Hu, Y. (2014a). “Investigating the causal relationships between causes of and vulnerabilities to corruption in the Chinese public construction sector.” *J. Constr. Eng. M.*, *140*(9), 05014007.

Le, Y., Shan, M., Chan, A. P., and Hu, Y. (2014b). “Overview of corruption research in construction.” *J. Manage. Eng.*, *30*(4), 02514001.

Lee, C. J., Wang, R., Lee, C. Y., Hung, C. C. W., and Hsu, S. C. (2018). “Board Structure and Directors’ Role in Preventing Corporate Misconduct in the Construction Industry.” *J. Manage. Eng., 34*(2): 04017067.

Lee, J. (1986). “An insight on the use of multiple logistic regression analysis to estimate association between risk factor and disease occurrence.” *Int. J. Epidemiol.*, *15*(1), 22-29.

Li, Y., Luo, P., Fan, Z. P., Chen, K., and Liu, J. (2017). “A utility-based link prediction method in social networks.” *Eur. J. Oper. Res.*, *260*(2), 693-705.

Lorsch, J. W., and MacIver, E. (1989). *Pawns and potentates: The reality of America's corporate boards*. Harvard University, Boston.

Luo, J. (2001). “Assessing management and performance of Sino-foreign construction joint ventures.” *Construction Management and Economics*, *19*(1), 109-117.

Mace, M. L. (1971). *Directors: Myth and reality*. Harvard University Graduate School of Business Administration, Mass, Cambridge.

May, D., Wilson, O., and Skitmore, M. (2001). “Bid cutting: an empirical study of practice in South-East Queensland.” *Engineering, Construction and Architectural Management*, *8*(4), 250-256.

Melcón, P., Merino, A., Sánchez, J. L., López, L., and García-Ortega, E. (2017). “Spatial patterns of thermodynamic conditions of hailstorms in southwestern France.” *Atmos. Res.*, *189*, 111-126.

Min, S. H., Lee, J., and Han, I. (2006). “Hybrid genetic algorithms and support vector machines for bankruptcy prediction.” *Expert Syst. Appl.*, *31*(3), 652-660.

Miozzo, M., and Dewick, P. (2002). “Building competitive advantage: innovation and corporate governance in European construction.” *Res. Policy*, 31(6), 989-1008.

Mishina, Y., Dykes, B. J., Block, E. S., and Pollock, T. G. (2010). “Why “good” firms do bad things: The effects of high aspirations, high expectations, and prominence on the incidence of corporate illegality.” *Acad. Manage. J.*, 53(4), 701-722.

Ostir, G. V., Markides, K. S., Black, S. A., and Goodwin, J. S. (2000). “Emotional well-being predicts subsequent functional independence and survival.” *J. Am. Geriatr. Soc.*, *48*(5), 473-478.

Pai, P. F., Hsu, M. F., and Wang, M. C. (2011). “A support vector machine-based model for detecting top management fraud.” *Knowl.-Based Syst.*, 24(2), 314-321.

Pelled, L. H. (1996). “Demographic diversity, conflict, and work group outcomes: An intervening process theory.” *Organ. Sci.*, *7*(6), 615-631.

Peng, C. Y. J., Lee, K. L., and Ingersoll, G. M. (2002). “An introduction to logistic regression analysis and reporting.” *J. Educ. Res.*, *96*(1), 3-14.

Perols, J. L., Bowen, R. M., Zimmermann, C., and Samba, B. (2016). “Finding needles in a haystack: Using data analytics to improve fraud prediction." *Account. Rev.*, 92(2), 221-245.

Porac, J. F., Thomas, H., and Baden-Fuller, C. (1989). "Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers." *J. Manage. Stud.*, *26*(4), 397-416.

Rebeiz, K. S. (2001). "Corporate Governance and Role of CEO in Engineering Companies." *J. Manage. Eng.*, *17*(1), 14-23.

