Predicting Disputes in Public-Private Partnership Projects: Classification and Ensemble Models

Jui-Sheng Chou[1] and Chieh Lin[2]

Abstract: Proactively forecasting disputes in the initiation phase of public-private partnership (PPP) projects can considerably reduce the effort, time, and cost of managing potential claims. This comprehensive study compared classification models for PPP project dispute problems. Performance comparisons included four machine learners, four classification and regression trees, two multivariate statistical techniques, and combinations of techniques that have performed best according to a historical database. Experimental results indicate that an ensemble technique (i.e., SVMs+ANNs+C5.0) provides better cross-fold prediction accuracy (84.33%) compared with all other individual classification models. Notably, SVM (support vector machine) is the best single model for classifying dispute propensity in terms of overall performance measures. This study demonstrates the efficiency and effectiveness of data-mining techniques for early prediction of dispute propensity in PPP projects pertaining to public infrastructure services. The modeling results provide proactive-warning and decision-support information needed for managing potential disputes before disputes occur. **DOI:** 10.1061/(ASCE)CP.1943-5487.0000197.

CE Database subject headings: Data collection; Private sector; Project management; Dispute resolution; Decision support systems.

Author keywords: Data mining; Classification; Public-private partnership; Project management; Dispute prediction.

Introduction

Public-private partnership (PPP) is a financial strategy for stimulating private investments in public services. Although the public sector is risk averse and tends to avoid financial guarantees, government support for PPP projects is common in developing countries compared with developed ones, especially after the 2008 global financial crisis significantly affected the construction industry, which is a pillar of the Taiwan economy. Since then, the Taiwan Public Construction Commission (TPCC) has actively promoted and encouraged private-sector participation in infrastructure and building construction throughout Taiwan.

Unlike the owner–general contractor relationship, PPP projects involve devoted stakeholders, including the promoter (the government), participating private investors, and financial institutions. Because of the high risks associated with the construction business, repeated challenges of stakeholders can cause delays in completing projects, budget overruns, and poor construction quality during the implementation, construction, operations, and transfer phases.

Although numerous studies (Abednego and Ogunlana 2006; Cheung 1999; Cheung et al. 2002; Gebken and Gibson 2006; Jones 2006) indicate that an efficient, effective, and fair dispute resolution process is essential for a successful PPP, this study focuses on providing a warning by predicting dispute propensity before project initiation. In addition to predicting propensity of construction claims, and unlike previous research in litigation outcome prediction (Arditi et al. 1998; Arditi and Pulket 2005, 2010; Arditi and Tokdemir 1999a, b; Chau 2007; Pulket and Arditi 2009a, b), the proposed prediction method gives supportive information needed by the governmental authority to furnish contract documents in the bidding phase.

On the basis of the partnerships, the Taiwan government functions as a promoter by constructing and operating public infrastructure or buildings with minimal out-of-pocket expense but with full administrative support. Private investors are expected to assume all risks and to obtain all funding from lenders (usually banks), who are expected to finance the private investors at a reasonable interest rate intermediated by the government. Given these seemingly bonded yet adversarial relationships between participants, the government, the investor, and the banker are likely to cooperate in successfully completing the project.

During the past decade, nevertheless, many PPP projects have failed because of project disputes occurring in the build–operate–transfer phases. The dispute rate was 23.6% during 2002–2009 according to statistics released by the Taiwan Public Construction Commission (PCC) (2011). Such disputes can be classified as mediation and nonmediation procedures. The nonmediation procedures include arbitration, litigation, negotiation, and resorting to higher authority, e.g., administrative appeals. In Taiwan, up to 84% of PPP projects are settled by mediation or negotiation within only 1–9 months, whereas arbitration or litigation costs all parties considerably more time and money (PCC 2011).

Taiwan has legally supported PPP projects for nearly 10 years. The National PPP Taskforce is usually responsible for nationwide policy making and sometimes advises on provisions for individual projects, whereas departments/local governments are generally responsible for PPP project delivery. For effective control of projects and to formulate proactive dispute management strategies, early knowledge and prediction of the PPP project dispute propensity is essential for providing the governmental PPP Taskforce with

[1]Professor in Project Management, Dept. of Construction Engineering, National Taiwan Univ. of Science and Technology, 43, Sec. 4, Keelung Rd., Taipei, 106, Taiwan (corresponding author). E-mail: jschou@mail.ntust.edu.tw

[2]Senior Specialist, Dept. of Planning, Public Construction Commission, Executive Yuan, Taiwan.

Note. This manuscript was submitted on May 13, 2011; approved on January 12, 2012; published online on January 14, 2012. Discussion period open until June 1, 2013; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, Vol. 27, No. 1, January 1, 2013. © ASCE, ISSN 0887-3801/2013/1-51-60/$25.00.

fundamental information for implementing a win-win PPP strategy and for preventing disputes.

Further, depending on the possible outcomes of a dispute, precautionary measures can be taken proactively when a project is in progress. Additional preparation in preventive actions can be beneficial once the disputes occur by reducing the future efforts, time, and costs of multiple parties during dispute settlement.

To achieve this goal (early knowledge of dispute propensity), this study presents classification models for predicting PPP dispute outcomes so as to alleviate future negative impacts on project delivery, operations, and transfer from the government perspective. The first step of the research procedure was to apply all historical data for PPP projects commenced during 2002–2009 to establish the functional relationships between characteristics of provided cases and the corresponding past claim occurrences. The major difference from conventional construction project disputes is that PPP project disputes may occur not only in the building phase, but also in the operations, rent, or transfer phases. The odds of a dispute are therefore much higher than those in a regular contracting procurement.

The remainder of this paper is organized as follows. The next section comprehensively reviews the literature on artificial intelligence and its application to predicting construction disputes and litigation resolution. The next two sections present the research methodology and evaluation methods, respectively, which provide a theoretical perspective of classification models adopted for subsequent investigation. Then, the next section articulates the project information and experimental analyses, including raw data preprocessing, descriptive statistics, and comparisons of model performance based on the PPP project database. Finally, concluding remarks and directions for future work are given.

Literature Background of Construction Dispute Prediction

In recent years, public infrastructure and building construction has often been financed by the PPP method because of the large and burdensome upfront investment costs (Clifton and Duffield 2006). Disputes among PPP participants may involve many issues, including surety bond issue, subcontractor qualification, license permit, investment scale, resident rights, government guarantee, excessive profits, operations period, taxation, and default loan commitment. Disagreements among parties can jeopardize the original project plan and cause the time-consuming dispute handling process that damage the government's PPP supporting image and the investors' future willingness to participate in public services.

When a dispute or claim occurs, the local government owners usually resort to adjudication by the central governmental authority (in this case, TPCC) if initial negotiations cannot resolve mutual conflicts. On agreement, an impartial committee may be suggested as a next option for mediating disputes (Jones 2006; Keith 1997). The timing to form up the committee, operation functions, and implementing methods should be clearly described in the contract before executing the project. Because mediation often cannot resolve the disputes, arbitration or litigation is often the only option based on the current law. However, because of the insufficient confidence of stakeholders about the current arbitration system, litigation is the agreed resolution mechanism in most PPP projects (PCC 2010). Because not all disputes or claims require costly and time-consuming dispute handling, an early-warning dispute prediction method is needed so that government owners and investors can take proactive measures during public construction projects.

