

## Article

# Classifying the Level of Bid Price Volatility Based on Machine Learning with Parameters from Bid Documents as Risk Factors

YeEun Jang, JeongWook Son  and June-Seong Yi \*

Department of Architectural & Urban Systems Engineering, Ewha Womans University, Seoul 03760, Korea; jyee@ewha.ac.kr (Y.J.); jwson@ewha.ac.kr (J.S.)

\* Correspondence: jsyi@ewha.ac.kr; Tel.: +82-2-3277-4454

**Abstract:** The purpose of this study is to classify the bid price volatility level with machine learning and parameters from bid documents as risk factors. To this end, we studied project-oriented risk factors affecting the bid price and pre-bid clarification document as the uncertainty of bid documents through preliminary research. The authors collected Caltrans's bid summary and pre-bid clarification document from 2011–2018 as data samples. To train the classification model, the data were preprocessed to create a final dataset of 269 projects consisting of input and output parameters. The projects in which the bid inquiries were not resolved in the pre-bid clarification had higher bid averages and bid ranges than the risk-resolved projects. Besides this, regarding the two classification models with neural network (NN) algorithms, Model 2, which included the uncertainty in the bid documents as a parameter, predicted the bid average risk and bid range risk more accurately (52.5% and 72.5%, respectively) than Model 1 (26.4% and 23.3%, respectively). The accuracy of Model 2 was verified with 40 verification test datasets.



**Citation:** Jang, Y.; Son, J.; Yi, J.-S. Classifying the Level of Bid Price Volatility Based on Machine Learning with Parameters from Bid Documents as Risk Factors. *Sustainability* **2021**, *13*, 3886. <https://doi.org/10.3390/su13073886>

Academic Editor: Sunkuk Kim

Received: 28 February 2021

Accepted: 25 March 2021

Published: 1 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** risk management; risk analysis; bid price volatility; uncertainty in bid documents; pre-bid clarification document; machine learning (ML), classification model; public project; sustainable project management

## 1. Introduction

Sustainability refers to the whole life cycle from siting to design, construction, operation, maintenance, renovation, and deconstruction [1,2]. Traditional research focused on the design and construction stages to maximize profits has gradually expanded to the entire life cycle of construction projects to realize sustainable development. Accordingly, many researchers have conducted valuable studies to minimize the impact on the environment by improving the energy performance of buildings and reducing waste. As a result, many studies on sustainability have developed remarkably around maintenance and subsequent steps. Recently, this trend has been further expanded to realize the results of many studies conducted so far [3]. Therefore, many studies have refocused on project management, which corresponds to the preceding stages in terms of sustainability [4]. The success or failure of a construction project starts from the initial stage of the project. More precisely, the feasibility at the bid and contract phases, stipulating plans for the future, enables the completion of a sustainable project.

Contracts for construction projects are created based on competitive bids. In general, the bidder who offers the lowest price is selected as the final winner. Therefore, determining the final price is crucial for bidders [5]. It is also difficult because the bid price affects the likelihood of gaining a satisfactory profit and winning the project [6]. The client provides a bid document to the bidders, who then examine it to estimate the bid price. Thus, the bid document plays an essential role in determining the bid price. If the content of the bid document is uncertain, the intention of the construction object may be ambiguous and cause mistakes during the construction phase, which may lead to construction rework and

claims [7,8]. Thus, bidders include the cost in the bid price to cover those risks. In general, the bid price can be expressed as follows (Equation (1)):

$$B_i = C_i(1 + M_i), \quad (1)$$

$C_i$  represents the construction cost and  $M_i$  is the markup (i.e., contingency), which means the risk cost due to uncertainty in the bid document [9]. In other words, if the risk cost increases owing to uncertainty in the bid documents, the bid price increases [10].

The uncertainty factor in bid documents that causes the risk cost must be determined and investigated. However, because reviewing all bid documents in a limited time frame is difficult [11], businesses often rely on their experience rather than quantitative uncertainty measurements [12–15]. Because uncertainty in bid documents is affected by complex factors that are difficult to measure quantitatively, most qualitative research studies have been conducted in academia. Therefore, determining the risk cost remains a tough challenge [16–20]. Bid prices that are not adequately set negatively affect bidders, clients, and users alike. Bidders take the risk to a severe degree, which not only does not yield the expected return but can lead to more serious financial difficulties. Simultaneously, such a bid price may increase the cost of completion due to frequent design changes during project execution, increasing the burden on the client. Eventually, the project quality completed by this process could be worse, causing great inconvenience to users.

Although reviewing all documents may be challenging, pre-bid clarification documents contain much more uncertain information than other documents. This document type includes inquiries and answers from bidders and clients about the uncertainty factors in the bid documents; this information can be used as an input parameter for a machine learning-based model to construct a bid price. This study aims to examine whether the uncertainty measured in pre-bid clarification documents affects the bid price. This uncertainty may change the mean value or variance of the bid price. In this paper, these two changes are operatively defined as “bid price volatility.”

In this study, a sample of data from the California Department of Transportation (Caltrans) in the US was used to see how uncertainty in the tender document changes bid price volatility. Analyzing the uncertainty of the bidding document is very difficult. In particular, the volume of bidding documents is enormous because of construction project size, making analysis difficult. Crucially, construction project data has an unstructured text format, making quantitative analysis even more difficult. However, the authors solved this problem using the pre-bid clarification document, which inquired about this uncertainty as a proxy, and used it with the bid summary. This study suggests that the uncertainty of the bid document affects the change in the bid price volatility. This allows bidders to execute the project at a reasonable price between earning a profit and winning the project. Further, this reasonable price can improve the project performance and realize the client’s satisfaction. More ultimately, it can extend sustainability in terms of the life cycle of the project.

## 2. Preliminary Research: Project Risk in Bid Phase and Uncertainty in Bid Documents

### 2.1. Definition of Project Risk in Bid Phase

The definition of risk depends on the subject and purpose in a field. Because risk is a concept defined to quantify the uncertainty regarding danger, it differs from the latter; it is defined as the “possibility of loss or injury” in dictionaries. In academia, the risk is more clearly defined as a factor or condition that can cause loss or injury owing to uncertainty; this definition focuses more on the possibility of risk rather than the risk itself, typically expressed in the following equation [21]:

$$\text{Risk Magnitude} = \text{Risk Severity} \times \text{Risk Likelihood}, \quad (2)$$

Risk magnitude is one of several attempts to measure risk [16]. It is a useful indicator for determining priorities among various risk factors using “risk” and “uncertainty” as

variables. However, this also acts as a limitation because relative comparisons between risk factors are possible, but absolute comparisons are impossible. Therefore, the quantitative relationship with the bidding price resulting from the risk in the actual bidding process is blurred. This suggests that a new indicator that can reflect the risk of bidding price is necessary for at least construction projects, and there have been many studies related to this. Abotaleb and El-Adaway [9] attempted to measure bidding risk as a percentage of markups. Besides the total construction cost, bidders present the total construction cost plus a specific rate as the bid price for pursuing profit while preparing for risks. In addition, a study was conducted to determine whether the successful bid price was a price that had more risk than necessary by using the contrast between the successful bid price and the average bid price [22]. Lee et al. [23] attempted to measure the bid risk by using an equation similar to the equation of Williams [22], but in which the engineer's estimate replaced the successful bid price. However, the previously suggested equations have limitations in that they are challenging to use in this study in the following aspects. First, it is a matter of the possibility of utilizing the markups. It is correct that the contingency is included in the price, but a third party such as researchers other than bidders cannot check from the bid history. This is because the contingency is included in one or more of several bid items of the bill of quantities (BOQ). Second, the successful bid price and the average bid price are values determined after the bidding date. It is difficult to predict similar projects' risks during the bid phase using these values.

