

Research Article

Naïve Bayesian Classifier for Selecting Good/Bad Projects during the Early Stage of International Construction Bidding Decisions

Woosik Jang,¹ Jung Ki Lee,¹ Jaebum Lee,² and Seung Heon Han¹

¹School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Republic of Korea

²Department of Civil and Infrastructure, Hyundai Engineering Co., Ltd., Seoul 140-2, Republic of Korea

Correspondence should be addressed to Seung Heon Han; shh6018@yonsei.ac.kr

Received 11 April 2015; Revised 17 July 2015; Accepted 26 July 2015

Academic Editor: Mohamed Marzouk

Copyright © 2015 Woosik Jang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the 1970s, revenues generated by Korean contractors in international construction have increased rapidly, exceeding USD 70 billion per year in recent years. However, Korean contractors face significant risks from market uncertainty and sensitivity to economic volatility and technical difficulties. As the volatility of these risks threatens project profitability, approximately 15% of bad projects were found to account for 74% of losses from the same international construction sector. Anticipating bad projects via preemptive risk management can better prevent losses so that contractors can enhance the efficiency of bidding decisions during the early stages of a project cycle. In line with these objectives, this paper examines the effect of such factors on the degree of project profitability. The Naïve Bayesian classifier is applied to identify a good project screening tool, which increases practical applicability using binomial variables with limited information that is obtainable in the early stages. The proposed model produced superior classification results that adequately reflect contractor views of risk. It is anticipated that when users apply the proposed model based on their own knowledge and expertise, overall firm profit rates will increase as a result of early abandonment of bad projects as well as the prioritization of good projects before final bidding decisions are made.

1. Introduction

In 2013, the international construction industry generated USD 8.9 trillion in annual construction spending [1]. The compound annual growth rate is predicted to reach 4.1% by 2021, suggesting future steady growth and thus implying that more contractors will enter the international construction market. Korean contractors have attained the highest revenues in this market, in which total international revenues over the last five decades have reached USD 600 billion [2]. In recent years, this volume has reached more than USD 70 billion per year, placing Korea sixth among countries generating the highest international revenues. In achieving this level of success, large Korean contractors have aggressively entered the international market, competing against other contractors from around the world.

However, given the presence of market volatility and other forms of uncertainty, particularly in developing and underdeveloped countries, international construction projects still carry a high degree of risk exposure, which affects

the overall financial soundness of the contractor. Korean international construction profit rate data show that the average dropped from 4.7% to 3.8% and then to 3.2% from 2010 to 2012. As Korean contractors have actively penetrated several countries and won bids for several international projects, the number of loss projects has also simultaneously increased, causing contractors to experience financial instability. According to actual data related to Korean contractors, as shown in Figure 1, 15% of approximately 300 projects were severely underperforming projects, and these projects accounted for 74% of all losses [2], demonstrating that a small portion of loss projects causes the majority of financial losses among international projects. For instance, Samsung Engineering's annual report [3] showed an increase in total revenues of KRW 9.8 trillion. However, the report revealed operating profit losses of KRW 1.03 trillion in 2013, which was mainly the result of a few major construction projects such as the UAE refinery project, Saudi gas projects, and the aluminum project. Despite the successful bidding that increases firms' overall revenues, a small number of projects

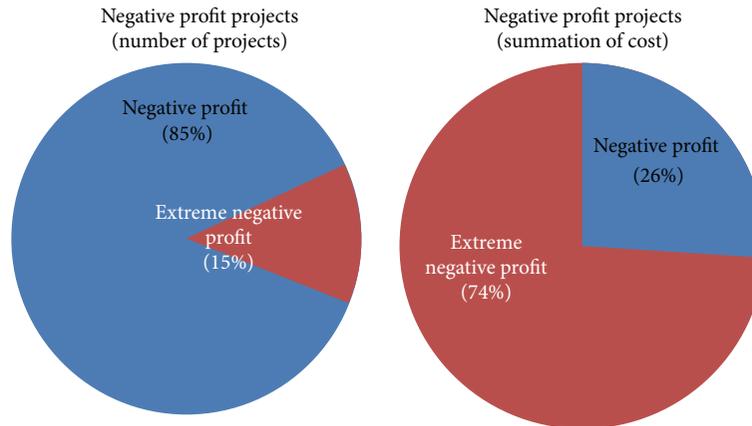


FIGURE 1: Distribution of bad projects in international construction.

such as these which cause negative profits after construction are completed. Other Korean firms have also experienced significant earnings shocks in recent years, illustrating the importance of screening projects that cause such losses.

Risk management has become central to successful project management, particularly in an uncertain international construction environment [4–7]. However, approaches to risk management focus primarily on the engineering and construction phases, that is, after a bidding decision has been made [8–12]. Moreover, these approaches are often limited by a reliance on historical data, which largely rely on the subjective judgments of users and hence do not provide comprehensive support for all levels of industry [13, 14]. To overcome these limitations, this paper examines retrospective data to elicit implicit meaning from the actual performance of previous cases. The paper also identifies risk factors that affect the profitability of international projects and examines the relationship between these factors and project profitability. To this end, the authors employ the Naïve Bayesian classifier using binomial variables, which improve practical applicability and thus the burden of screening out bad projects through simple yet accurate methodology and results for industry practitioners that are similar to actual project performances. Through the lens of this relationship, the proposed model evaluates whether a project is likely to be good or bad based on limited information available before final bidding decisions are made.

The procedure for developing a project classification model is shown schematically in Figure 2. First, the content of previous research is examined to enhance the theoretical aspects of this study. Based on a review of previous approaches, the limitations of existing strategies are also identified.

Second, risk attributes that affect the soundness of construction projects in the early stages are identified in interviews with relevant experts; risk types are categorized based on this information. Third, case-based surveys are conducted to relate risk attributes with project profitability. Case projects were performed from 1999 to 2010 by Korean international contractors. Questionnaires assessed risk attributes of a given

project and actual levels of profitability, forming the basis of the proposed model.