Because of the unique and dynamic nature of PPP projects, claims and disputes are regulated by the executive government unit before the project construction commences (PCC 2010). Unlike conventional construction projects, PPP projects involve not only a building (B) stage, but also operations (O), rent (R), transferring (T), and own (On) stages (e.g., BOT, OT, ROT, BOOn, BTO). In post-construction stage, the likelihood of incomplete contracts usually increases, especially in long-term and complex PPP projects. Some contractual agreements designed to maximize ex-ante efficiency may result in ex-post inefficiency because the value of the contract performance to the promisee is lower than the cost of performance incurred by the promisor (Solino and Vassallo 2009).

Specifically, the opportunistic behaviors of individual parties trying to exploit a situation for their own advantage increase the potential for increased transaction costs in the postconstruction stage of PPP projects (Chang and Ive 2007; Irwin 2007; Zhang 2005). Moreover, as the highest public construction authority, TPCC encounters an extremely high volume of planning and promoting PPP projects and regular construction activities. Hence, management personnel would benefit if the TPCC or PPP Task Force had a decision-support tool for estimating dispute propensity and for planning how disputes should be resolved before initiating the project.

Several related attempts have been explored to minimize construction litigation by predicting the likelihood of court decisions. In Arditi et al. (1998), a network was trained using Illinois appellate courts data, and a 67% prediction accuracy was obtained. The authors believed that, if the parties to a dispute know with some certainty how the case would be resolved in court, the number of disputes could be reduced greatly.

In another series of studies, artificial intelligence techniques achieved superior prediction accuracy with the same data set, 83.33% in the case-based reasoning study (Arditi and Tokdemir 1999b), 89.95% in the boosted decision trees (Arditi and Pulket 2005), and 91.15% in integrated prediction modelling (Arditi and Pulket 2010). These studies used artificial intelligence (AI) to enhance prediction of outcomes in conventional construction procurement litigation.

However, Chau (2007) found that, other than the previously noted case studies, AI techniques are rarely applied in the legal field (Chau 2007). The author thus applied particle swarm optimization-based AI techniques to predict construction litigation outcomes, a field in which new data mining techniques are rarely applied. The presented network achieved an 80% prediction accuracy rate, which is much higher than mere chance. Nevertheless, the author suggested the use of additional case factors related to cultural, psychological, social, environmental, and political characteristics in future works.

For construction disputes triggered by change orders, Chen (2008) proposed a K Nearest Neighbor (KNN) pattern classification to identify potential lawsuits on the basis of a nationwide study of U.S. court records. The authors showed that the KNN approach has an 84.38% classification accuracy (Chen 2008). Chen and Hsu (2007) further applied hybrid ANN (artificial neural network)-CBR (case-based reasoning) model with disputed change order dataset to obtain early-warning information. The classifier reached a similar prediction rate of 84.61% (Chen and Hsu 2007).

Despite the numerous studies of CBR and its variations for identifying similar dispute cases for use as references in dispute settlements, Cheng et al. (2009) further refined and improved the conventional case-based reasoning approach by combining fuzzy-set theory with a new similarity measurement that fuses Euclidean distance and cosine angle distance (Cheng et al. 2009). The proposed model successfully extracted the knowledge and experience of experts from 153 historical construction dispute cases manually collected from multiple sources.

Generally speaking, all the past studies analyzed in that study focused on either specific change order disputes or on conventional contracting project using a single accuracy performance measure. Characteristics and environments of construction projects under PPP strategy, however, are much different from the above general contractor–owner relationships and need another insightful study of AI or data mining (DM) techniques with rigorous model performance measures to assist governmental agencies in predicting and preventing disputes before they occur.

Because disputes always involve numerous complex and interconnected factors and are difficult to rationalize, use of DM techniques is now among the most effective methods of determining hidden relationships between the available or accessible attributes and dispute-handling methods (Arditi and Pulket 2005, 2010; Arditi and Tokdemir 1999a; El-Adaway and Kandil 2010; Kassab et al. 2010; Pulket and Arditi 2009b). Identifying these attributes and methods would give practitioners improved understanding of the complex nature of PPP project claims.

The DM or AI-based approaches are related to computer system designs that attempt to resolve problems intelligently by emulating human brain processes. As AI technology enhances the ability of computer programs to handle tasks for which humans are still superior (Haykin 1999), AI models are typically used to solve prediction or classification problems. Researchers in various scientific and engineering fields have recently combined different AI paradigms to enhance their efficacy. Numerous studies confirm that hybrid AI schemes have promising applications in various industries (Arditi and Pulket 2010; Chen 2007; Chou et al. 2010, 2011; Kim and Shin 2007; Lee 2009; Li et al. 2005; Min et al. 2006; Nandi et al. 2004; Wu et al. 2009; Wu 2010). However, because selecting the most appropriate combinations is difficult and time-consuming, further attempts are not worthwhile unless significant improvements are made.

This study thereby constructed machine learning [ANNs, SVMs, DL (decision list), TAN (tree-augmented naïve Bayesian)], classification and regression-based techniques [CART (classification and regression technique), QUEST (quick, unbiased and efficient statistical tress), C5.0, Exhaustive CHAID (Chi-squared automatic interaction detection)], multivariate statistical methods [DA (discriminant analysis), LR (logistic regression)], and ensemble models by combining the best of the above approaches to classify PPP project dispute propensity and to compare their prediction performance. Notably, prediction accuracy depends mainly on the amount and quality of information available at the time of estimation. Thus, providing accurate preliminary estimates is extremely challenging when using these data mining models, particular during the project initiation phases (Chou 2009).

Classification Models and Research Methodology

When the response variable is categorical rather than continuous, prediction problems become data classification problems. Classification (or supervised learning) techniques are based on learning by examples that map input vectors into one of several desired output classes. In supervised learning, a target categorical variable (hereafter, dispute propensity), is partitioned into binary classes (i.e., dispute and no dispute). Although classification techniques are widely used in various disciplines, their effectiveness and efficiency are rarely exploited in the construction industry, particularly in the PPP-related domain. The classification techniques proposed in this study are concisely described along with their evaluation methods in the following sections.

Machine Learners

Neural Net

A neural net [or artificial neural network (ANN)] consists of information-processing units similar to neurons in the human brain except that a neural network consists of artificial neurons (Haykin 1999). Neural networks learn by experience, generalize from previous experiences to new ones, and make decisions. A neural network is a group of neural and weighted nodes, each representing a brain neuron, and the connections among these nodes are analogous to the synapses connecting brain neurons.

Specifically, multilayer perceptron (MLP) neural networks are the standard neural network models. In an MLP network, the input layer contains a set of sensory input nodes, one or more hidden layers contain computation nodes, and an output layer contains computation nodes. The input nodes/neurons are the feature values of an instance, and the output nodes/neurons function as discriminators between the class of the instance and those of all other instances.