In this study, the risk is defined as the quantitative uncertainty regarding risk, and the risk factor represents a factor that causes uncertainty regarding time, cost, and quality risk. We use two metrics that match our definition of risk in Section 4.1.2. The scope of this study covers construction projects, and the project risk corresponds to the uncertainty regarding risks that arise from the characteristics of the construction project. The bid varies according to the project delivery method in actual construction projects. In this study, the bid phase is considered as the period in which the construction bid is made with the design–bid–build method. Uncertainty in a bid document is one of the many project risk factors in the bid phase.

## 2.2. Project Risk Factors Affecting the Bid Price

Several researchers have studied project risk factors that affect the bid price. Construction projects can be classified into several types depending on the case, and any project type can include risk. Therefore, researchers have analyzed the risk without considering the project type. In this study, risk factors that affect the bid price are extracted from 13 reviews in the field of transportation (Table 1).

The above studies are of great significance in that they have substantially advanced the critical risk identification stage in risk management. Many studies have extracted common factors as considerations when bidding for projects. Existing studies have facilitated more detailed risk management by deriving or breaking down the priorities of risks to be considered when performing projects based on surveys of most experts. On the other hand, some studies have analyzed how the number of bidders affects the bid price and predicts the bid price through simulation using multiple variables instead of one variable. However, there is a limitation in not considering how the project risk is integrated into the project's initial bid price.

**Table 1.** Project Risk Factors Affecting Bid Price According to Previous Studies.

No.	Author	Year	Risk Factors
1	Ahmad & Minkarah [24]	1988	Degree of difficulty; type of project; client; location; design document quality; size of job; competition; contingency; duration
2	Shash [25]	1993	Number of competitors with bidding experience in such projects; client identity; contract conditions; project type; project size; bid method
3	Chua & Li [26]	2000	Degree of technological difficulty; identity of client/consultant; size of project; completeness of drawings and specifications; consultants' interpretation of the specification; project timescale and penalty for non-completion; time for bid preparation
4	Wanous et al. [27]	2000	Financial capability of client; relationship to and reputation of client; project size; availability of time for bidding; site clearance of obstructions; project duration
5	Han & Diekmann [28]	2001	Geography and climate conditions of country; document issues and contract conditions; project cost uncertainty; project schedule uncertainty
6	Egemen & Mohamed [18]	2007	Project size (total project value); current financial capability of client; type of work; technological difficulty of project
7	Zeng et al. [21]	2007	Communication; layout and space; site constraints; work scheduling; condition
8	Bageis & Fortune [29]	2009	Financial capacity of client; contract conditions; experience with similar projects; size of contract
9	Chan & Au [19]	2009	Poor employer's reputation to honor payment on time; very tight contract period; onerous contract conditions and rigid specifications
10	El-Mashaleh [17]	2012	Project type; project size (contract price); quality of bid documents (e.g., drawings, and specifications); client's reputation
11	Leśniak & Plebankiewicz [30]	2013	Type of work; contract documents; client's reputation; project value; project duration; criteria of bid selection; location of project; time window for bid preparation; degree of complexity of works
12	Ahiaga-Dagbui & Smith [31]	2014	Bid strategy; site information; ground conditions; type of soil; information; scope of project
13	Delaney [32]	2018	Project size; project type; level of competition; contract document quality; project-specific items; unforeseen conditions

### 2.3. Uncertainty in Bid Document as a Project Risk Factor

Uncertainty in bid documents is one of the most crucial risk factors. The bid phase is the first stage of a project contract. The bidder submits the bid price after reviewing the extensive bid documents, which contain information on the following three aspects: (1) the bid procedure (e.g., the announcement, guide, participation application, participation notice, and bid), (2) contract (e.g., the general and special conditions), and (3) construction (e.g., the drawings, specification, and pre-bid clarification document). Each bid document has a different scope and form. For example, the specification document contains a set of documented requirements and the drawings that present the building requirements. The special conditions are contract clauses that apply only to the project subject to the contract; they are created by changing, adding, or deleting existing content in the General Conditions section. In other words, the bid documents present standards and procedures regarding the design, construction method, materials, and inspection for the completion of the construction object; thus, the bid documents constitute the basis for calculating the bid price. In addition, because bid documents are contract documents, they are the basis for judgment when legal problems arise in the future. Cost overrun can occur if the bidder fails to review the bid document's risk factors in advance [23]. Therefore, analyzing the uncertainty in bid documents is crucial.

Discrepancies, errors, and omissions cause uncertainty in bid documents; these are the leading causes of legal adjustment, arbitration, or litigation regarding the project costs. According to Tanaka [33], 74.4% of construction-related claims in the United States are due to uncertainty in bid documents; Erdis and Ozdemir [34] studied the dispute between a client and bidder, arguing that uncertain expressions in a bid document could lead to construction disputes.

Public projects in the US include a pre-bid clarification procedure that can resolve all uncertainty in bid documents before the bid. If bidders find uncertainty in a bid document during the quotation, they can contact the client, who must respond within the deadline. Relieving all uncertainty in bid documents through this approach helps bidders present the correct bid price [35]. New York State in the US emphasized that the pre-bid clarification is a significant procedure for the client and bidder [36]. The former can calculate the project cost more accurately with less uncertainty. Pre-bid clarification is an institutional method that helps present accurate project costs and prevents possible future design changes, extensions, additional construction costs, and disputes [37].

#### 2.4. Pre-Bid Clarification Document as a Proxy for Uncertainty in Bid Documents

Uncertainty in bid documents includes (1) unclear communication caused by discrepancies, errors, or the omission of information or (2) unclear requirements regarding the project object. In general, the bid process for a construction project involves many bid documents [23]. Each document may independently contain risk factors; besides this, they may interact with each other, creating risk factors. Hence, all bid documents must be carefully reviewed to determine the uncertainty level. However, it is complicated for bidders to identify all hidden risks within a short bid preparation time [16–20,23]. Therefore, in the actual field and academia, the uncertainty of bid documents has been considered a complex problem to solve [11,38] and risk beyond control [28].

The uncertainty that arises in the pre-bid clarification procedure is caused by factors, which occur in all the bid documents that the bidders read. These documents are incorporated into the pre-bid clarification document, which can serve as a proxy variable that gauges the entire bid document's uncertainty. For example, Daoud and Allouche [39] analyzed pre-bid clarification documents to examine which uncertainty factors occur in the bid documents of construction projects.

#### 2.5. Hypothesis Development

From the literature review, there is a widely believed proposition: uncertain things during the bid phase affect bid price on the theoretical plane. However, the problem is that the factors classified as risks are mixed with what can be measured and what is not, what can be controlled and what is not possible, making quantitative analysis impossible. For this reason, when practitioners calculate prices, these uncertainties are guessed and reflected in prices without a factual basis. We made the following two assumptions to establish the hypothesis: (1) As the uncertainty increases, the bidders will reflect this in their prices, causing an increase in the overall average bid price. (2) The greater the uncertainty, the more significant the difference in prices offered by bidders will also increase, resulting in an increase in the range of bidding prices formed. Under these assumptions, we set up the following two hypotheses on the empirical plane.

**Hypothesis 1 (H<sub>1</sub>).** *Factors derived from bid summary and pre-bid clarification document affect  $F_1(x)$ , representing the volatility of bid price.*

**Hypothesis 2 (H<sub>2</sub>).** *Factors derived from bid summary and pre-bid clarification document affect  $F_2(x)$ , representing the volatility of bid price.*

In H<sub>1</sub> and H<sub>2</sub>, the factors consist of seven independent parameters obtained from the bid summary and pre-bid clarification document. Then,  $F_1(x)$  of H<sub>1</sub> becomes Bid Average Risk (Equation (3)), and  $F_2(x)$  of H<sub>2</sub> becomes Bid Range Risk (Equation (4)) discussed in Section 4.