Third, the correlations between each risk attribute level and project profitability are analyzed using a probabilistic classification model. To enhance the practical applicability of the model, the Naïve Bayesian classifier is used as a proper probabilistic tool to support the screening of good/bad projects in the early stages of a project. The advantages and detailed structure of this approach are also described.

Finally, cross validation is employed to test its usability. To validate the project classification model, actual project data are used to compare results obtained from the model with actual performance data, testing the accuracy and reliability of the proposed model.

2. Research Background

Risk management has been widely investigated by various researchers focusing on different phases of construction for the management scope. For example, Hastak and Shaked [15] developed a model based on country, firm, and risk assessment projects. Han et al. [16] also developed a fully integrated risk management system (FIRMS) that considers the enterprise and individual risk management project level. Furthermore, various methodologies have been used to develop risk management systems, including case based reasoning [17], artificial neural networks [18], the total risk index [19, 20], and the fuzzy analytic hierarchy process [4, 21].

As international construction volumes have grown rapidly in recent decades, various studies have assessed project conditions based on the assumption that a full risk assessment of a given project can be performed prior to the bidding stage by analyzing the information collected. However, in reality, it is impossible for contractors to consider all of the risks that affect project profitability at this early stage. Accordingly, del Caño and de la Cruz [22] identified certain project characteristics, including complexity, size, and owner capabilities and suggested that risk analysis results should be derived based on project characteristics. The authors also emphasized that an appropriate risk management method

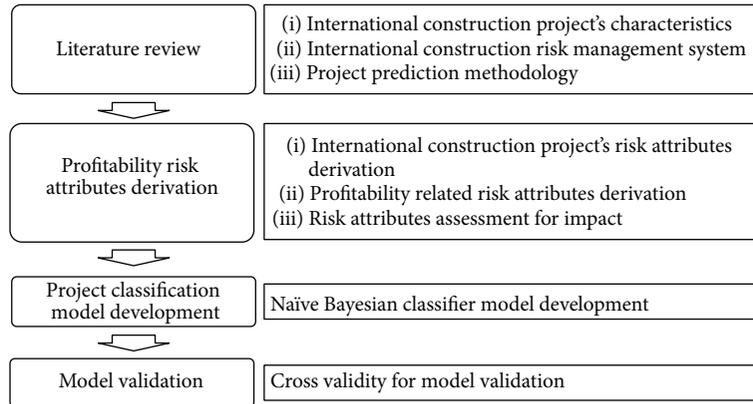


FIGURE 2: Research procedure.

based on representing attributes should be employed for effective project screening modeling. Thus, a systematic methodology that predicts project profitability during this early stage is imperative.

For developing the prediction models, the various theories and tools can be categorized as statistical or learning methods. Several researchers have used regression models, through which the impact of each risk attribute is assessed on a Likert scale by experts [23, 24]. However, such assessment is usually plagued by ambiguousness from maintaining consistency with different standards depending on expert's cognitive limitations or the cooperation cultures during surveys. In addition, input data usually include subjective assessments that often distort practitioner's decision-making and judgment [15, 25]. As such, previous studies that have attempted to classify profit levels by Likert or numeric scale have not yet reached a statistical level of significance in profit prediction due to their compound mechanism and a large number of inputs for understanding complicated project information [15, 26]. To develop a more reliable prediction model and improve its applicability for industry practitioners, expert assessments should be consistent, and a more simplified risk assessment model is imperative.

Data-based learning tools have also been widely applied to support the reliability of classifications such as Matlab for Artificial Neural Network (ANN), C5.0 for decision tree, and Linear Discriminant Analysis (LDA). ANN and C5.0 are considered representative data mining techniques for detecting dependent variables from the loads of input data with complex calculations [27, 28]. LDA utilizes the linear functions between various classes where the data can be categorized according to the appropriate number of classes and its relationships [29]. Unfortunately, these data-mining tools do not properly explain the underpinning rule of prediction results. In addition, these methods strictly require a large number of data for improved prediction. Hence, the Naïve Bayesian classifier is considered a suitable method for this research that uses a relatively small amount of historical data to determine whether a given project is likely to be bad or good.

To sum up, previous studies on profit prediction have been limited to simple Likert scale assessments based on experts' subjective judgments of contractor experience, potentially generating skewed results. These not only contain unrealistic assumptions and a high burden for practical applicability but also show limited accuracy when such complex calculations with a large number of input data are utilized [29]. The authors thus present a simple binary classification model based on the updated conditional probability data from the actual project's performance. This approach is expected to reduce bias while improving practical applicability.

3. Risk Attributes of Project Profitability

Project classification model development begins by incorporating overall risk attributes for international projects that are present prior to early bidding stages. Risks identified as important in previous studies were selected and analyzed via meta-analysis and frequency measurements. Studies by Hastak and Shaked [15], the American Construction Industry Institute [30], the International Construction Association of Korea [2], and Han et al. [16, 25] identified risk factors that affect construction project profitability. In addition to conducting a literature review, the authors performed interviews with eight industry experts; as a result, 20 risk factors were added to the analysis. Table 1 presents a total of 71 risk attributes derived from the aforementioned procedures.

A case-based questionnaire with refined risk attributes was designed to identify relevant information for the proposed model. The questionnaire measured the effect of 71 risk factors on profit levels on a 7-point Likert scale, in which -2 refers to a positive impact (in the sense of gain), 0 indicates no impact, 2 refers to a negative impact, and 4 refers to a critical impact on project profitability. Sample projects were chosen among actual international projects performed by the most prominent 20 Korean contractors since 2000. Approximately 900 projects were available; of these, 140 projects, which included sufficient data for prediction model development, were compiled for further analysis. The collected data were

TABLE 1: Risk attributes affecting project profitability.