In the multilayer architecture, input vector x passes through the hidden layer of neurons in the network to the output layer. The weight connecting input element i to hidden neuron j is denoted by W_{ji}, and the weight connecting hidden neuron j to output neuron k is denoted by V_{kj}. The net input of a neuron is obtained by calculating the weighted sum of its inputs, and its output is determined by applying a sigmoid function. Therefore, for the jth hidden neuron

$$\mathrm{net}_j^h = \sum_{i=1}^{N} W_{ji} x_i \quad \text{and} \quad y_i = f(\mathrm{net}_j^h) \tag{1}$$

whereas for the kth output neuron

$$\mathrm{net}_k^o = \sum_{j=1}^{J+1} V_{kj} y_i \quad \text{and} \quad o_k = f(\mathrm{net}_k^o) \tag{2}$$

The sigmoid function $f(\mathrm{net})$ is the logistic function

$$f(\mathrm{net}) = \frac{1}{1 + e^{-\lambda \mathrm{net}}} \tag{3}$$

where λ controls the gradient of the function.

For a given input vector, the network produces an output o_k. Each response is then compared with the known desired response of each neuron d_k. The weights in the network are then modified continuously to correct or reduce errors until the total error from all training examples is maintained below a predefined tolerance level.

For the output layer weights V and the hidden layer weights W, the update rules are given by the 4th and 5th equations, respectively:

$$V_{kj}(t+1) = v_{kj}(t) + c\lambda(d_k - o_k)o_k(1 - o_k)y_j(t) \tag{4}$$

$$W_{ji}(t+1) = w_{ji}(t) + c\lambda^2 y_j(1 - y_j)x_i(t) \times \left[\sum_{k=1}^{K}(d_k - o_k)o_k(1 - o_k)v_{kj}\right] \tag{5}$$

Support Vector Machines

Support vector machines (SVMs), which were introduced by Vapnik (1998), perform binary classification, i.e., they separate a set of training vectors for two different classes $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ where $x_i \in R^d$ denotes vectors in a d-dimensional feature space and $y_i \in \{-1, +1\}$ is a class label. The SVM model is generated by mapping the input vectors onto a new higher dimensional feature space denoted as $\Phi{:}R^d \to H^f$ where $d < f$. An optimal separating hyperplane in the new feature space is then constructed by a kernel function $K(x_i, x_j)$, which is the product of input vectors x_i and x_j and where $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$.

Decision List

A decision list (DL) finds a group of individuals with a distinct behavior pattern by using an if–then rule that has two parts: an antecedent part and a consequent part. The antecedent is a Boolean expression of predictors, and the consequent is the predicted value of the target field when the antecedent is true (Klivans and Servedio 2006). The DL algorithm can be summarized as follows (IBM 2010):

1. Find candidate rules in the original dataset.
2. Append the best rules to the decision list.
3. Remove records covered by the decision list from the dataset.
4. Find new rules based on the reduced dataset.
5. Repeat the process until one or more of the stopping criteria are met.

Notably, the order of decision rules is significant. A case that is governed by a rule is ignored by subsequent rules.

Tree Augmented Naïve Bayesian Classifier

The tree-augmented naïve Bayesian (TAN) classifier improves on the naïve Bayes model by succinctly describing the joint probability distribution for a given set of random variables. It allows predictors that are dependent on other predictors in addition to the target variable. The TAN has demonstrated superior classification accuracy and simplicity in comparison with the general Bayesian network model (Friedman et al. 1997). The algorithm for the TAN classifier learns a tree structure over $\mathbf{X}$ by using mutual information conditioned on categorical target variable Y, where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ represents a categorical predictor vector. It then adds a link from the target node to each predictor node. The TAN learning procedure is performed in the following steps (Friedman et al. 1997; SPSS 2007):

1. Use training data D, $\mathbf{X}$ and Y as input.
2. Learn a tree-like network structure over $\mathbf{X}$ by using a maximum weighted spanning tree method.
3. Add Y as a parent of each X_i where $1 \leqq i \leqq n$.
4. Learn the parameters of the TAN network.

Classification and Regression-Based Techniques

Classification and Regression Tree and Quick, Unbiased and Efficient Statistical Tree

Classification and regression tree (CART) technique (Breiman et al. 1984) partitions the data into two subsets so that the records within each subset are more homogeneous than those in the previous subset. In this recursive process, each of the two subsets is then split again, and the process repeats until the homogeneity criterion is reached or until some other stopping criterion is satisfied (Witten and Frank 2005). A relatively new binary tree-growing technique, QUEST (Quick, Unbiased and Efficient Statistical Tree), is another alternative binary-split decision tree algorithm for data classification. The QUEST algorithm is similar to CART except that QUEST uses an unbiased variable selection technique by default and uses imputation instead of surrogate splits to deal with missing values. Therefore, QUEST can easily handle categorical predictor variables with many categories (Loh and Shih 1997).

C5.0

Another recent classification technique is C5.0, which was developed by J. Ross Quinlan (Quinlan 2007). This greedy algorithm obtains decision trees featuring boosting technology for improving accuracy in identifying samples. The top-down approach (divide and conquer) to decision tree induction starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is constructed (Tan et al. 2006). The main difference between CART and C5.0 is that the former performs only binary splits, which gives binary trees, whereas the latter performs splits that are as large as the number of categories, which gives a "bushlike" structure. Notably, an alternative solution to limiting tree growth is pruning the full-grown tree. The CART and CART-like procedures use validation data to prune deliberately overgrown trees by using training data, whereas C5.0 uses training data for both growing and pruning the tree (Shmueli et al. 2007).

Exhaustive Chi-Squared Automatic Interaction Detection

To avoid overfitting the full-grown tree to the training data, Chi-squared automatic interaction detection (CHAID) is the standard method for setting stop rules to prevent the tree from growing excessively and overfitting the training data. The algorithm also uses a recursive partitioning method that predates CART technique and is widely applied in diverse domains (Shmueli et al. 2007). It tests for independence by Chi-square test to assess whether splitting a node obtains a statistically significant improvement in purity. Particularly, the predictor with the strongest association (according to p-value) with the response variable at each node is used as a split node. If the tested predictor does not show a statistically significant improvement, no split is performed, and the algorithm stops.

This study, however, proposes the use of Exhaustive CHAID, a modification of CHAID developed to address the weakness of the CHAID technique (Biggs et al. 1991), to classify the target field. Specifically, CHAID may not always find the optimal split for a predictor variable because it stops merging categories as soon as it finds that all remaining categories significantly differ. Exhaustive CHAID remedies this by continuing to merge categories of the predictor variable until only two super categories remain. It then examines the series of merges for the predictor and finds the set of categories that gives the strongest association with the target variable and computes an adjusted p-value for that association. Thus, Exhaustive CHAID can find the best split for each predictor and then choose which predictor to split by comparing their adjusted p-values (SPSS 2007).