#### 2.6. Research Gaps and Research Questions

According to the Project Management Body of Knowledge [40], risk management research is based on (1) risk identification, (2) risk assessment, and (3) risk plan and control. The risk assessment, which is a leading step in risk planning and control, quantifies the



potential impact of these uncertain factors [11]. However, many variables to be considered and interrelated make the analysis in the actual field and research studies difficult [14,24].

The fundamental reason is risk identification (which is a leading step); the general approach is to subdivide all project risk factors into controllable units based on specific criteria. Analyzing segmented risks can reduce uncertainty in the bid phase [37]. However, this approach has not been both quantitatively and qualitatively studied for risk factors in bid documents [23] because they contain vast amounts of information and differ in content depending on the project. Therefore, uncertainty in bid documents has been classified as an uncontrollable risk [28]. As mentioned in Section 2.2, researchers have only progressed to Level 1 by suggesting that uncertainty in bid documents is a risk factor; there are no sufficient specific studies on Level 2 [23]. In other words, published management studies have mainly focused on high-level risk factors and surveys with expert groups [11,38]. However, these data [41] only serve as references for determining the bid price in the bid phase.

The following factors must be investigated: first, regarding the social background, a reasonable bid price is crucial for establishing a reasonable project budget for the bidder and client [20]. This requires a decision support tool that can be used by practitioners who encounter difficulties in the bid price prediction. In research, a more quantitative study based on actual bid data is required to assess whether uncertainty in bid documents affects bid prices. This study aims to meet both academic and practical needs by analyzing whether uncertainty in bid documents affects the bid price.

Uncertainty in a bid document is expected to have the following effects on the bid price. First, each bidder will represent this risk factor in his/her bid price, thereby increasing the project's overall bid price (i.e., the bid price average). Besides this, the other bidders represent this risk factor in prices, which increases the range of the established bid price bands. In this study, these two X are defined as "bid average risk" and "bid range risk," respectively. Further, the bid price volatility comprises the two types of bid price fluctuations due to uncertainty in a bid document (i.e., the increase in the bid price average and range). This study aims to provide answers to the following two questions:

- *Research Question 1:* Does the uncertainty in the pre-bid clarification document increase the bid average risk and bid range risk?
- *Research Question 2:* Does the parameter of uncertainty in bid documents improve the classification of the bid price volatility?

The results of this study are two types of bid price volatility level classification models:

- *Model 1:* level classification model without uncertainty in the bid documents;
- *Model 2:* level classification model with uncertainty in the bid documents.

In this study, the performance of Model 2 is evaluated to support decision-making about bid prices.

### 3. Materials and Methods: Modeling Approach

Regarding risk management, this study on assessment is different from risk plan and control, which supports decision-making on participation in the bid phase. This is because the decision-making process regarding bid prices of bidders who have already decided to participate is supported in this study. In general, risk assessment studies can be classified into studies of  $B_i$  and studies of  $M_i$  (Equation (1)). Because it is difficult to collect and analyze sufficient data, mainly  $M_i$  has been studied; in this study,  $B_i$  is empirically evaluated based on the actual bid results. In addition, in other published studies, the uncertainty factors of bid documents were analyzed with a proxy (i.e., a pre-bid clarification document).

Uncertainty factors in bid documents are natural phenomena because construction projects are typically one-off projects. Considering the toxin clause that partially exists in the special conditions, an uncertainty factor in bid documents is problematic because the artificial content clearly defines who should be responsible in certain circumstances.

Therefore, the uncertainty factors in bid documents considered in this study are limited to those that occur naturally because of specific characteristics of the construction industry.

### 3.1. Materials

The data from the bid results regarding the project risk factors discussed in Section 2.2 are the variables of interest. To study their effects, the data in which the influence of other factors can be minimized should be analyzed. The public construction project of Caltrans meets this purpose because of the following reasons: first, the uncertainty in the bid documents can be analyzed. Because Caltrans includes a pre-bid clarification process in the bid phase, the pre-bid clarification document is publicly available. Second, the quantity and quality of available project data are sufficient. Caltrans invests approximately \$1.7 billion per year in approximately 450 projects, which is the largest of the 50 US states. The thousands of standardized project datasets of Caltrans have led to large amounts of high-quality data and excellent project management capabilities based on experience. Third, the absence of special conditions reduces influences from other than the variables of interest. Standard contracts used worldwide include the FIDIC (Fédération Internationale Des Ingénieurs-Conseils), JCT (Joint Contracts Tribunal), NEC (New Engineering Contract), and AIA (American Institute of Architects), mainly applied to private projects. By contrast, Caltrans uses federal-aid construction contracts (FHWA-1273) for public projects. In this case, only general conditions without special conditions (unlike private projects) are applied, which means that the projects are relatively standardized. The bid document of a standardized project reduces the influences of numerous external factors; it is considered suitable for observing the effect of uncertainty in bid documents on the bid price because of the absence of special conditions.

Caltrans has published all the bid results online since 2004 (they provide all bid documents, bid summaries, and important information). However, the online services for pre-bid clarification documents have been operated since 2011. The number of projects since the access date was 3584 during 2011–2018. In total, 3578 datasets were collected (six cases were excluded because they could not be accessed owing to system errors).

### 3.2. Methods

*Pre-Data Analysis (Data Preprocessing Based on Bid Summary and Pre-Bid Clarification Document):* in this step, information that can be obtained from the bid result is preprocessed into input parameters (IPs) and output parameters (OPs). Caltrans has published a bid summary containing the critical details of the bid results. In this study, the data related to project risk factors affecting the bid price (which was discussed in Section 2.2) are extracted from the bid summary (B. S.) and pre-bid clarification document (P. C. D.). Subsequently, the final dataset is constructed from the raw data.

*Data Analysis (Two Classification Models of Bid Price Volatility Based on Machine Learning):* Methods of analyzing data can be classified into several categories depending on the purpose of the study and the characteristics of the data. When analyzing data that is large and composed of various factors, techniques such as data mining through machine learning (ML) are mainly used, and the data mining method is actively applied in recent risk analysis studies [42]. Such data mining can be classified mainly into a prediction technique that derives a regression equation, such as statistical analysis, and a classification technique that determines the category of data. Therefore, this study uses a machine learning-based data mining classification technique as a data analysis technique to classify the level of bidding risk with a large amount of data composed of various variables. In this study, machine learning can be used to classify the OPs in data consisting of multiple IPs. In this study, the class of the OP is designated such that the model algorithm learns to classify the levels of bid price volatility with MATLAB. To evaluate the model performance during training and validate it through validation tests in a post-data analysis, the pre-data analysis's final dataset is classified into training and validation data (for the training validation and validation test, respectively). As a result, Model 1 (which does not include the uncertainty

in the bid documents in the IPs) and Model 2 (which includes the uncertainty in the bid documents in the IPs) are generated.

*Post-Data Analysis (Validation):* Models 1 and 2 are tested in a validation test to determine whether the models created in the data analysis step show similar performance characteristics for new data other than the data used for training. The test results are presented in a confusion matrix, which is analyzed and discussed.

#### 4. Model for Classification of Level of Bid Price Volatility

##### 4.1. Pre-Data Analysis: Preprocessing of Data from Bid Summary and Pre-Bid Clarification Document

###### 4.1.1. Input Parameters

Because of the wide range of types and the number of bid documents, Caltrans's bid summary contains significant information about the bid. The pre-bid clarification document contains uncertain details of risk on the bid documents. Based on the factors discussed in Section 2.2, highly relevant parameters to the bid price are extracted from these two documents (Table 2).