Category	Risk code	Risk attribute	Cronbach's α value
Political and social risks	R1	High degree of corruption, collusion, and illegal transaction	0.787
	R2	Lack of familiarity with local cultures, customs, and traditions	
	R3	Frequent policy and postlegislative changes	
	R4	Insufficient legal system	
Economic and inflation risks	R5	High material unit price volatility	0.735
	R6	High equipment unit price volatility	
	R7	High local currency exchange rate volatility	
Owner risks	R8	Unsatisfactory owner management capabilities	0.869
	R9	Insufficient design information from the owner	
	R10	Lacking specifications and bidding documents provided by the owner	
	R11	Approval and permission delays on documentation and construction	
	R12	Illegal requirements without contracts	
Bidding and contract risks	R13	Insufficient time for bidding and construction quotation preparation	0.834
	R14	Inadequate consideration of variations in contract duration	
	R15	High-level local content requirements	
	R16	Complex tax and tariff regulations	
	R17	Low-level inflation compensation conditions	
	R18	Inadequate warranty-related conditions	
	R19	Inadequate payment regulations relative to international contracts	
	R20	Ineffective dispute resolution procedures and regulations	
Procurement risks	R21	Inadequate local labor supplies and procurement conditions	0.905
	R22	Inadequate local labor supplies and demand conditions	
	R23	Inadequate material production supplies and procurement conditions	
	R24	Inadequate material production supplies and demand conditions	
	R25	Inadequate material production from third-world-country supplies and demand conditions	
	R26	Inadequate equipment from local country supplies and procurement conditions	
	R27	Inadequate local equipment supply procurement and demand conditions	
	R28	Inadequate local subcontractor procurement conditions	
Geographic risks	R29	Inadequate geographical conditions of materials and equipment delivery	0.853
	R30	Inadequate logistical conditions (i.e., electricity, water, gas, and communication)	
	R31	Inadequate business environment and welfare conditions	
	R32	Inadequate weather conditions for construction	
	R33	Inadequate geological site conditions	
	R34	Severe complaints regarding construction throughout project implementation	

TABLE 1: Continued.

Category	Risk code	Risk attribute	Cronbach's α value
Construction performance risks	R35	Lacking design experience	0.917
	R36	Lacking bid estimation experience	
	R37	Unsatisfactory cash flow management capabilities	
	R38	Unsatisfactory site manager leadership and organization management capabilities	
	R39	Unsatisfactory site manager experience relative to similar construction projects	
	R40	Unsatisfactory international project planning and management capabilities	
	R41	Unsatisfactory quality management capabilities	
	R42	Unsatisfactory resource management capabilities	
	R43	Unsatisfactory conflict management capabilities	
	R44	Unsatisfactory contract management capabilities	
	R45	Lacking connectivity between sites and headquarters	
	R46	Unsatisfactory localization	
	R47	Lacking trust between project participants	
	R48	Poor communication between project participants	
	R49	Poor project management capabilities based on IT	
	R50	Inadequate construction methods for the specific project	
	R51	Insufficient subcontractor project performance	

first tested for reliability, stability, consistency, predictability, accuracy, and dependency [31]. Cronbach's coefficient alpha internal consistency method was employed, and results ranging from 0.8 to 0.9 were considered reliable. The PASW Statistics 21 program was used to analyze the data. As Table 1 shows, missing values in the raw data were excluded from the reliability test to increase reliability. As a result, a total of seven categories with 51 attributes were selected; an average reliability of 0.843 was obtained, indicating a high level of reliability. A total of 121 datasets for the 140 projects were used to develop the Naïve Bayesian classifier, indicating that approximately 20% of the responses were excluded due to data inconsistency.

4. Naïve Bayesian Classifier

A Naïve Bayesian classifier based on Bayes' theorem is employed to derive a probabilistic classification model from previously assessed attributes. It is most commonly used for antispam mail filtering, which trains the filter to automatically separate spam mail and legitimate messages in a binary manner [32, 33]. In doing so, each attribute corresponds with words included in the spam and legitimate mails, thus distinguishing between the two; in turn, the former are blocked by a trained filter.

4.1. Conceptual Structure of a Proposed Model. Because the objective of this research is to screen bad or good projects in the early stages of international construction projects, the result of the proposed model should also be binary (good or bad). Thus, the authors transformed the original conditions

of each risk attribute (7-point Likert scale) into a binary scale to develop the Naïve Bayesian model. In addition, the early stages of the construction possess a high level of uncertainties which prohibit analysis of exact ranges of profit rates or detailed numeric classification of project conditions. Accordingly, a specific profit level or numeric number of each risk attribute is not essential, particularly during this early phase. To this end, authors consider the unity of the binary distribution of the dependent variable as well as the preassessed risk attributes to improve its practical application.

The benefits of employing the Naïve Bayesian classifier are the following [34–36]. First, the model is less sensitive to outliers. Generally, when an outlier exists in the data, the results may skew. The reduction of outlier effects is thus required for data analysis. Because the Naïve Bayesian classifier uses a probabilistic distribution, it enables minimizing the outlier effects of the original data and prediction results are less sensitive to outliers. Second, less data is required to predict the parameter. By using the utility function, relearning the probability of a given conditional probability is not strictly required in the Naïve Bayesian classifier, thus minimizing data collection efforts for risk prediction. Third, the classifier is more applicable to the industry practitioners by reducing the burden of data collection. Although the model utilizes simple assumptions and algorithms, the classification results are powerful even though the specified assumptions may contain errors. Finally, the models allow for the consideration of additional attributes by combining existing characteristics of the attributes, thus increasing the accuracy of the model's predictions. Based on the aforementioned benefits of the Naïve Bayesian method, the authors have chosen this

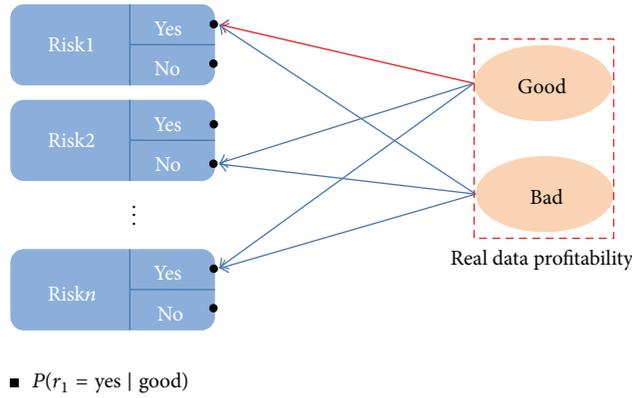


FIGURE 3: Conceptual model of the Naïve Bayesian classifier.

approach as the main mechanism for the proposed prediction model.