Multivariate Statistical Analyses

Discriminant Analysis

The goal of discriminant analysis (DA) is to obtain rules that describe the separation between groups of observations (Hubert and Van Driessen 2004). Moreover, it predicts values of a categorical-dependent variable on the basis of a linear combination of interval independent variables, $X_1, X_2, \ldots, X_m$. The classification functions can predict which group in each case most likely belongs by calculating the resulting classification score, CS_i, which is given by

$$\mathrm{CS}_i = b_{i1}X_1 + b_{i2}X_2 + b_{i3}X_3 + \cdots + b_{im}X_m + k_i \tag{6}$$

where the subscript i denotes the respective group; subscripts $1, 2, \ldots, m$ denote the m discriminating variables; k_i = constant for the ith group, b_{ij} = regression coefficients for the jth variable when computing the classification score for the ith group; and X_j = observed value for the respective case for the jth variable. The regression coefficients are selected by maximizing the distance between the means of the dependent variable or, alternatively, by minimizing the distance between actual and predicted outputs (Katos 2007). For a two-group problem, the interpretation closely follows the logic of ordinary multiple regression.

Logistic Regression

Logistic regression (LR), also known as logistic model or logit model, extends linear regression to problems in which the dependent variable is categorical. It predicts the probable occurrence of an event by fitting data onto a logistic curve. Like regression analysis, its predictor variables may be either numerical or categorical. Given an input z, the logistic function [output $f(z)$] is

$$f(z) = \frac{1}{1 + e^{-\lambda z}} \tag{7}$$

where λ = parameter for controlling the gradient of the function.

The variables of z are usually defined as

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{8}$$

where β_0 is called the intercept; and β_1, β_2, and so on are called the regression coefficients of x_1, x_2, respectively. The intercept is the value of z when the value of all independent variables is zero. Positive and negative regression coefficients indicate that the explanatory variable increases and decreases, respectively, the probability of the outcome. Additionally, a large regression coefficient means that the risk factor has a stronger effect on the probability of an outcome.

Notably, both DA and LR determine which variables to retain or remove from related theories. Independent variables are selected for the regression equation through either forced entry or stepwise regression. The forced entry method does not filter independent variables and enters all inputs needed by the model to calculate the coefficients of all variables. Stepwise approach, however, repeatedly calculates the effect of each independent variable on the dependent variable; the variables are prioritized for selection or removal according to their contribution. Here, the process continues until all variables in the equation comply with the screening criteria.

Ensemble Model

Ranking a set of the preceding candidate models enables selection of the best performing models, which can be combined into a single ensemble model. This approach often yields more accurate predictions compared with conventional models because it aggregates the benefits of multiple models (Alexandre et al. 2001). Models combined in this manner typically perform at least as well as the best of the individual models and often better. If the primary goal is maximizing automation of the prediction process, the ensemble model usually provides practitioners with a sufficiently robust model without having to delve specifics as required in the preceding models.

The ensemble structure (Fig. 1) can specify the combined method used to determine the ensemble score, such as "simple voting," "confidence-weighted voting," and "highest confidence wins methods." In simple voting, for example, if two out of three models predict *dispute*, then *dispute* wins by a vote of two to one. In the case of confidence-weighted voting (CWV), the votes are weighted according to the confidence level for each prediction. Thus, if one model predicts *no dispute* with a higher confidence than the two *dispute* predictions combined, then *no dispute* wins. In this study, the combined method for dispute outputs is CWV method.

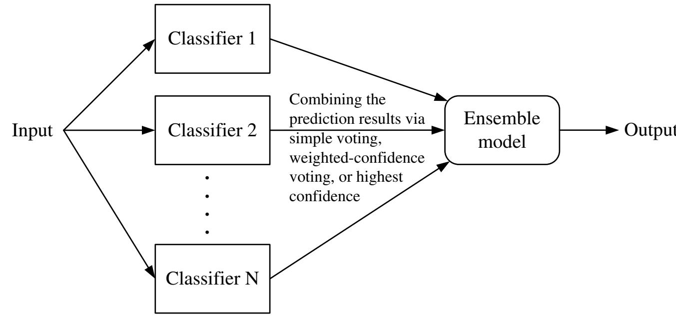

Fig. 1. Ensemble model

Model Validation and Evaluation Methods

Cross-Fold Validation

When comparing the predictive accuracy of two or more methods, researchers often use k-fold cross-validation to minimize bias associated with the random sampling of the training and holdout data samples. Since cross-validation requires random assignment of individual cases into distinct folds, a common practice is stratifying the folds themselves. In stratified k-fold cross-validation, the proportions of predictor labels (responses) in the folds are expected to approximate those in the original dataset.

Empirical studies show that, compared with regular k-fold cross-validation, stratified cross-validation tends to reduce bias in the comparison results (Han and Kamber 2001). Kohavi (1995) further confirmed that 10-fold validation testing was optimal in terms of computation time and variance (Kohavi 1995). Thus, a stratified 10-fold cross-validation approach was used to assess model performance in this study. The entire data set was divided into ten mutually exclusive subsets (or folds) with class distributions approximating those of the original data set (stratified). The subsets were extracted in five steps:

1. Randomize the dataset.
2. Extract one-tenth of the original data set size from the randomized data set (single fold).
3. Remove the extracted data from the original data set.
4. Repeat steps 1–3 eight times.
5. Assign the remaining portion of the data set to the last fold (10th fold).

After using this procedure to obtain 10 distinct folds, each fold was holdout rotationally for performance tests of the single flat and ensemble classification models, and the left over nine folds were used for training in turns, which obtained 10 independent performance estimates. The cross-validation estimate of overall accuracy was calculated by simply averaging the k individual accuracy measures for cross-validation accuracy.

Performance Evaluation Methods

The classification performance can be evaluated by computing the number of correctly recognized class examples (true positives; tp), the number of correctly recognized examples that do not belong to the class (true negatives; tn), and the number of examples that were either incorrectly assigned to the class (false positives; fp) or that were unrecognized as class examples (false negatives; fn) (Sokolova and Lapalme 2009). The four counts constitute a confusion matrix (Fig. 2), which can generate commonly used measures [e.g., accuracy, precision, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve] for binary

		Predicted	
		Positive	Negative
Actual	Positive	a (tp)	b (fn)
	Negative	c (fp)	d (tn)

Fig. 2. Confusion matrix

classification (Ferri et al. 2009; Horng 2010; Kim 2010; Sokolova and Lapalme 2009).