Table 2. Metadata of Input Parameters.

Risk Type	No.	Input Parameter	Relevant Risk Factors in Previous Studies	Origin	Type	Unit	Coding
Time	IP-1	Working Days	Duration (No. 1, 2, 4, 7, & 11); project schedule uncertainty (No. 5)	B. S.	Num.	Days	47–1530
Time	IP-2	Project Location	Location of the project (No. 1 & 12)	B. S.	Categ.	County	1–12
Cost	IP-3	Engineer's Estimate	Size of project (No. 1 <sup>1</sup> , 2, 3, 4, 7, 10, 12, & 13); size of contract (No. 8); value of project (No. 11)	B. S. <sup>2</sup>	Num. <sub>4</sub>	\$	10,000,000–280,000,000
Cost	IP-4	Bid Preparation Days	Available time for bid preparation (No. 3, 4, 9, & 12)	B. S.	Num.	Day	18–237
Cost	IP-5	Number of Bidders	Competition (No. 1 & 13); the number of competitors with bidding experience in such projects (No. 2)	B. S.	Num.	#of Bidders	2–12
Quality	IP-6	Project Type	Project type (No. 1, 2, 10, 11, & 13)	B. S.	Categ. <sub>5</sub>	Road or Bridge	1–2
Quality	IP-7	Uncertainty in Bid Documents	Document quality (No. 1, 3, 5, 8, 9, 10, 11 & 13); communication (No. 7)	P. C. D. <sup>3</sup>	Num.	# of U. B. I. <sup>6</sup>	2–59

<sup>1</sup> Numbers in "Relevant Risk Factors . . ." match numbers in Table 1; <sup>2</sup> B. S.: bid summary; <sup>3</sup> P. C. D.: pre-bid clarification document;

<sup>4</sup> Num.: numeric. <sup>5</sup> Categ.: categorical; <sup>6</sup> U. B. I.: unsolved bid inquiries.

Meanwhile, the information extracted from the bid summary requires preprocessing to be used as a model parameter. Further, text data must be standardized through nominalization, and numeric data must be filtered based on a chosen range such that the outliers do not affect the model performance. The pre-bid clarification document contains the inquiries and answers for the project (Table 3).

Most inquiries aim to accurately estimate the bid prices by resolving uncertainty; because some inquiries do not, they must be preprocessed to include only those related to uncertainty. If these inquiries can be resolved with appropriate answers, they are excluded. The following describes the seven input parameters presented in Table 2 in detail for each risk type.



**Table 3.** Examples of Inquiry and Answer in Pre-Bid Clarification Document.

No.	Inquiry	Answer
1	(Inquiry #6) Please check the shoulder backing quantity. Please check the tack coat quantity. Please check the HMA quantity.	(Response #1): Unsolved Bid per current contract documents.
2	(Inquiry #36) Bolted Connections and Faying Surfaces: 1. Are all A325 and A490 HS bolted connections considered slip critical or bearing connections? If so, provide reference in plans and specs.	(Response #4): Solved 1. Refer to Addendum No 4, dated 13 February 2018. Section 55–1.02E(6)(c) of the Special Provisions has been modified to define the HS bolted joint type.

### Time Risk

*IP-1 (Working Days):* IP-1 represents the period of completion of the project required by the client. If IP-1 is relatively short considering the size of the project, the bid price may increase owing to required rush or night work. The projects considered in this study have IP-1 values between 47 and 1530.

*IP-2 (Project Location):* the IP-2 of the raw data is an address close to the construction site. IP-2 is related to the local price index, affecting the bid price. In the US, the price index is generally determined at the state level; however, differences in prices can occur within a single state. Because California is a large state in the US, Caltrans divides its administration into 12 districts separate from their counties. Accordingly, in this study, IP-2 is coded as 1–12 according to the district.

### Cost Risk

*IP-3 (Engineer's Estimate):* the project cost to which bidders can refer for the bid price is IP-3 at the time of the announcement; because the raw data have a too wide IP-1 distribution, the range must be adjusted. In this study, projects in the range of \$10,000,000–\$280,000,000 are used in the model.

*IP-4 (Bid Preparation Days):* IP-4 is when bidders have to review the bid document (including the uncertainties); thus, this time affects the accuracy of the bid price. IP-4 is calculated as the period from the bid announcement date to the bid opening date extracted from the raw data (values between 18 and 237).

*IP-5 (Number of Bidders):* to use IP-5 as an input parameter, it must be checked whether the information is known before the bid opening. Researchers have argued that the variable IP-5 influences the bid price [24,25,32]. When it increases, the bidders deliberately lower the bid price to win the project [43,44]. Thus, the bidders are aware of the number of competitors in advance; Christodoulou [43] studied the optimal  $M_i$  (Equation (1)) based on this premise. Therefore, IP-5 is included as an input parameter with values from 2 to 12.

### Quality Risk

*IP-6 (Project Type):* IP-6 can be mainly classified into roads (e.g., highways, freeways, or roadways) and bridges; the numbers “1” and “2” represent a road and bridge, respectively.

*IP-7 (Uncertainty of Bid Documents):* 3578 raw datasets are screened through Section 4.1.1, which result in 269 final datasets with 6682 bid inquiries. As mentioned in Section 2.4, there are two uncertainty factors in bid documents: unclear communication (BI. 1–3) and unclear requirements (BI. 4–5). Uncertain communication includes discrepancies, errors, and the omission of information in the bid document, each of which has overlapping meanings. For example, omission means that necessary information is missing owing to an error; thus, it can be interpreted as an error itself. Therefore, each term is clearly classified according to the mutually exclusive and collectively exhaustive principle. When certain identical information in various bidding documents causes conflicts, the case corresponds to case BI. 1: discrepancy. The case in which being inquired by an error of single information itself is categorized as case BI. 2: error. Uncertainty due to the lack of specific information

is classified as case BI. 3: omission. Furthermore, uncertain requirements are classified into two types that ask for insufficient but non-essential information (BI. 4: insufficient information) or accept alternatives to the existing guidelines (BI. 5: alternative information). That is, only inquiries corresponding to BI. 1-5 among the pre-bid clarification document content are regarded as uncertainty factors. Through this process, 52 bid inquiries are excluded. After excluding the questions, the uncertainty of which has been resolved with appropriate answers (4336), the number of unsolved bid inquiries is 1994 with values between 2 and 59 for each project (i.e., the IP-7).

All coded IPs are used to train the model in the normalized form.

#### 4.1.2. Output Parameters

As stated in Section 2.6, the output parameters of the models are the bid average risk (OP-1) and bid range risk (OP-2), which are based on the bid price of the raw data. The bid average risk is the ratio of the average price to the engineer's estimate (Equation (3)):

$$\text{Bid Average Risk} = \frac{\text{Average Bid Price}}{\text{Engineer's Estimate}} = F_1(x). \quad (3)$$