Rebeiz, K. S., and Salameh, Z. (2006). "Relationship between governance structure and financial performance in construction." *J. Manage. Eng.*, *22*(1), 20-26.

Richard, O., Barnett, T., Dwyer, S., and Chadwick, K. (2004). "Cultural diversity in management, firm performance, and the moderating role of entrepreneurial orientation dimensions." *Acad. Manage. J.*, *47*(2), 255–266.

Russell, J. S., and Jaselskis, E. J. (1992). "Predicting construction contractor failure prior to contract award." *J. Constr. Eng. M.*, *118*(4), 791-811.

Ryan, H. E., and Wiggins, R. A. (2004). "Who is in whose pocket? Director compensation, board independence, and barriers to effective monitoring." *J. Financ. Econ.*, 73(3), 497-524.

Schaubroeck, J. M., Hannah, S. T., Avolio, B. J., Kozlowski, S. W., Lord, R. G., Treviño, L. K., ... and Peng, A. C. (2012). "Embedding ethical leadership within and across organization levels." *Acad. Manage. J.*, 55(5), 1053-1078.

Schnatterly, K. (2003). "Increasing firm value through detection and prevention of white-collar crime." *Strategic Manage. J.*, 24(7), 587-614.

Su, C. T., and Chen, Y. C. (2012). "Rule extraction algorithm from support vector machines and its application to credit screening." *Soft Comput.*, 16(4), 645-658.

Sun, J., Shang, Z., and Li, H. (2014). "Imbalance-oriented SVM methods for financial distress prediction: a comparative study among the new SB-SVM-ensemble method and traditional methods." *J. Oper. Res. Soc.*, *65*(12), 1905-1919.

Sundaramurthy, C., and Lewis, M. (2003). "Control and collaboration: Paradoxes of governance." *Acad. Manage. Rev.*, 28(3), 397-415.

Tabish, S. Z. S., and Jha, K. N. (2011). "Analyses and evaluation of irregularities in public procurement in India." *Constr. Manage. Econ.*, *29*(3), 261–274.

The Indian Express. (2016). "Kolkata flyover collapse: IVRCL top official held, total nine." <http://indianexpress.com/article/cities/kolkata/kolkata-flyover-collapse-ivrcl-top-official-held-total-nine/> (3 January 2018).

Transparency International. (2005). "The Global Corruption Report 2005". Pluto Press, London. <https://www.transparency.org/whatwedo/publication/global_corruption_report_2005_corruption_in_construction_and_post_conflict> (3 January 2018).

Tserng, H. P., Lin, G. F., Tsai, L. K., and Chen, P. C. (2011). "An enforced support vector machine model for construction contractor default prediction." *Automat. Constr.*, *20*(8), 1242-1249.

Van De Bunt, H. (2010). "Walls of secrecy and silence: The Madoff Case and Cartels in the Construction Industry." *Criminol. Public Pol.*, *9*(3), 435-453.

Vance, S. C. (1964). *Boards of directors: Structure and performance*. School of Business Administration, University of Oregon, Eugene, Oregon State.

Vaughan, D. (1999). "The dark side of organizations: Mistake, misconduct, and disaster." *Annu. Rev. Sociol.*, 25(1), 271-305.

Wang, T., and Hsu, C. (2013). "Board composition and operational risk events of financial institutions." *J. Bank Financ.*, 37(6), 2042-2051.

Westphal, J. D., and Zajac, E. J. (1995). "Who shall govern? CEO/board power, demographic similarity, and new director selection." *Admin. Sci. Quart.*, *40*(1), 60-83.

Zarkada-Fraser, A., and Skitmore, M. (2000). "Decisions with moral content: collusion." *Construction Management and Economics*, *18*(1), 101-111.

Zhao, Y., Yang, Y., Lu, H., Liu, J., Leung, H., Wu, Y., Zhou, Y., and Xu, B. (2016). "Understanding the value of considering client usage context in package cohesion for fault-proneness prediction." *Automat. Softw. Eng.*, *24*(2), 1-61.