Prediction accuracy, which is defined as the percentage of records that is correctly predicted by the model relative to the total number of records among the classification models, is a primary evaluation criterion. The classification *accuracy* can be obtained by

$$\text{Accuracy} = \left(\frac{a+d}{a+b+c+d}\right) \tag{9}$$

Precision and *sensitivity* are two extended versions of accuracy. *Precision* can be considered a measure of exactness or fidelity, whereas sensitivity is a measure of completeness. The *precision* in Eq. (10) is defined as the proportion of the true positives compared with both the true positives and false positives given by the classifier. *Sensitivity* (a.k.a. *recall*), which is given by Eq. (11), is the number of correctly classified positive examples divided by the number of positive examples in the data. Sensitivity is useful for evaluating the effectiveness of a classifier in identifying positive labels.

$$\text{Precision} = \left(\frac{a}{a+c}\right) \tag{10}$$

$$\text{Sensitivity} = \left(\frac{a}{a+b}\right) \tag{11}$$

Another performance measure for binary classification is specificity, which is the proportion of correctly identified negatives in a binary classification. This measure shows how effectively a classifier identifies negative labels. Eq. (12) is the formula for specificity:

$$\text{Specificity} = \left(\frac{d}{c+d}\right) \tag{12}$$

Moreover, ROC curves indicate the ability of a classifier to avoid false classification. The ROC curve captures a single point, area under the curve (AUC), in the analysis of model performance. The AUC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (Fawcett 2006). The AUC, sometimes referred to as balanced accuracy (Sokolova and Lapalme 2009), is easily derived by Eq. (13):

$$\text{AUC} = \frac{1}{2}\left[\left(\frac{a}{a+b}\right) + \left(\frac{d}{c+d}\right)\right] \tag{13}$$

To compound the effect of preceding measures, an overall average performance score (S) for the distinct classification models is proposed:

$$S = \frac{1}{m}\sum_{i=1}^{m} P_i \tag{14}$$

where m = number of distinct performance measures; and $P_i = i$th performance measure. The S range is 0–1; the size of the coefficient is positively related to the effectiveness of the overall evaluation measures.

In this study, IBM SPSS Modeller (IBM 2010), a powerful and versatile data analytics workbench, was used for developing the preceding classification models and for evaluating PPP project dispute propensity prediction. All model parameters were set to the default to demonstrate the easier automation and implementation of the tool compared with other techniques.

Experimental Models for PPP Project Dispute Prediction

In 2000, Taiwan Legislative Yuan promulgated the *Act for Promotion of Private Participation in Infrastructure Projects*. Subject to the *Act*, privately owned organizations can invest in and operate 13 public infrastructure sectors, including environmental pollution prevention facilities; sewerage, water supply and water conservancy facilities; sanitation and medical facilities; social and labor welfare facilities; cultural and education facilities; major facilities for tour sites; power facilities and public gas and fuel supply facilities; sport facilities; parks facilities; major industrial, commercial and hi-tech facilities; development of new towns; and agricultural facilities.

To demonstrate the applicability and efficiency of the dispute classification schemes, this study used PPP project data collected by the TPCC, the highest authority of public services and infrastructure construction in Taiwan, in the proposed classification models for predicting dispute tendency. Modeling parameters are set to default while experimenting with various settings to produce baseline comparisons. Doing so ensures that building a prediction model is objective, easy, and satisfactory in terms of utilization and accuracy.

Data Collection and Preprocessing

The study database contains 584 PPP projects promoted by TPCC in 2002–2009. Of the 584 surveys issued, 569 were returned completed, and the response rate was 97.4%. The high return rate from the dispersed governmental units may have been attributable to the TPCC taskforce, the highest official authority, leading the survey activity. The questionnaire included items regarding the social demographics of respondents, background information, project characteristics, and project dispute resolutions. As necessary, survey data were preprocessed into a structural spreadsheet format by recording the project attributes with the dispute resolution methods.

In several projects, disputes occurred more than once (up to nine disputes in a single project) at various project phases. Thus, the overall data set included data for $N = 645$ cases (i.e., $N_2 = 493$ nondispute cases; $N_1 = 152$ dispute cases). Through experts' feedback, project attributes relevant to the prediction output of interest were initially identified through the questionnaire survey. However, quantitative techniques were still needed to build and validate the hidden relationships between the selected project predictors and the response (output) variable.

Unfortunately, the licensed operations duration (LOD) data were unavailable for one surveyed project that was abruptly canceled because of contract termination, although its remaining attributes were intact. Notably, since each project is valuable, extrapolating the missing data is both feasible and worthwhile. The LOD assumedly correlated with private capital investment (PCI). Thus, the missing LOD value was curve fitted through 11 estimation models (Table 1). One outlier and 33 zero capital data were eliminated during curve fitting. The best estimation was obtained by the Power model with R-squared. 719 and statistical significance at 1% alpha level. When using the Power model, the LOD derived for this project was 18.46 years.

Table 2 summarizes the statistical profile of categorical labels and numerical ranges for the resulting study samples. For PPP-orient procurement, specifically, 59.5% of projects were performed by the central government. Over the past eight years, most public construction has focused on cultural and education facilities (25.3%), sanitation and medical facilities (20.8%), transportation facilities (18.1%), and major tour-site facilities (10.5%). In accordance with the economic planning and development policy,

Table 1. Curve Estimating Model Summary for the Missing LOD Value

Model	Equation	R^2	F^a
Linear	$\mathrm{LOD} = 9.95 + 3.275E - 6(\mathrm{PCI})$	0.345	320.075
Logarithmic	$\mathrm{LOD} = (-27.034) + (3.888)\ln(\mathrm{PCI})$	0.633	1,047.697
Inverse	$\mathrm{LOD} = 13.775 - 3791.977/\mathrm{PCI}$	0.067	43.725
Quadratic	$\mathrm{LOD} = 7.945 + 9.978E - 6(\mathrm{PIC}) - 4.439E - 13(\mathrm{PCI})^2$	0.573	407.946
Cubic	$\mathrm{LOD} = -6.945 + 1.676E - 5(\mathrm{PCI}) - 1.955E - 12\,(\mathrm{PCI})^2 + 6.179E - 20(\mathrm{PCI})^3$	0.671	412.287
Compound	$\mathrm{LOD} = 6.842 \times 1.000^{\mathrm{PCI}}$	0.224	175.135
Power	$\mathrm{LOD} = 0.458 \times \mathrm{PCI}^{0.280}$	0.719	1,552.634
S	$\mathrm{LOD} = e^{(2.191-362.139/\mathrm{PCI})}$	0.134	94.102
Growth	$\mathrm{LOD} = e^{(1.923+1.782E-7(\mathrm{PCI}))}$	0.224	175.135
Exponential	$\mathrm{LOD} = 6.842 \times e^{1.782E-7(\mathrm{PCI})}$	0.224	175.135
Logistic	$\mathrm{LOD} = 1\{(1/u) + [0.146 \times (1.000^{\mathrm{PCI}})]\}$	0.224	175.135

[a]p-value < 0.001; LOD: licensed operations duration; PCI: private capital investment; u: upper boundary value.