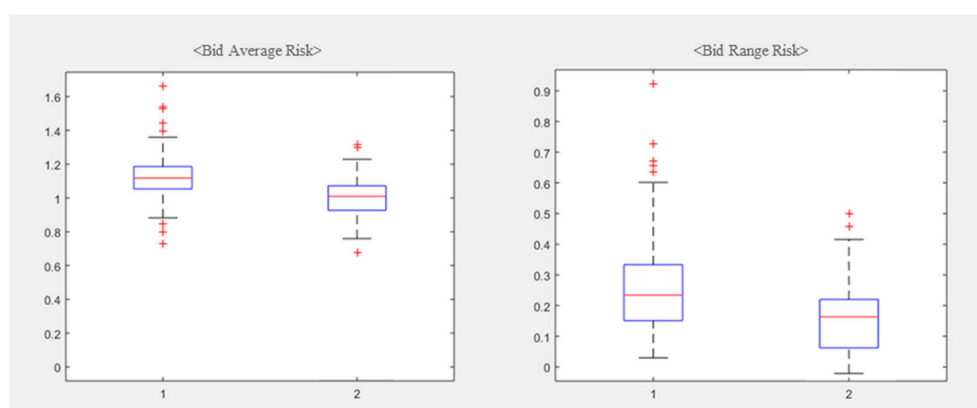
For example, if the engineer's estimates of projects A and B are \$10 billion and the respective average bid prices are \$10 billion and \$13 billion, the bid average risks are 1.0 and 1.3, respectively. Thus, it can be assumed that the bidders expect a higher risk for project B. Moreover, the bid range risk (OP-2) refers to the difference between the maximal and minimal bid price concerning the engineer's estimate (Equation (4)):

$$\text{Bid Range Risk} = \frac{\text{Max. Bid Price} - \text{Min. Bid Price}}{\text{Engineer's Estimate}} = F_2(x). \quad (4)$$

For projects A and B with estimates of \$10 billion and \$1 billion, respectively, the differences between the maximal and minimal bid prices would be identical (\$2 billion); however, the differences between the uncertainties of the two projects cannot be considered identical: project B is riskier than A. Finally, as mentioned in Section 2, the bid average risk and bid range risk defined in this section act as  $F_1(x)$  of  $H_1$  and  $F_2(x)$  of  $H_2$ .

#### 4.1.3. Impact of IP-7 on Bid Price Volatility

The raw data of 3578 projects are preprocessed to create the final dataset of 269 projects. To answer research question 1 in Section 2.6, the final dataset should be classified into groups with and without uncertainty factors. In this study, projects with IP-7 values between 2 and 59 are considered a group with uncertainty factors, and projects with IP-7 values between 0 and 1 are considered a group without uncertainty factors. The OP-1 and OP-2 in each group are presented in Figure 1 and Table 4.



**Figure 1.** Comparison between OP-1 and OP-2 of Projects with IP-7 Values of (1) 2–59 and (2) 0–1.

**Table 4.** Comparison between Data Statistics of OP-1 and OP-2 of Projects 1 and 2.

No.	Output Parameter	Data Statistics							
		Project 1					Project 2		
		Med.	Max.	Min.	Mean	Med.	Max.	Min.	Mean
OP-1	Bid Average Risk	1.1172	1.6612	0.7295	1.1303	0.9759	1.2787	0.6482	0.8653
OP-2	Bid Range Risk	0.2339	0.9222	0.0297	0.2684	0.2004	0.5338	0.0174	0.1761

As shown in Figure 1 and Table 4, projects with uncertainty (Project 1) score higher values for the bid average risk and bid range risk than projects without uncertainty (Project 2). In other words, uncertainty in bid documents increases the bid price volatility.

#### 4.2. Data Analysis: Two Classification Models Based on Machine Learning for Bid Price Volatility

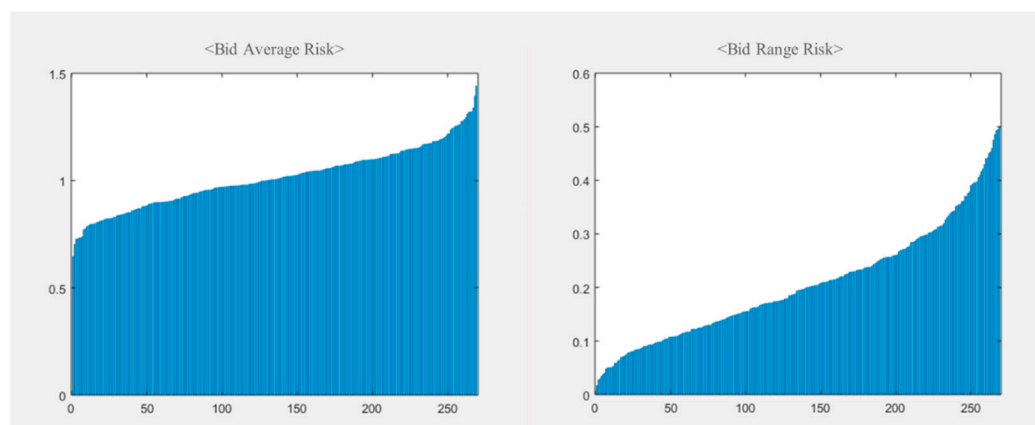
##### 4.2.1. Design

In the data analysis stage, the final data in Section 4.1 is used to implement a model for classifying the level of volatility in bid prices. This study presupposes that the uncertainty of the bid document is related to the bid price and ultimately tries to improve the accuracy of the bidding price volatility level classification model by adding the variable of the uncertainty of the bid document to the existing bid-related variables.

SPSS, MATLAB, R, and Python are mainly used for machine learning-based data mining classification. In this study, data analysis and model development were performed using Mathworks' MATLAB software as a data analysis tool.

Because this model predicts the classes of OP-1 and 2 with input parameters, the models are trained through supervised learning after the class designation. A suitable number of classes is significant because too many classes decrease the reliability of the prediction results; too few classes make the interpretation of the results difficult. In this study, four levels according to the bid price volatility distribution are considered.

When the boundary between the classes is set, a natural breakpoint is preferable; the breakpoint is the point at which the distribution of the OP values suddenly breaks. If there is no natural breakpoint, a boundary should be set such that the data of each class is evenly distributed to ensure reliability. Figure 2 shows the distributions of OP-1 and OP-2 of the 269 final projects; the parameters do not have natural breakpoints. The set boundaries between the classes of the models are presented in Table 5. The OP-1 and OP-2 have a total of 4 classes: OP-1 is a “-” class that is much smaller than 0, a slightly smaller class is “-,” a slightly larger class is “+,” and a much larger class is “++.” On the other hand, classes of OP-2 were named with the symbols “+,” “++,” “+++,” and “++++” in the order of close to 0.

**Figure 2.** Distribution of Bid Price Volatility of 269 Projects.

**Table 5.** Class Designation for Output Parameters.

No.	Output Parameter	Class Designation				
OP-1	Bid Average Risk	Class	–	–	+	++
		Range	(min., 0.9)	(0.9, 1.0)	(1.0, 1.1)	(1.1, max.)
		Ratio	22%	26%	27%	25%
OP-2	Bid Range Risk	Class	+	++	+++	++++
		Range	(min., 0.27)	(0.2, 0.27)	(0.27, 0.31)	(0.31, max.)
		Ratio	25%	23%	24%	28%

Not all final datasets are used to train the model. The remaining datasets are used to check whether the classification model performs consistently for new data (typically 30% of the total data). Half these data is used for the validation during training, and the rest is applied in the validation tests (Section 4.3). Accordingly, 269 datasets are allocated to 40 for the training validation, 40 for the validation test, and 189 for the model training with machine learning. There are several machine learning algorithms for training classification models; in this study, neural net (NN), which shows good performance, is used. In addition, three algorithms are combined to evaluate the results derived with neural net. In total, four classification algorithms are used for the models: NN, decision tree (Tree), support vector machine (SVM), and K-nearest neighbor (KNN).

#### 4.2.2. Results

In the bid price volatility level classification based on the model design in Section 4.2.1, Model 1 is trained without IP-7, whereas Model 2 is trained with IP-7. The performance of the model is expressed as the accuracy (i.e., the number of correct classifications compared to the total number of classifications):

$$Accuracy (\%) = \frac{\text{Number of Correct Classifications}}{\text{Total Number of Classifications}} \quad (5)$$

In this section, the accuracies of Model 1 trained without IP-7 and Model 2 trained with IP-7 are compared to answer research question 2 in Section 2.6 (Table 6).