To structure the Naïve Bayesian classifier, we considered the unity of the binary distribution of the dependent variables. As mentioned earlier, the effect of each risk attribute on the level of profit is assessed by using a 7-point Likert scale (−2 to 4) to improve the model's practical application. In this study, a binary approach is used to identify the presence of risk, and final project profitability is predicted via the Naïve Bayesian classification. Because this research is focused on the early stages of the construction project, screening and recognizing a bad project are the priority rather than analyzing the specific value of the project. Therefore, the assessment range from −2 to 1 correlated with absence of risk or a “No” answer, and a score from 2 to 4 was correlated with the presence of risk or a “Yes” answer, as shown in Figure 3.

As the assessment approach is fairly simple, users perform fewer risk assessment tasks, increasing the accuracy of predictions. The benefit of a Naïve Bayesian classifier is the fact that all risk probabilities are assumed to be independent; according to Murphy [35], although these assumptions may not be correct, developing a model using the Naïve Bayesian classifier is feasible and generates highly reliable results for classification compared to other methodologies.

Equation (1) presents the classification group c set (i.e., good or bad project), where $r_1, r_2, r_3, \dots, r_D$ are vectors ascribed to each independent risk variable in the Naïve Bayesian classifier. This categorization of attributes maximizes the probability of each classification result. Consider

$$P = (c \mid r_1, r_2, \dots, r_D). \quad (1)$$

Equation (2) shows the prediction mode that applies (1) to Bayes' theorem. Consider

$$\frac{P(c) \times P(r_1, r_2, \dots, r_D \mid c)}{P(r_1, r_2, \dots, r_D)}. \quad (2)$$

As discussed above, the Naïve Bayesian classifier assumes that each attribute divided into each classification is independent; thus, (3) can be derived accordingly, where v is the target value of the group c set (i.e., the profit rate). Consider

$$P(r_2 \mid c, r_1) = P(r_2 \mid c),$$

$$P(r \mid v=c) = \prod_{i=1}^D P(r_i \mid v_j=c). \quad (3)$$

Hence, the Naïve Bayesian classifier is considered a suitable method that uses a relatively small amount of historical data to determine whether projects are likely to be bad or good. The next section describes the proposed model, which uses the Naïve Bayesian classifier to examine the actual project data collected previously.

4.2. Naïve Bayesian Project Classification Model. In developing the Naïve Bayesian model, the aforementioned risk attributes were used as relevant independent variables for predicting good or bad projects. An overview of the model is presented in Figure 4.

The dependent variable was predicted over two or three classification intervals, and the two models are presented accordingly. The model with two intervals is based on a profitability of less than 0% for a bad project and more than 0% for a good project. The model with three intervals represents projects with less than 0% profitability—those generating a 0% to 4.5% rate of return—and a profit rate above 4.5%, which is considered a good project. A marginal profit rate of 4.5% was chosen because the average profit rate for good projects out of the total sample of projects was calculated at this range. Because projects with negative profitability are all considered “bad,” “moderate” and “good” projects are differentiated based on the reference profit value of 4.5%. Table 2 represents intervals divided based on the results.

Using the training dataset, each risk attribute is categorized into a specific class to develop the project-screening model. Therefore, the conditional probability of each attribute

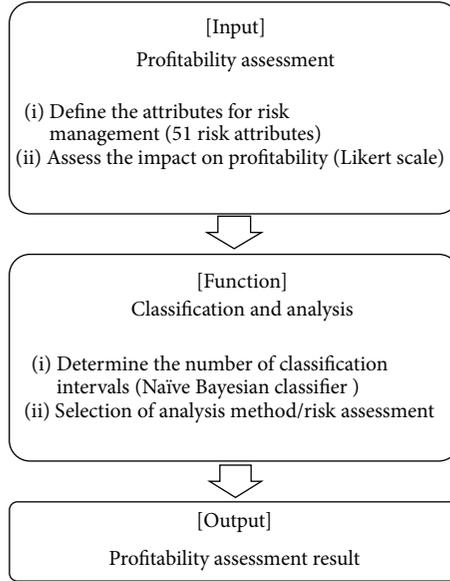


FIGURE 4: Overview of project classification model.

TABLE 2: Classification of profitability.

	Bad project	Moderate project	Good project
2 classification intervals	Below 0%	—	Above 0%
3 classification intervals	Below 0%	0%–4.5%	Above 4.5%

given to the bad project in advance is shown in the following equation:

$$P(r_i | v_j) = \frac{n_c}{n}, \quad (4)$$

P is estimated conditional probability, r_i is risk attributes i , v_j is the condition of profit rate on Bad or Good based on the training data ($v = \text{below } 0$), n_c is number of bad projects under the specific risk a_i in the training data ($v = v_j$ and $r = r_i$), and n is number of bad projects in the training data ($v = v_j$).

However, whereas this study examines 140 units of data, the number of training data is insufficient relative to the number of risk attributes. Unbalanced datasets also lead to an inappropriate result in the statistical analysis [37]. It is thus necessary to augment the small datasets in the training set which decreases the accuracy of the model. To overcome this, the authors borrowed the concept of the m -estimate from [38–40]. By setting the value of a suitable “ m ” number, the shifting of probability toward parameter “ p ” is controlled; this also further increases model accuracy as shown in the following equation:

$$P'(r_i | v_j) = \frac{n_c + mp}{n + m}, \quad (5)$$

P' is estimated conditional probability using m -estimate, m is equivalent data size, and p is prior probability based on training data (4).

Zadrozny and Elkan [38] recommend that the value of mp be set to less than 10, thus using a probabilistic equation to cross-validate 140 data units; 70 units are set as training data, and the remaining 70 units of m parameter are set to an equivalent data size to develop the project screening model. In turn, parameter p becomes $10/70$ or $1/7$ for (5). Of the 140 data units collected, 20 units were therefore classified as bad projects, and 120 were deemed good projects. Therefore, the two datasets—training and equivalent data size—consist of 10 bad projects and 60 good projects that are chosen via random selection.