Table 2. Project Attributes and Their Descriptive Statistics ($N = 645$)

Attribute	Data range, categorical label, or statistical description
Input variables	
Type of government agency in charge	Central authority (59.5%); municipality (11.5%); local government (29%)
Type of public construction and facility	1: transportation facilities (18.1%); 2: common conduit (0%); 3: environmental pollution prevention facilities (2.3%); 4: sewerage (1.1%); 5: water supply facilities (0.5%); 6: water conservancy facilities (2.5%); 7: sanitation and medical facilities (20.8%); 8: social welfare facilities (3.9%); 9: labor welfare facilities (1.2%); 10: cultural and education facilities (25.3%); 11: major tour-site facilities (10.5%); 12: power facilities (0%); 13: public gas and fuel supply facilities (0%); 14: sports facilities (3.3%); 15: parks facilities (2.5%); 16: major industrial facilities (0.5%); 17: major commercial facilities (1.9%); 18: major hi-tech facilities (0.2%); 19: new urban development (0%); 20: agricultural facilities (5.6%).
Project location	North (48.5%); center (21.2%); south (24.5%); east (5.3%); isolated island (0.5%)
Executive authority	Central authority (36.0%); municipality (36.1%); local government (27.9%)
Type of invested private sector	Standard industry classification-primary (0.2%); secondary (38.6%); tertiary (50.7%); quaternary (10.5%)
Planning and design unit	Government provides land and plans facility (91.0%); government provides land and private investor designs facility (5.9%); private provides land and designs facility (3.1%)
PPP contracting strategy	BOT (23.7%); OT (52.7%); ROT (23.6%)
Major public infrastructure/facility	Promoted as major public infrastructure/facility in PPP Act (80.1%); not major infrastructure/facility (19.9%)
Project scale	Range: 0–60,000,000; sum: 5.43E8; mean: 841337.1776; standard deviation: 3.52061E6 (thousand NTD; USD:NTD is about 1:30 as of April 2011)
Government capital investment	Range: 0–9,600,000; sum: 40,975,392.41; mean: 63527.7402; standard deviation: 5.11192E5 (thousand NTD)
Private capital investment	Range: 0–60,000,000; sum: 5.02E8; mean: 777809.4374; standard deviation: 3.32433E6 (thousand NTD)
Private capital investment ratio (PCIR)	Range: 0–100; mean: 91.4729; standard deviation: 25.42269 (%)
Licensed operations duration	Range: 0–60; mean: 11.9778; standard deviation: 13.39007 (year)
Output variables	
Dispute propensity	No dispute occurred (76.4%); dispute occurred (23.6%)

48.5% of projects were located in northern Taiwan. By the definition of standard industry, most private sector investment was in industrial departments (38.6%) and service departments (50.7%). In most cases (91.0%), the government provided land and facility designs to attract the investors.

Over the past practice, the three major PPP strategies for delivering public services were BOT (23.7%), OT (52.7%), and ROT (23.6%). Specifically, The World Bank Group (WBG) (2011) defines BOT (build, operate, and transfer) as a strategy in which a private sponsor builds a new facility, operates the facility, and then transfers the facility to the government at the end of the contract period. The government usually provides revenue guarantees through long-term take-or-pay contracts. Where a private sponsor rehabilitates an existing facility, then operates and maintains the facility at its own risk for the contract period, the PPP strategy is ROT (rehabilitate, operate, and transfer) according to the WBG classifications. Projects involving only management and lease contracts are classified as OT (operate and transfer) projects.

Further, flagship infrastructure projects refer to those that are important and fairly large in scale. The average project value is approximately 841 million NTD (new Taiwan dollar). On the basis of the collected data, the overall procurement amount through PPP was about 543 billion NTD. The mean capital investment by the government and private sectors per project was 63.5 million NTD and 777.8 million NTD, respectively. Notably, average private capital investment ratio was as high as 91.4%. The mean duration of licensed facility operations was about 12.0 years (maximum, 60 years).

To measure the dependencies among the categorized data, contingency table analyses were compared between the distinct predictors and response variable through Chi-square testing to infer the relationships (Table 3). All tests obtained statistically significant results at least or near 5% alpha level except the variable (i.e., planning and design unit), that was a rejection of the null hypothesis, i.e., no relationship was observed between the row variable (input variables) and the column variable (output variable).

For example, among the disputed cases ($N_1 = 152$), the central government had a higher probability of encountering disputes (67.1% probability) compared with municipal (15.1%) and local agencies (17.8%). Particularly, in Nos. 1, 6, 7, 10, 11, and 20 in type of public construction and facility (Table 2), disputes occurred in 76.4% of projects. The data showed that 85.5% disputes occurred in northern and southern Taiwan.

Interestingly, 92.1% of the disputes occurred when the government provided the land and was in charge of designing the facility, whereas merely 2% occurred when private investors provided the land and designed the facilities themselves. Among the PPP strategies, the probability of disputes was higher in BOT (49.3%) than in OT (32.2%) and ROT (18.4%). Notably, once the project was legally promoted as major infrastructure, the likelihood of a dispute involving PPP (38.8%) was lower than that in nonmajor infrastructure (61.2%).

Moreover, once the project value exceeded 50 million NTD, the dispute propensity was 4.33 times higher than that for projects valued between 5–50 million NTD or less than 5 million NTD. However, when the private sector investment exceeded 75%, the dispute tendency increased to 92.8%. Notably, dispute patterns were significantly related to licensed operation period. Table 3 summarizes the statistical results of the cross-analysis.

Table 3. Contingency Table and Chi-Square Test Results for Disputed Cases

Project attributes	p-value	Dispute occurred (%)
Agency	0.002	
Central authority		67.1
Municipality		15.1
Local government		17.8
Type of public construction	0.000	
Transportation facilities		10.5
Water conservancy facilities		9.9
Sanitation and medical facilities		17.1
Cultural and education facilities		13.2
Major tour-site facilities		14.5
Agricultural facilities		11.2
Planning and design unit	0.657	
Government provides land and plans facility		92.1
Government provides land and private investor designs facility		5.9
Private investor provides land and designs facility		2.0
PPP strategy	0.000	
BOT		49.3
OT		32.2
ROT		18.4
Major public infrastructure	0.000	
No		61.2
Yes		38.8
Project scale (thousand NTD)	0.000	
<5,000		15.8
5,000–50,000		15.8
>50,000		68.4
PCIR (%)	0.057	
<25		3.3
25–50		0.0
50–75		3.9
>75		92.8
LOD (year)	0.000	
<5		19.7
5–10		23.0
10–15		5.9
15–20		13.8
>20		37.5

Analytical Results and Model Performance

The analyses were reproduced by cross-fold method. In each fold experiment, 56 models were generated to train the models of each DM method. The testing fold was then used to verify and retain the best model of each DM method. The procedure was then rotated to the next fold until every fold was tested. The coincident matrices (rows show actuals; columns display predictions) for the individual models were used to calculate model performance in terms of five measures: accuracy [Eq. (9)], precision [Eq. (10)], sensitivity [Eq. (11)], specificity [Eq. (12)], and AUC [Eq. (13)]. Additionally, a composite index S [Eq. (14)] was derived to evaluate the overall performance of the distinct classification models.