**Table 6.** Accuracies of Classification Models.

Algorithm	Accuracy (%)			
	Model 1		Model 2	
	Bid Average Risk	Bid Range Risk	Bid Average Risk	Bid Range Risk
NN	37.5	42.5	63.9	65.8
Tree	31.5	34.3	61.7	60.6
SVM	34.6	40.5	59.5	60.5
KNN	32.8	36.1	58.4	56.1
Average	34.1	38.4	60.9	60.8

According to Table 6, Model 2 performs better than Model 1 for all algorithms, including the NN. Second, the classification Models 1 and 2 result in higher accuracies for OP-2 than for OP-1 in most cases. Third, the NN exhibits the best performance in Models 1 and 2.

Model 1 results in accuracies with the order of SVM > KNN > Tree, and Model 2 results in accuracies with the order of Tree > SVM > KNN. Thus, IP-7 (a parameter of the uncertainty in bid documents) improves the performance of the model that classifies the level of the bid price volatility. Regarding the NN algorithm classifier, the performance scores of Model 1 (without IP-7) are 37.5% (OP-1) and 42.5% (OP-2), whereas those of Model 2 (with IP-7) are 63.9% (OP-1) and 65.8% (OP-2); these results are 26.4% and 23.3% higher than those of Model 1, respectively. This trend is identical for the averages of all

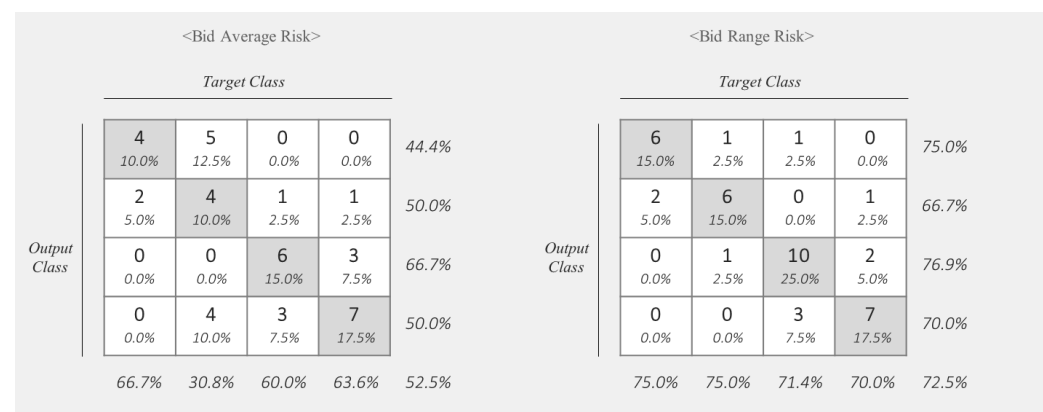
algorithms: the accuracies of Model 1 are 34.1% (OP-1) and 38.4% (OP-2), while those of Model 2 are 60.9% (OP-1) and 60.8 % (OP-2).

#### 4.3. Post-Data Analysis: Validation

In the validation test, the accuracy of a classification model is evaluated with data that have not been used for training. In this section, the training performance of Model 2 with IP-7 is validated based on the NN. There are various approaches for validating a machine learning model (e.g., the N-fold cross-validation, bootstrap, and sliding window methods); N-fold cross-validation is the most used approach. The results of the validation test are presented in a confusion matrix, which enables the determination of the accuracy of the model and true positive rate (TPR, which represents the recall and sensitivity):

$$\text{True Positive Rate (TPR)} = \frac{TP (\text{True Positive})}{TP (\text{True Positive}) + FN (\text{False Negative})} \quad (6)$$

where “True Positive” (TP) refers to samples in which the positive cases are correctly classified, whereas “False Negative” (FN) refers to samples in which the negative cases are incorrectly classified. In other words, the TPR is the ratio of correct samples to total samples classified by the model. Figure 3 presents the validation results of Model 2 trained with the NN for the classification of the bid price volatility.



**Figure 3.** Confusion Matrices of Model 2: Bid Average Risk and Bid Range Risk.

The validation accuracies of Model 2 for OP-1 and OP-2 are 52.5% and 72.5%, respectively; thus, the model correctly classifies the bid price volatility levels of 21 and 29 out of 40 projects. First, the TPR of Model 2 for OP-1 is highest in the “+” class (66.7%) and lowest in the “-” class (44.4%). However, according to the confusion matrix, incorrectly classified samples are classified into groups of similar rather than entirely different classes. This is because successive values are cut by the artificial boundaries of the class designation. Moreover, the accuracy of the model for the test data (52.5%) is slightly lower than the accuracy of the model for the training data (69.3%). Second, the TPR of Model 2 for OP-2 is highest in the “+++” class (76.9%) and lowest in the “++” class (66.7%). For OP-1, the model classifies most of the samples into same or similar classes. In addition, the performance of the model for the validation test data is 72.5%, which is better than the accuracy of the model for the training data (64.6%).

## 5. Discussion

From a background in Introduction and Research Gaps, a decision support tool for bid price budgeting for the actual field and research studies was created based on quantitative data analyses with project-based bid results from other studies. Based on these requirements, two research questions were established, and Caltrans’s bid results were used to select related parameters based on previous studies. Based on the 269 final project datasets obtained after preprocessing, the uncertainty level in bid documents was



quantified as the number of inquiries corresponding to BI. 1-5 in the pre-bid clarification document that have not been resolved.

To answer research question 1, the 269 projects were classified into two groups: a group with IP-7 values of 0–1 and another group with IP-7 values of 2–59. The group with uncertainty in the bid documents generally had a higher bid average risk and bid range risk than the group without uncertainty. Thus, uncertainty in bid documents increases the bid price volatility. It is noteworthy that the project with an IP-7 value of 1 was also considered a project without uncertainty in this study because the considered uncertainty level depends on the stakeholders' opinion. For example, for projects with IP-7 values of 1, some bidders may believe that there is little uncertainty, whereas others may believe that the uncertainty level is relatively high. Therefore, the boundary of uncertainty can be expressed as follows: {project without uncertainty: IP-7 value of  $x \mid x = 0$ }  $\cup$  {project with uncertainty: IP-7 value of  $x \mid x > 0$ } for risk-averse bidders and as {project without uncertainty: IP-7 value of  $x \mid 0 \leq x \leq n$ , where  $n = 1, \dots, 59$ }  $\cup$  {project with uncertainty: IP-7 value of  $x \mid x > n$ } for risk-takers. In other words, for projects with IP-7 values of  $n$ , the level of perceived uncertainty depends on the stakeholder; the case  $n = 1$  was presented as an example in this study.

Furthermore, Models 1 and 2 were trained with an NN to examine whether IP-7 affects the bid price volatility classification. The accuracies of Model 1 for the bid average risk and bid range risk were 37.5% and 42.5%, respectively; those of Model 2 were 63.9% and 65.8%, respectively.

The accuracies of Model 1 exceeded 25%, which corresponds to the mathematical probability when one of four classes is randomly selected; nevertheless, the accuracies were too low for decision making. This means that parameters that are not included in the model have a more substantial influence on the bid price. By contrast, the accuracies of Model 2 were better, and similar phenomena were observed for the three other algorithms and NN. In other words, the influence of IP-7 is relatively strong compared to those of the other parameters included in the model. However, the fact that the accuracy of the models has stopped at the current level proves that there are still other parameters that are not included in the models and that significantly affect the bid price. However, this effect can be attributed to the nature of IP-7 itself; IP-7 represents the number of unsolved bid inquiries; this surrogate endpoint is set because it is impossible to cost the situation mentioned in each inquiry. For example, two inquiries might be worth \$100 and \$10,000, respectively; however, they are treated equally as a value of 1 in the models. This may be why the models score higher accuracies for the bid range risk than for the bid average risk. From a superficial point of view, this implies that the model's application may be limited due to its accuracy. However, the more time it takes to consider accurately the cost of each inquiry, the less time it takes to identify more risks, which in turn increases the uncertainty of the bidding document. In this respect, obtaining such a relatively high accuracy only with the number of unsolved bid inquiries is a good sign.