Five different datasets are created randomly to develop and validate the model in view of cross validation. Random data are selected using Microsoft Excel, which generates data in a range from 0 to 1 for random probability. In addition, for the training dataset, the actual risk exposure level for each attribute in a given project is used to develop the model. In this paper, only one set of training data results out of five different datasets is presented in detail, as the other results were found to be almost equivalent.

The Naïve Bayesian classifier consists of an m -estimator parameter, which gives a random value of m to control the classification bias. This study uses m value from the “moderate” interval of project profitability, which had the highest error percentage, for further validation. According to the results of the first analysis, m value in the Naïve Bayesian classifier was reduced in the second analysis: by modifying m value, the moderate profitability project produces a more stable result. In addition, the weights given to the bad projects increased, which results in both increased stability of the classification and reduction of bad project’s classification error. Therefore, the analysis of the moderate classification results was reanalyzed. To identify a suitable m value, a sensitivity analysis was also performed for the changes as m value decreases.

TABLE 3: Part of the conditional probability of Naïve Bayesian classifier model.

Category	Risk code	$P'(r_i \text{BAD})$	$P'(r_i \text{GOOD})$
Political and social risks	R1	35%	23.4%
	R3	15%	5.6%
Economic and inflation risks	R5	15%	8.9%
	R7	20%	24.2%
Bidding and contract risk	R15	10%	27.4%
	R19	15%	15.3%
Procurement risks	R22	30%	36.2%
	R24	5%	18.5%
Geographic risks	R29	40%	22.6%

By using the derived model using m -estimate, each risk factor is calculated for the conditional probability for Bad and Good. The two classification intervals model is used to analyze each risk factor. Table 3 shows the conditional probability of the key risk factors from each category that has a severely negative impact (assessed as 3 or 4 points) on the total sample data (140 projects). For example, the R24 risk had an 18.5% probability of occurring in a “Good” project and a 5% probability of occurring in a “Bad” project. To summarize Table 3, external (political or economic) factors more frequently appear in the “bad” projects, while internal (bidding or procurement) factors more frequently occur in the “good” projects.

5. Model Analysis and Validation

5.1. Model Validation. In employing the classifier model, two datasets (training and sample test datasets) for 70 projects were used to compare actual and predicted risk impacts on the dependent variable. Actual risk effects in a given project were used to determine the profitability of the 70 projects. In addition to evaluating the performance level of the model, the model’s usability and accuracy were tested using a confusion matrix of Bayes’ theorem. To determine the accuracy of the model, the dependent variable, profitability, included a sample test dataset of 70 projects for comparison with the actual project performance and predicted performance of a given project. The 2 and 3 classification intervals models were also tested by classification into bad (PB: Predicted Bad), moderate (PM: Predicted Moderate), and good (PG: Predicted Good) categories; the training dataset for the 70 projects also employs the actual profit level, which is classified as bad (AB: Actual Bad), normal (AM: Actual Moderate), or good (AG: Actual Good). Tables 4 and 5 present the results of the two models.

The aforementioned five models were randomly tested for cross validation. In the test models, a dataset of 70 projects was randomly selected to determine the model’s accuracy. As an initial measure, each project was deemed bad or good. As shown in Table 4, the overall correct level is found to be 79.7% by dividing the bold numbers with the total number of

TABLE 4: Validation of profitability at 2 classification intervals.

Model 1 out of five cross validations	Classification level			Total
	Bad (PB)	Good (PG)		
Actual Level				
Bad (AB)	5	5		10
Good (AG)	9	51		60
Total	14	56		70

TABLE 5: Validation of profitability at 3 classification intervals.

Model 1 out of five cross validations	Classification level			Total
	Bad (PB)	Moderate (PM)	Good (PG)	
Actual Level				
Bad (AB)	3	5	2	10
Moderate (AM)	3	16	13	32
Good (AG)	3	13	12	28
Total	9	34	27	70

samples. Other cross validation models also showed highly accurate levels of 78.6%, 77.1%, 82.7%, and 80%, respectively.

In contrast, the model with three classified intervals presented relatively low levels of 44.3%, 42.9%, 51.4%, 48.6%, and 50.0%, with a mean percentage of 44.8%, reflecting less accurate results (refer to Table 5). Models with 2 and 3 classification intervals showed standard deviations of 2.07 and 3.07, respectively; this suggests the presence of a few differences and consistent results between the two models. The 2 classification intervals model is considered an excellent vehicle for screening purposes, as it presents highly accurate results and less deviation. Although the 3 classification intervals model shows a relatively low level of accuracy, in a real business environment, the users also may apply the 3 classification intervals model for screening out a moderate range of projects by sacrificing the prediction accuracy.

Additional analysis to validate the Naïve Bayesian classification model is performed through comparing with linear regression model using SPSS 21. The authors used the same binary method of dividing the profit rate into good and bad (2 classification intervals). By utilizing the same 140 projects’ units of data that is composed of 70 training units of data and 70 test units of data, the accuracy result of linear regression model is found to be 52.2%. This result shows a significantly lower accuracy than the 2 classification intervals Naïve Bayesian model (79.7%) and slightly higher accuracy than the 3 classification intervals model (44.8%).

In summary, this study proposed two types of models depending on the user’s need and degree of risk exposure. The models were validated through comparing the accuracy of the prediction result with other learning methods. In addition, the user can flexibly select the proper model that best suited to a given firm, whether he or she requires an accurate classification between good and bad project or is contend to identify a moderate or good project.