Table 4 shows the cross–fold modeling performance. The C5.0, CHAID, and ANNs models were the most accurate according to accuracy value, which is the most single common measure of model performance. In terms of overall performance measure S (0.781), SVMs ranked highest in four performance measures followed by ANNs, C5.0, and CART Interestingly, C5.0 was the best single flat model for predicting dispute/no-dispute outcomes (accuracy = 83.25%) and for classifying no-dispute examples (sensitivity = 95.58%), whereas SVMs performed best at measuring classification fidelity for no-dispute examples (precision = 87.67%) and identifying dispute cases (specificity = 60.42%). Moreover, SVMs were also the best classifier in terms of avoiding false classification (AUC = 0.7391).

Table 4. Cross–Fold Modeling Performance

Category	Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC	S
Machine learners	ANNs	**82.18** **(3)**	84.56	93.80	45.26	0.6953	**0.751** **(2)**
	SVMs	80.93	**87.67** **(1)**	87.40	**60.42** **(1)**	**0.7391** **(1)**	**0.781** **(1)**
	DL	16.95	7.96	8.67	45.33	0.2700	0.212
	TAN	58.17	79.81	60.76	51.32	0.5256	0.605
Classification and regression techniques	CART	80.00	85.25	89.30	50.39	0.6985	**0.750** **(3)**
	Exhaustive CHAID	**82.63** **(2)**	84.47	94.42	44.01	0.6920	0.749
	QUEST	79.06	81.86	93.36	33.68	0.6371	0.703
	C5.0	**83.25** **(1)**	84.24	**95.58** **(1)**	42.62	0.6910	**0.750** **(3)**
Multivariate statistical methods	LR	79.69	82.46	93.20	35.41	0.6430	0.710
	DA	77.38	86.38	83.64	56.93	0.7028	0.749

Note: (1)–(3) denotes performance ranking.

Table 5. Cross-Fold Ensemble Model Performance

Ensemble model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC	S
SVMs+ANNs+C5.0	84.33	85.60	95.26	48.82	0.7229	0.773

The three best performing single models (SVMs, ANNs, and C5.0) were combined (ensemble model) using a confidence-weighted voting method to assess the aggregation power. Table 5 shows that average cross-fold accuracy (84.33%) was higher than that in any individual models (Table 4). However, the overall performance measure was not as good as the best of the three individual models in this case.

To sum up, an ensemble model (SVMs+ANNs+C5.0) performed best for cross-fold testing data for accuracy, whereas SVMs was the best overall model for classifying PPP project dispute propensity. Notably, DL was the worst classifier in almost all measurements and was clearly inappropriate for project dispute classification.

Conclusions

The objective of this study was not to propose alternative dispute resolution (ADR) methods for handling disputes but to predict dispute propensity in PPP projects by using classification models. For government agencies, the advantages of early warning of dispute propensity include reducing the time and effort needed to prepare a rule set for preventing potential disputes by improving understanding among the parties of the essential issues on each side of the potential disputes.

This study presents a number of different classifiers, uses the three best performed models, and combines them to predict dispute propensity. For accuracy, although the C5.0 (83.25%), CHAID (82.63%), and ANNs (82.18%) models perform well when using cross-fold testing data, the ensemble model (84.33%) performs even better. Interestingly, in terms of overall performance measurement score (S), SVMs (0.781), ensemble model (0.773), and ANNs (0.751) are the three best classification models.

Although this study comprehensively compared the effectiveness of data mining techniques in terms of PPP project dispute prediction, some classification techniques and their variations were not evaluated in this study. Future work may investigate whether a hierarchical ensemble approach combining multiple classifications and clustering techniques in a parallel or series form can improve model performance.

Integration of proactive strategy deployment and preliminary countermeasures are also worthy of further study in PPP project dispute early-warning systems. Another potential research direction is a second model for use once the propensity of a dispute is identified. On the basis of the disputed cases, such a model is needed to predict which dispute category and which resolution methods are likely to be used at which phases of the project lifecycle by mapping the hidden association rules.

Acknowledgments

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research. Gratitude is also extended to Taiwan Public Construction Commission whose generosity in data, knowledge, and experience sharing makes the study possible.

References

Abednego, M. P., and Ogunlana, S. O. (2006). "Good project governance for proper risk allocation in public-private partnerships in Indonesia." *Int. J. Proj. Manage.*, 24(7), 622–634.

Alexandre, L. A., Campilho, A. C., and Kamel, M. (2001). "On combining classifiers using sum and product rules." *Pattern Recognit. Lett.*, 22(12), 1283–1289.

Arditi, D., Oksay, F. E., and Tokdemir, O. B. (1998). "Predicting the outcome of construction litigation using neural networks." *Comput. Aided Civ. Infrastruct. Eng.*, 13(2), 75–81.

Arditi, D., and Pulket, T. (2005). "Predicting the outcome of construction litigation using boosted decision trees." *J. Comput. Civ. Eng.*, 19(4), 387–393.

Arditi, D., and Pulket, T. (2010). "Predicting the outcome of construction litigation using an integrated artificial intelligence model." *J. Comput. Civ. Eng.*, 24(1), 73–80.

Arditi, D., and Tokdemir, O. B. (1999a). "Comparison of case-based reasoning and artificial neural networks." *J. Comput. Civ. Eng.*, 13(3), 162–169.

Arditi, D., and Tokdemir, O. B. (1999b). "Using case-based reasoning to predict the outcome of construction litigation." *Comput. Aided Civ. Infrastruct. Eng.*, 14(Compendex), 385–393.

Biggs, D., de Ville, B., and Suen, E. (1991). "A method of choosing mutiway partitions for classification and decision trees." *J. Appl. Stat.*, 18(1), 49–62.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*, Chapman & Hall/CRC, New York.

Chang, C.-Y., and Ive, G. (2007). "The hold-up problem in the management of construction projects: A case study of the Channel Tunnel." *Int. J. Proj. Manage.*, 25(4), 394–404.

Chau, K. W. (2007). "Application of a PSO-based neural network in analysis of outcomes of construction claims." *Autom. Construct.*, 16(5), 642–646.

Chen, J.-H. (2008). "KNN based knowledge-sharing model for severe change order disputes in construction." *Autom. Construct.*, 17(6), 773–779.

Chen, J.-H., and Hsu, S. C. (2007). "Hybrid ANN-CBR model for disputed change orders in construction projects." *Autom. Construct.*, 17(1), 56–64.

Chen, K.-Y. (2007). "Forecasting systems reliability based on support vector regression with genetic algorithms." *Reliab. Eng. Syst. Saf.*, 92(4), 423–432.

Cheng, M.-Y., Tsai, H.-C., and Chiu, Y.-H. (2009). "Fuzzy case-based reasoning for coping with construction disputes." *Expert Syst. Appl.*, 36(2), 4106–4113.

Cheung, S.-O. (1999). "Critical factors affecting the use of alternative dispute resolution processes in construction." *Int. J. Proj. Manage.*, 17(3), 189–194.

Cheung, S.-O., Suen, H. C. H., and Lam, T.-I. (2002). "Fundamentals of alternative dispute resolution processes in construction." *J. Construct. Eng. Manage.*, 128(5), 409–417.

Chou, J.-S. (2009). "Generalized linear model-based expert system for estimating the cost of transportation projects." *Expert Syst. Appl.*, 36(3, Part 1), 4253–4267.