Moreover, the results were validated with test data, and the answer to research question 2 was found. Bidders should determine the bid price by integrating project-related information. Before this, the bidders should add the cost of project uncertainty in the markups along with  $C_i$  (Equation (1)), which is calculated based on construction budgeting. Because "uncertainty" literally means "lack of knowledge," there is no raw data for calculating it; thus, Model 2 can be applied. For example, if the project's bid average risk is classified as "–" the average of the bid price is likely to be much lower than the engineers' estimate. If the currently calculated  $C_i$  is high,  $M_i$  must be lowered, or  $C_i$  must be adjusted to win the bid. If the results are below the lower bound of the expected profit, the bidder may stop participating in the project. Likewise, if the bid range risk is classified as "+++", many bidders may present prices that deviate much from the engineers' estimate; thus, the bidder may use this strategy to adjust  $C_i$  and  $M_i$ .

Since there are too many uncertainties affecting the bid price of a construction project and many of them are risks, researchers have performed many valuable studies, and

practitioners have been through trial and error. However, when a specific point in time and space is determined, many risks are eventually settled. For example, a city's price index itself fluctuates over time, but finally, the city's price index is determined as one at the time and place of the project. So, for bidders, one dilemma between winning bids and project profits is, therefore, whether they should call more or less than expected. These fine-grained adjustments are ultimately determined by project-intrinsic-risks that have not been finalized. The bidding document is a kind of contract document, and all the explicit contents contained in it are directly related to the price. However, there is a problem that uncertain content cannot be dealt with one-to-one with the price. Moreover, since the contents of the bidding document are all different for each project, it was impossible to solve this with the existing linear method. On the other hand, the authors measured the uncertainty of each bid document, not the content of the bid document, and analyzed the risk through the NN algorithm. This study provided an answer to how much the uncertainty in the contents of bid documents increases or decreases the already expected price.

Nevertheless, the method presented in this study has several limitations, which must be considered in future research studies. First, the data are limited. Caltrans's project data were used to minimize the influence of external factors; however, owing to the characteristics of standardized data, the impact of uncertainty in bid documents (which is one of the internal factors) may be relatively weak. In this respect, it might be difficult to generalize the results of this study. Another limitation is related to the preprocessing of IP-7. Because the pre-bid clarification document contains unstructured text data, which is challenging to be computationally processed, there is no automated method for quantifying it; thus, the authors manually analyzed 6682 datasets. Because this process can introduce human error, the authors mutually verified the results. Third, there are obvious limitations in that this paper does not address all of these factors and only deals with the risk of uncertainty in the bid document. The risk of fluctuations in the material and workforce market is a significant factor affecting the bid amount and must be considered in the bidding stage. The bidders scrutinize these through market research. Meanwhile, the risk of fluctuations in the material and labor market varies with time and location, which means they are variables. If they are determined, the risk of volatility could be determined, too. Since the bidders bid simultaneously for work performed at the same location, these costs will reach some agreed value. However, this cannot be confirmed as a single value, so it remains a risk of the bid price. Future studies have to further consider the remaining major fluctuation risks and use them as parameters. In that case, significant improvements are expected to improve model accuracy.

## 6. Conclusions

In this study, a classification model for the bid price volatility level was developed by analyzing the relationship between the uncertainty in bid documents and bid price based on bid data. The model results in this study reveal that uncertainty in bid documents is causing bid prices to rise or fall more than necessary. Therefore, it is essential to conduct a thorough review of items causing uncertainty before bidding. We present these items as discrepancy, error, omission, insufficient information, and alternative information. Besides this, the inclusion of a pre-bid clarification process allows the price of the project to converge more appropriately if the remaining uncertainties are eliminated at the time of bidding. This study has the following contributions: the first step in quantifying the uncertainty in bid documents. The results of published qualitative studies of risk identification were evaluated with bid result data in this quantitative study. The model proposed in this paper enables risk management at a lower level: in the new approach, the uncertainty in bid documents considered to have an uncontrollable risk is analyzed with a pre-bid clarification document. Through this, the theoretical gaps are closed. The results of this study can help bidders to determine the bid price. According to the results, the pre-bid clarification in the bid phase is an essential process because the resolution of uncertainty in the bid documents can reduce the bid average risk and bid range risk. Accordingly, it is

expected that the bid price of the project, the risk of which has been resolved during the pre-bid clarification process, will converge to a more acceptable price. The construction objectives created with this price will improve bidders' profitability and meet the client's expectations, which ensures a successful construction project. The final contribution of this study is that the concept of sustainability has been further expanded in construction projects. In construction, the idea of sustainability is meaningful in that it extends the project management, which focused on the design and construction stages, to the entire life cycle. Until now, researchers have carried out valuable studies related to sustainable energy use, noting the importance of maintenance and operation phase after completion. However, the early stages of the project are also of great importance. This study tried to ensure the success and sustainability of the project through research on a reasonable project price. In the future, the authors will establish a method that comprehensively analyzes the uncertainty in unstructured text data from public projects of various institutions that provide pre-bid clarification documents for the automatic extraction of the IP-7 content. Further, the authors will combine this with the results of this study to establish a more general and highly accurate risk classification model.

**Author Contributions:** Conceptualization, Y.J. and J.-S.Y.; methodology, Y.J. and J.S.; software, Y.J.; validation, Y.J., J.-S.Y. and J.S.; formal analysis, Y.J.; investigation, Y.J.; resources, Y.J.; data curation, Y.J. and J.-S.Y.; writing—original draft preparation, Y.J., J.-S.Y. and J.S.; writing—review and editing, Y.J., J.-S.Y. and J.S.; visualization, Y.J.; supervision, J.-S.Y. and J.S.; project administration, J.-S.Y. and J.S.; funding acquisition, J.-S.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21ORPS-B158109-02). This study is also supported by the Ewha Womans University Scholarship of 2019.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Caltrans's raw data (2011–2018) used in this study can be accessed through the following link: <http://ppmoe.dot.ca.gov/des/oe/project-bucket.php> (accessed on 27 March 2021).

**Acknowledgments:** The authors would like to thank Lee, Assistant Professor of Department of Civil & Environmental Engineering and Construction of University of Nevada, Las Vegas, for helpful advice on various technical issues examined in this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Tomczak, M.; Jaśkowski, P. New approach to improve general contractor Crew's work continuity in repetitive construction projects. *J. Constr. Eng. Manag.* **2020**, *146*, 04020043. [[CrossRef](#)]
2. Dasović, B.; Galić, M.; Klanšek, U. A survey on integration of optimization and project management tools for sustainable construction scheduling. *Sustainability* **2020**, *12*, 3405. [[CrossRef](#)]
3. Czarnecki, L.; Kaproń, M. Definiowanie zrównoważonego budownictwa. Cz. 2. *Mater. Bud.* **2010**, *nr2*, 46–47.
4. Hermarij, J. *Better Practices of Project Management based on IPMA competences*; Van Haren: Hertogenbosch, The Netherlands, 2013.
5. KPMG International. *Global Construction Survey 2015*; Climbing the curve; KPMG International: Amstelveen, The Netherlands, 2015.
6. Ishii, N.; Takano, Y.; Muraki, M. A bidding price decision process in consideration of cost estimation accuracy and deficit order probability for engineer-to-order manufacturing. In *Technical Report No. 2011-1*; Tokyo Institute of Technology: Tokyo, Japan, 2011.
7. Doloi, H. Cost overruns and failure in project management: Understanding the roles of key stakeholders in construction projects. *J. Constr. Eng. Manag.* **2012**, *139*, 267–279. [[CrossRef](#)]
8. Maemura, Y.; Kim, E.; Ozawa, K. Root Causes of Recurring Contractual Conflicts in International Construction Projects: Five Case Studies from Vietnam. *J. Constr. Eng. Manag.* **2018**, *144*, 05018008. [[CrossRef](#)]
9. Abotaleb, I.S.; El-Adaway, I.H. Construction bidding markup estimation using a decision theory approach. *J. Constr. Eng. Manag.* **2016**, *143*, 04016079. [[CrossRef](#)]
10. Miller, R.; Lessard, D.R. *Evolving Strategy: Risk Management and the Shaping of Mega-Projects*; Edward Elgar: Cheltenham, UK, 2008.