Profitability assessment			
Method	1	Naïve Bayesian classifier	
Category	Risk code	Risk attributes	Risk
Country's political and social risk	R1	Country's corruption, collusion, and illegal transaction	O
	R2	Culture, customs, and tradition difference	X
	R3	Policy and postlegislative changes	X
	R4	Country's overall legal system's maturity	X
Country's economic and inflation risks	R5	Average material unit price change	X
	R6	Average equipment unit price change	X
	R7	Local currency exchange rate change	X
Stakeholder risks	R8	Stakeholder's understanding and management ability	O
	R9	Design error provided to stakeholder	O
	R10	Clarity of specifications and bidding documents provided to stakeholder	X
	R11	Stakeholder's administrative and construction consent delay	X
	R12	Requirements regardless of contract from stakeholder	X
Bidding and contract terms	R13	Bidding and quotation preparation time for construction	X
	R14	Contract of construction duration for construction	X
	R15	Required degree of local contents	X
	R16	Tax and tariff regulations	X
	R17	Inflation compensation conditions compared to international contract	X
	R18	Warranty-related conditions compared to international contract	X
	R19	Payment regulations compared to international contract	X
	R20	Dispute resolution procedures regulation compared to international contract	X

FIGURE 5: Excel-based risk assessments.

5.2. *Excel-Based Model Development.* Based on the aforementioned basic algorithm, the authors developed an Excel-based system that serves as a continuous risk management platform for project screening for the international construction sector. To initiate project screening, risk attributes related to a given project are assessed for their impact on project profit depending on the amount of information acquired. As shown in Figure 5, system inputs for risk assessment are based on two scenarios: the specified risk either affects the project or does not. In addition, users may select their model depending on their objectives: the more accurate Naïve Bayesian classifier with a higher project classification rate (2 classification intervals model) or the less accurate model with a more conservative classification rate (3 classification intervals model).

The authors present a sample case in which the developed program is utilized to test the project screening model through the user interface. This case considers a Korean contractor working on a harbor expansion project that lasts 22 months. Figure 5 shows step 1, which is the actual user

interface of the program generated with case study data inputs for good/bad project screening. As shown in Figure 5, the user first evaluates the risk attribute using “O” for “risk exists,” indicating the presence of a given risk, and “X” for “risk does not exist,” indicating an absence of project risk. The user then selects a model, that is, the number of classification intervals. Depending on the model selected, the classification model categorizes projects as bad, moderate, or good. The model classifies the good projects as ranging from 0% to 4.5% of the profit rate. Project screening is conducted by selecting profit interval ranges depending on firm risk behavior. Model results are obtained from the training sample via random selection, through which classification results are selected from the highest predicted value from 100 prediction cycles. This process increases reliability, and the results are shown in percentage form for ease of comprehension.

Figure 6 shows the procedural steps in the process of the Excel-based model. Risk assessment is performed primarily in step 1. Then, in step 2, the model counts for the risk existence based on the historical dataset of 140 real projects.

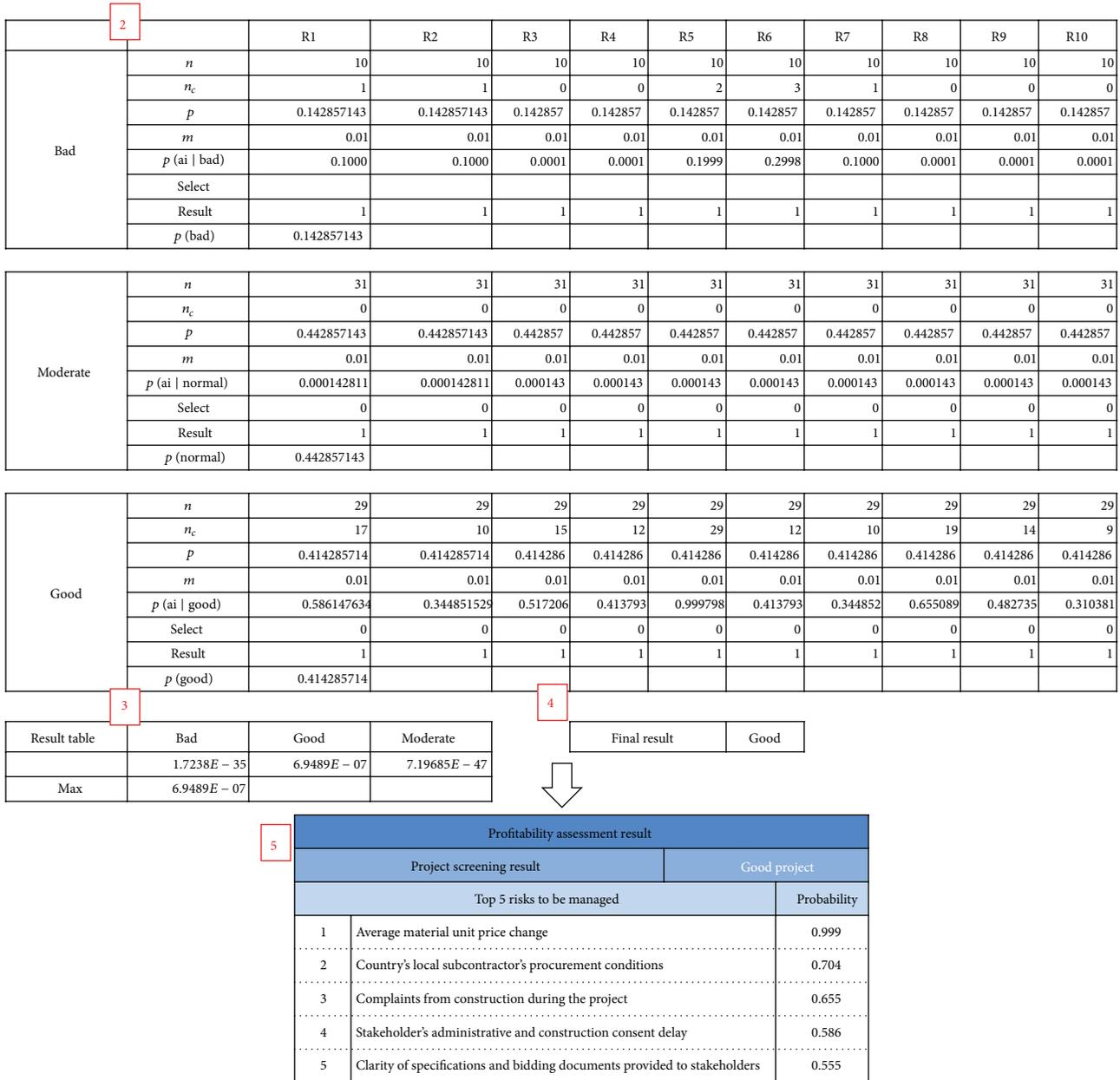


FIGURE 6: Process for assessment result.