Zhou, Z., Xiao, T., Chen, X., and Wang, C. (2016). "A carbon risk prediction model for Chinese heavy-polluting industrial enterprises based on support vector machine." *Chaos, Solitons & Fractals*, *89*, 304-315.

Table 1. Prediction results of the SVM model

Company's Label y_i	Testing Sample		Accuracy (%)
	Actual Number	Prediction Number	
1	20	18	90%
-1	70	47	67.14%
Total Number	90	65	72.22%

Note: “1” refers to the label of companies conducting illegal behaviors in the focal year; while “-1” refers to the label of companies without illegal behaviors in the focal year.

Table 2. Results of Logistic Regression Analysis

Variable	B	S.E.	Wald χ^2	df	p	Exp(B)
Constant	3.150	1.110	8.050	1	0.005	23.336
AT	-0.067	0.040	2.771	1	0.096	0.935
TV	0.010	0.004	4.865	1	0.027	1.010
NFD	-0.138	0.079	3.021	1	0.082	0.871
BI	0.073	0.114	0.410	1	0.522	1.076
BS	0.025	0.042	0.345	1	0.557	1.025
DS	-0.098	0.281	0.121	1	0.728	0.907
$(DS)^2$	0.004	0.040	0.011	1	0.917	1.004
IE	-1.302	0.615	4.475	1	0.034	0.272
ED	-0.027	0.532	0.003	1	0.960	0.974
ALE	-0.852	0.326	6.814	1	0.009	0.426
Goodness-of-fit test	χ^2		df		p	
Hosmer & Lemeshow	5.773		8		0.673	

Note: For the effect of variables, it is significant when $p<0.1$.

Table 3. Classification results of Logistic Regression Model

Observed	Predicted		Accuracy (%)
	1	-1	
1	116	84	42.0
-1	197	53	78.8
Overall Accuracy			62.4

Note: “1” refers to the label of companies conducting illegal behaviors in the focal year; while “-1” refers to the label of companies without illegal behaviors in the focal year.

Table 4. Results of Comparative Analysis when including or excluding directors' experience and education background

	Model 1	Model 2
Best CVaccuracy	79.72%	73.33%
Best c	2.8284	7.4643
Best g	26.9087	142.0249
Testing accuracy for the whole	72.22% (65/90)	63.33% (57/90)
Testing accuracy when $y_i = 1$	90% (18/20)	80% (16/20)
Testing accuracy when $y_i = -1$	67.14% (47/70)	58.57% (41/70)

Note: (1) Best CVaccuracy is the best cross-validation accuracy based on training data. It was calculated to select the two parameters, c and g, in SVM. (2) Model 1 was conducted based on all the data used in the current study, while Model 2 was conducted based on partial data, namely excluding the information about the industry experience and educational background of directors.

Table 5. Results of Comparative Analyses to explore the contribution of insignificant variables in Lee et al. (2017)

	Model 1	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Best CVaccuracy	79.72%	79.17%	80.83%	79.72%	79.72%	76.67%	70.83%
Best c	2.8284	90.5097	6.7272	1.6245	19.0273	16	38.0546
Best g	26.9087	9.5137	13.4543	36.7583	6.7272	11.3137	128
Testing accuracy for the whole	72.22%	67.78%	70%	66.67%	71.11%	66.67%	56.67%
Testing accuracy when $y_i = 1$	90%	75%	70%	90%	80%	80%	70%
Testing accuracy when $y_i = -1$	67.14%	65.71%	70%	60%	68.57%	62.86%	52.86%

Note: Model 1 includes all of the variables; Model 3 includes all variables except Education Level; Model 4 includes all variables except Board Independence; Model 5 includes all variables except Number of Female Directors; Model 6 includes all variables except Tenure Variation; Model 7 includes all variables except Average of Tenure; Model 8 includes the significant variables only.