11. Dikmen, I.; Budayan, C.; Talat Birgonul, M.; Hayat, E. Effects of Risk Attitude and Controllability Assumption on Risk Ratings: Observational Study on International Construction Project Risk Assessment. *J. Manag. Eng.* **2018**, *34*, 04018037. [[CrossRef](#)]
12. Messner, J.I. An Information Framework for Evaluating International Construction Projects. Ph.D. Thesis, Pennsylvania State University, University Park, PA, USA, 1995.
13. Fayek, A.; Ghoshal, I.; AbouRizk, S. A survey of the bidding practices of Canadian civil engineering construction contractors. *Can. J. Civ. Eng.* **1999**, *26*, 13–25. [[CrossRef](#)]
14. Chua, D.K.H.; Li, D.Z.; Chan, W.T. Case-based reasoning approach in bid decision making. *J. Constr. Eng. Manag.* **2001**, *127*, 35–45. [[CrossRef](#)]
15. El-Mashaleh, M.S. Decision to bid or not to bid: A data envelopment analysis approach. *Can. J. Civ. Eng.* **2010**, *37*, 37–44. [[CrossRef](#)]
16. An, M.; Baker, C.; Zeng, J. A fuzzy-logic-based approach to qualitative risk modeling in the construction process. *World J. Eng.* **2005**, *2*, 1–12.
17. El-Mashaleh, M.S. Empirical framework for making the bid/no-bid decision. *J. Manag. Eng.* **2012**, *29*, 200–205. [[CrossRef](#)]
18. Egemen, M.; Mohamed, A.N. A framework for contractors to reach strategically correct bid/no bid and mark-up size decisions. *Build. Environ.* **2007**, *42*, 1373–1385. [[CrossRef](#)]
19. Chan, E.H.; Au, M.C. Factors influencing building contractors' pricing for time-related risks in bids. *J. Constr. Eng. Manag.* **2009**, *135*, 135–145. [[CrossRef](#)]
20. Williams, T.P.; Gong, J. Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Autom. Constr.* **2014**, *43*, 23–29. [[CrossRef](#)]
21. Zeng, J.; An, M.; Smith, N.J. Application of a fuzzy based decision making methodology to construction project risk assessment. *Int. J. Proj. Manag.* **2007**, *25*, 589–600. [[CrossRef](#)]
22. Williams, T.P. Bidding ratios to predict highway project costs. *Eng. Constr. Archit. Manag.* **2005**, *12*, 38–51. [[CrossRef](#)]
23. Lee, J.; Yi, J.S. Predicting Project's Uncertainty Risk in the Bid Process by Integrating Unstructured Text Data and Structured Numerical Data Using Text Mining. *Appl. Sci.* **2017**, *7*, 1141. [[CrossRef](#)]
24. Ahmad, I.; Minkarah, I. Questionnaire survey on bid in construction. *J. Manag. Eng.* **1988**, *4*, 229–243. [[CrossRef](#)]
25. Shash, A.A. Factors considered in tendering decisions by top UK contractors. *Constr. Manag. Econ.* **1993**, *11*, 111–118. [[CrossRef](#)]
26. Chua, D.K.H.; Li, D. Key factors in bid reasoning model. *J. Constr. Eng. Manag.* **2000**, *126*, 349–357. [[CrossRef](#)]
27. Wanous, M.; Boussabaine, A.A.; Lewis, J. To bid or not to bid: A parametric solution. *Constr. Manag. Econ.* **2000**, *18*, 457–466. [[CrossRef](#)]
28. Han, S.H.; Diekmann, J.E. Approaches for making risk-based go/no-go decision for international projects. *J. Constr. Eng. Manag.* **2001**, *127*, 300–308. [[CrossRef](#)]
29. Bagies, A.; Fortune, C. Bid/no-bid decision modelling for construction projects. In Proceedings of the 22nd Annual ARCOM Conference, Birmingham, UK, 4–6 September 2006.
30. Leśniak, A.; Plebankiewicz, E. Modeling the decision-making process concerning participation in construction bidding. *J. Manag. Eng.* **2015**, *31*, 04014032. [[CrossRef](#)]
31. Ahiaga-Dagbui, D.D.; Smith, S.D. Dealing with construction cost overruns using data mining. *Constr. Manag. Econ.* **2014**, *32*, 682–694. [[CrossRef](#)]
32. Delaney, J.W. The Effect of Competition on Bid Quality and Final Results on State DOT Projects. Ph.D. Thesis, State University of New York at Buffalo, Buffalo, NY, USA, 2018.
33. Tanaka, T. Analysis of Claims in Us Construction Projects. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 1988.
34. Erdis, E.; Ozdemir, S.A. Analysis of technical specification-based disputes in construction industry. *KSCE J. Civ. Eng.* **2013**, *17*, 1541–1550. [[CrossRef](#)]
35. Trost, S.M.; Oberlender, G.D. Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *J. Constr. Eng. Manag.* **2003**, *129*, 198–204. [[CrossRef](#)]
36. OGS. Pre-Bid Inquiry & Response Policy. Available online: <https://online.ogs.ny.gov/dnc/contractorconsultant/esb/prebidinquiryresponsepolicy.asp> (accessed on 27 November 2020).
37. Duzkale, A.K.; Lucko, G. Exposing uncertainty in bid preparation of steel construction cost estimating: II. Comparative analysis and quantitative CIVIL classification. *J. Constr. Eng. Manag.* **2016**, *142*, 04016050. [[CrossRef](#)]
38. Baloi, D.; Price, A.D. Modelling global risk factors affecting construction cost performance. *Int. J. Proj. Manag.* **2003**, *21*, 261–269. [[CrossRef](#)]
39. Daoud, O.E.; Allouche, E.N. Bid queries as a gauge for quality control of documents. In Proceedings of the Canadian Society for Civil Engineering, Moncton, NB, Canada, 4–7 June 2003.
40. Project Management Institute (PMI). *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*, 4th ed.; Project Management Institute: Newtown Square, PA, USA, 2008.
41. Winch, G.M. *Managing Construction Projects*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
42. Katal, A.; Wazid, M.; Goudar, R.H. Big data: Issues, challenges, tools and good practices. In Proceedings of the 2013 Sixth International Conference on Contemporary Computing (IC3), Noida, India, 8–10 August 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 404–409.

- 
43. Christodoulou, S. Optimum bid markup calculation using neurofuzzy systems and multidimensional risk analysis algorithm. *J. Comput. Civ. Eng.* **2004**, *18*, 322–330. [[CrossRef](#)]
  44. Lo, W.; Lin, C.L.; Yan, M.R. Contractor's opportunistic bid behavior and equilibrium price level in the construction market. *J. Constr. Eng. Manag.* **2007**, *133*, 409–416. [[CrossRef](#)]