Through the sum of its risk existence, the conditional probability is calculated given each classification of a project (Good, Moderate, or Bad) separately.

Step 3 shows the result table for each classification that has been selected from the profitability assessment. This is the step in which users are to perform the risk assessment from Figure 6, where selected risks are calculated for conditional probability. The result is calculated from the product of each probability from the 51 risks attributes. Step 4 then shows the final result by selecting the highest probability from each classification. In the illustrative example, “Good” had the highest probability. In addition, as shown in step 5 in the user interface, the risks that most directly affect project profitability can be prioritized, which should be

continuously monitored and controlled by the user. These five risk attributes serve as the basis of a risk response strategy.

The proposed model prioritizes the top 5 risks that affect overall construction profitability. According to Halawa et al. [41], construction projects are dramatically influenced by this financial evaluation and awareness of the financial status improves; it offers a better chance of determining project feasibility. Therefore, this study incorporates the Naïve Bayesian classifier with an Excel-based model to give practitioners a quantitative analysis result (statistical references) and qualitative analysis results (the top risk attributes to be managed). It should also be noted that although the result of the model shows an absolute value of bad, moderate, or good, it is up to the managers and decision-makers to make a bad project,

a moderate project, or even good project by careful and proactive risk management as presented in the illustrative case application.

6. Conclusions

This study predicted the profitability of early-stage international construction projects and identified good projects that may affect firm's financial stability. The current state of the international construction market was first analyzed; losses were shown to account for 73% of the entire market, reflecting the importance of predicting project profitability in advance. Risk attributes that affect project profitability were identified, and a risk management system was developed to manage, prioritize, and monitor higher impact risks. The Naïve Bayesian classifier method was used to improve the practical application of the project-screening model. When risk attribute assessments for determining risk exposure levels are completed, assessed risk exposure is classified into a binary parameter that presents an approach to risk management that is supported by a probabilistic methodology.

This study used case study data to create a model using Microsoft Excel. Depending on the Naïve Bayesian classifier m -estimator value of the weight application and the user's risk behaviors, users can choose the model that best suits a given firm. The results show that when a 70-unit training dataset and 70-unit sample dataset area are used (a 140-unit set in total) to cross-validate the Naïve Bayesian classifier, the model has on average a 79.7% prediction accuracy for the 2 classification intervals model and a 44.8% accuracy for the 3 classification intervals model. Additional analysis to validate the model is performed through linear regression analysis to show the higher precision result from the Naïve Bayesian classification model.

For project screening, conditional probability is used with information acquired during early project stages. Although risk attributes were largely derived in the model, risk attributes that more clearly affect projects can be further extracted, elucidating the effects of bad projects.

As for the limitations of this study, analytical results for the application of the developed program to a more diverse scenario will be essential; for example, different types of construction projects may be examined for superior results. In addition, if contract types were categorized as contractor, subcontractor, and joint ventures, this may have altered the profitability of each project. A new-entry market scenario will also be analyzed separately due to the unique characteristics of new markets. Given a range of unique construction project qualities, a more realistic and practical application can be obtained in future research. Furthermore, if additional data on both project risk assessment and the actual reliable profit rates are accumulated, a diverse segmentation of the intervals can be obtained to examine each project in a precise measure.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by a grant (14IFIP-B089065-01) funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

- [1] Global Insight, *Global Construction Outlook*, Global Insight, Lexington, Mass, USA, 2014.
- [2] International Contractors Association of Korea (ICAK), *International Construction Information Service. Construction Statistics*, 2013, http://www.icak.or.kr/sta/sta_0101.php.
- [3] Samsung Engineering, *Business Results Report: Third Quarter of 2013*, Samsung Engineering, Seoul, Republic of Korea, 2013.
- [4] M. Abdelgawad and A. R. Fayek, "Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP," *Journal of Construction Engineering and Management*, vol. 136, no. 9, pp. 1028–1036, 2010.
- [5] G. Aydogan and A. Koksul, "An analysis of international construction risk factors on partner selection by applying ANP approach," in *Proceedings of the International Conference on Construction and Real Estate Management (ICCREM '13)*, pp. 658–669, Karlsruhe, Germany, October 2013.
- [6] B.-G. Hwang, X. Zhao, and L. P. Toh, "Risk management in small construction projects in Singapore: status, barriers and impact," *International Journal of Project Management*, vol. 32, no. 1, pp. 116–124, 2014.
- [7] A. E. Yildiz, I. Dikmen, M. T. Birgonul, K. Ercoskun, and S. Alten, "A knowledge-based risk mapping tool for cost estimation of international construction projects," *Automation in Construction*, vol. 43, pp. 144–155, 2014.
- [8] O. Taylan, A. O. Bafail, R. M. S. Abdulaal, and M. R. Kabli, "Construction projects selection and risk assessment by fuzzy AHP and fuzzy TOPSIS methodologies," *Applied Soft Computing*, vol. 17, pp. 105–116, 2014.
- [9] S. Mu, H. Cheng, M. Chohr, and W. Peng, "Assessing risk management capability of contractors in subway projects in mainland China," *International Journal of Project Management*, vol. 32, no. 3, pp. 452–460, 2014.
- [10] F. Acebes, J. Pajares, J. M. Galán, and A. López-Paredes, "A new approach for project control under uncertainty. Going back to the basics," *International Journal of Project Management*, vol. 32, no. 3, pp. 423–434, 2014.
- [11] Y. Zhang and Z.-P. Fan, "An optimization method for selecting project risk response strategies," *International Journal of Project Management*, vol. 32, no. 3, pp. 412–422, 2014.
- [12] M. Rohaninejad and M. Bagherpour, "Application of risk analysis within value management: a case study in dam engineering," *Journal of Civil Engineering and Management*, vol. 19, no. 3, pp. 364–374, 2013.
- [13] Q. Yang and Y. Zhang, "Application research of AHP in project risk management," *Advanced Materials Research*, vol. 785–786, pp. 1480–1483, 2013.
- [14] O. Kapliński, "The utility theory in maintenance and repair strategy," *Procedia Engineering*, vol. 54, pp. 604–614, 2013.
- [15] M. Hastak and A. Shaked, "ICRAM-I: model for international construction risk assessment," *Journal of Management in Engineering*, vol. 16, no. 1, pp. 59–69, 2000.
- [16] S. H. Han, W. Jung, D. Y. Kim, and H. Park, "A study for the improvement of international construction risk management,"