Chou, J.-S., Chiu, C.-K., Farfoura, M., and AL-Taharwa, I. (2011). "Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data mining techniques." *J. Comput. Civ. Eng.*, 25(3), 242–253.

Chou, J.-S., Tai, Y., and Chang, L.-J. (2010). "Predicting the development cost of TFT-LCD manufacturing equipment with artificial intelligence models." *Int. J. Prod. Econ.*, 128(1), 339–350.

Clifton, C., and Duffield, C. F. (2006). "Improved PFI/PPP service outcomes through the integration of Alliance principles." *Int. J. Proj. Manage.*, 24(7), 573–586.

El-Adaway, I. H., and Kandil, A. A. (2010). "Multiagent system for construction dispute resolution (MAS-COR)." *J. Construct. Eng. Manage.*, 136(3), 303–315.

Fawcett, T. (2006). "An introduction to ROC analysis." *Patt. Recognit. Lett.*, 27(8), 861–874.

Ferri, C., Hernández-Orallo, J., and Modroiu, R. (2009). "An experimental comparison of performance measures for classification." *Patt. Recognit. Lett.*, 30(1), 27–38.

Friedman, N., Geiger, D., Goldszmidt, M., Provan, G., Langley, P., and Smyth, P. (1997). "Bayesian network classifiers." *Mach. Learn.*, 29(2/3), 131–163.

Gebken, R. J., and Gibson, G. E. (2006). "Quantification of costs for dispute resolution procedures in the construction industry." *J. Prof. Issues Eng. Educ. Pract.*, 132(3), 264–271.

Han, J., and Kamber, M. (2001). *Data mining: Concepts and techniques*, Morgan Kaufmann Publishers, San Diego.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*, 2nd Ed., Prentice Hall, New York.

Horng, M.-H. (2010). "Performance evaluation of multiple classification of the ultrasonic supraspinatus images by using ML, RBFNN and SVM classifiers." *Expert Syst. Appl.*, 37(6), 4146–4155.

Hubert, M., and Van Driessen, K. (2004). "Fast and robust discriminant analysis." *Comput. Stat. Data Anal.*, 45(2), 301–320.

IBM. (2010). *PASW Modeler*, IBM Corporation, New York.

Irwin, T. C. (2007). *Government guarantees-allocating and valuing risk in privately financed infrastructure projects*, The World Bank, Washington, DC.

Jones, D. (2006). "Construction project dispute resolution: options for effective dispute avoidance and management." *J. Prof. Issues Eng. Educ. Pract.*, 132(3), 225–235.

Kassab, M., Hegazy, T., and Hipel, K. (2010). "Computerized DSS for construction conflict resolution under uncertainty." *J. Construct. Eng. Manage.*, 136(12), 1249–1257.

Katos, V. (2007). "Network intrusion detection: Evaluating cluster, discriminant, and logit analysis." *Inf. Sci. (N. Y.)*, 177(15), 3060–3073.

Keith, F. S. (1997). "Alternative dispute resolution in government." *J. Manage. Eng.*, 13(5), 25–28.

Kim, H.-J., and Shin, K.-S. (2007). "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets." *Appl. Soft Comput.*, 7(2), 569–576.

Kim, Y. S. (2010). "Performance evaluation for classification methods: A comparative simulation study." *Expert Syst. Appl.*, 37(3), 2292–2306.

Klivans, A. R., and Servedio, R. A. (2006). "Toward attribute efficient learning of decision lists and parities." *J. Mach. Learn. Res.*, 7(Compendex), 587–602.

Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection." *IJCAI'95 Proc., 14th International Joint Conference on Artificial Intelligence*, Vol. 2, Morgan Kaufmann Publishers Inc., San Francisco, 1137–1143.

Lee, M.-C. (2009). "Using support vector machine with a hybrid feature selection method to the stock trend prediction." *Expert Syst. Appl.*, 36(8), 10896–10904.

Li, L., Jiang, W., Li, X., Moser, K. L., Guo, Z., Du, L., Wang, Q., Topol, E. J., Wang, Q., and Rao, S. (2005). "A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset." *Genomics*, 85(1), 16–23.

Loh, W.-Y., and Shih, Y.-S. (1997). "Split selection methods for classification trees." *Statistica Sinica*, 7, 815–840.

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.

Nandi, S., Badhe, Y., Lonari, J., Sridevi, U., Rao, B. S., Tambe, S. S., and Kulkarni, B. D. (2004). "Hybrid process modeling and optimization strategies integrating neural networks/support vector regression and genetic algorithms: study of benzene isopropylation on Hbeta catalyst." *Chem. Eng. J.*, 97(2–3), 115–129.

Public Construction Commission (PCC). (2010). *The strategies to the promotion of public-private partnerships*, Executive Yuan, Taipei, Taiwan (R.O.C.) (in Chinese).

Public Construction Commission (PCC). (2011). *Engineering evaluation forum of PPP strategy*, Executive Yuan, Taipei, Taiwan (R.O.C.) (in Chinese).

Pulket, T., and Arditi, D. (2009a). "Construction litigation prediction system using ant colony optimization." *Constr. Manage. Econ*, 27(3), 241–251.

Pulket, T., and Arditi, D. (2009b). "Universal prediction model for construction litigation." *J. Comput. Civ. Eng.*, 23(3), 178–187.

Quinlan, J. R. (2007). "C5." ⟨http://www.rulequest.com/see5-unix.html⟩.

Shmueli, G., Patel, N. R., and Bruce, P. C. (2007). *Data mining for business intelligence*, Wiley, Hoboken, NJ.

Sokolova, M., and Lapalme, G. (2009). "A systematic analysis of performance measures for classification tasks." *Inf. Process. Manage.*, 45(4), 427–437.

Solino, A. S., and Vassallo, J. M. (2009). "Using public-private partnerships to expand subways: Madrid-Barajas International Airport case study." *J. Manage. Eng.*, 25(1), 21–28.

SPSS. (2007). *Clementine 12.0 algorithms guide*, Integral Solutions Limited, Chicago.

Tan, P.-N., Steinbach, M., and Kumar, V. (2006). *Introduction to data mining*, Pearson Education, Upper Saddle River, NJ.

Vapnik, V. N. (1998). *Statistical learning theory*, Wiley, New York.

World Bank Group (WBG). (2011). "Private participation in infrastructure database." ⟨http://ppi.worldbank.org/resources/ppi_glossary.aspx⟩ (Apr. 5, 2011).

Witten, I. H., and Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann, San Francisco.

Wu, C.-H., Tzeng, G.-H., and Lin, R.-H. (2009). "A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression." *Expert Syst. Appl.*, 36(3, Part 1), 4725–4735.

Wu, Q. (2010). "The hybrid forecasting model based on chaotic mapping, genetic algorithm and support vector machine." *Expert Syst. Appl.*, 37(2), 1776–1783.

Zhang, X. (2005). "Financial viability analysis and capital structure optimization in privatized public infrastructure projects." *J. Construct. Eng. Manage.*, 131(6), 656–668.