- Conference of Korean Society of Civil Engineering (KSCE)*, no. 10, pp. 791–794, 2009.
- [17] V. Kumar and N. Viswanadham, “A CBR-based decision support system framework for construction supply chain risk management,” in *Proceedings of the 3rd IEEE International Conference on Automation Science and Engineering (CASE '07)*, pp. 980–985, IEEE, Scottsdale, Ariz, USA, September 2007.
- [18] B. Yang, L. X. Li, H. Ji, and J. Xu, “An early warning system for loan risk assessment using artificial neural networks,” *Knowledge-Based Systems*, vol. 14, no. 5-6, pp. 303–306, 2001.
- [19] P. Sanchez, “Neural-risk assessment system for construction projects,” in *Proceedings of the Construction Research Congress*, pp. 1–11, San Diego, Calif, USA, April 2005.
- [20] B. R. Fortunato III, M. R. Hallowell, M. Behm, and K. Dewlaney, “Identification of safety risks for high-performance sustainable construction projects,” *Journal of Construction Engineering and Management*, vol. 138, no. 4, pp. 499–508, 2012.
- [21] J. Li and P. X. W. Zou, “Fuzzy AHP-based risk assessment methodology for PPP projects,” *Journal of Construction Engineering and Management*, vol. 137, no. 12, pp. 1205–1209, 2011.
- [22] A. del Caño and M. P. de la Cruz, “Integrated methodology for project risk management,” *Journal of Construction Engineering and Management*, vol. 128, no. 6, pp. 473–485, 2002.
- [23] A. P. C. Chan, D. C. K. Ho, and C. M. Tam, “Design and build project success factors: multivariate analysis,” *Journal of Construction Engineering and Management*, vol. 127, no. 2, pp. 93–100, 2001.
- [24] D. J. Lowe and J. Parvar, “A logistic regression approach to modelling the contractor’s decision to bid,” *Construction Management and Economics*, vol. 22, no. 6, pp. 643–653, 2004.
- [25] S. H. Han, D. Y. Kim, and H. Kim, “Predicting profit performance for selecting candidate international construction projects,” *Journal of Construction Engineering and Management*, vol. 133, no. 6, pp. 425–436, 2007.
- [26] A. S. Bu-Qammaz, I. Dikmen, and M. T. Birgonul, “Risk assessment of international construction projects using the analytic network process,” *Canadian Journal of Civil Engineering*, vol. 36, no. 7, pp. 1170–1181, 2009.
- [27] T. Hegazy and A. Ayed, “Neural network model for parametric cost estimation of highway projects,” *Journal of Construction Engineering and Management*, vol. 124, no. 3, pp. 210–218, 1998.
- [28] RuleQuest, Data mining tools, 2008, <http://www.rulequest.com>.
- [29] W. Jung, *Three-Staged Risk Evaluation Model for Bidding on International Construction Projects*, Graduate School, Yonsei University School of Civil and Environmental Engineering, Seoul, Republic of Korea, 2011.
- [30] Construction Industry Institute (CII), “Risk assessment for international projects,” Research Report 181-11, Construction Industry Institute (CII), Austin, Tex, USA, 2003.
- [31] G. Young, “Glossary and discussion of terms,” in *Malingering, Feigning, and Response Bias in Psychiatric/Psychological Injury*, vol. 56 of *International Library of Ethics, Law, and the New Medicine*, pp. 777–795, Springer, Dordrecht, The Netherlands, 2014.
- [32] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, “A Bayesian approach to filtering junk e-mail,” in *Learning for Text Categorization: Papers from the 1998 Workshop*, vol. 62, pp. 98–105, 1998.
- [33] I. Androutsopoulos, J. Koutsias, K. V. Chandrinou, G. Paliouras, and C. Spyropoulos, “An evaluation of Naïve Bayesian anti-spam filtering,” in *Proceedings of the Workshop on Machine Learning in the New Information Age, 11th European Conference on Machine Learning*, G. Potamias, V. Moustakis, and M. van Someren, Eds., pp. 9–17, Barcelona, Spain, 2000, <http://arxiv.org/pdf/cs/0006013.pdf>.
- [34] H. Zhang, “The optimality of naïve Bayes,” in *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS '04)*, Miami Beach, Fla, USA, May 2004.
- [35] K. P. Murphy, *Naive Bayes Classifiers*, University of British Columbia, Vancouver, Canada, 2006.
- [36] R. R. Yager, “An extension of the naive Bayesian classifier,” *Information Sciences*, vol. 176, no. 5, pp. 577–588, 2006.
- [37] D. Lowd and P. Domingos, “Naive Bayes models for probability estimation,” in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pp. 529–536, ACM, Bonn, Germany, August 2005.
- [38] B. Zadrozny and C. Elkan, “Learning and making decisions when costs and probabilities are both unknown,” in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 204–213, ACM, August 2001.
- [39] L. Jiang, H. Zhang, Z. H. Cai, and D. Wang, “Weighted average of one-dependence estimators,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 24, no. 2, pp. 219–230, 2012.
- [40] J. Wu and Z. Cai, “A naive Bayes probability estimation model based on self-adaptive differential evolution,” *Journal of Intelligent Information Systems*, vol. 42, no. 3, pp. 671–694, 2014.
- [41] W. S. Halawa, A. M. K. Abdelalim, and I. A. Elrashed, “Financial evaluation program for construction projects at the pre-investment phase in developing countries: a case study,” *International Journal of Project Management*, vol. 31, no. 6, pp. 912–923, 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

