A DATA-DRIVEN APPROACH TO IDENTIFY-QUANTIFY-ANALYSE CONSTRUCTION RISK FOR HONG KONG NEC PROJECTS

Ming-Fung Francis SIU[*] , Wing-Yan Jacqueline LEUNG, Wai-Ming Daniel CHAN

Department of Building and Real Estate, Hong Kong, China

Abstract. Project risks must be managed to deliver construction projects on time and within budget. In recent years, the New Engineering Contract (NEC) provides an alternate contracting method for procuring construction projects. As stipulated in the NEC contract, NEC risk register must be used to record any project risks. The risk register is designed to record each risk item in the context of textual description, likelihood, and consequence. However, it is time-consuming to identify, quantify, and analyse NEC project risks based on experience, questionnaire, simulation, and data-mining approach. Any method to fully utilise the records of NEC risk registers of past projects for managing NEC project risks remains unexplored. As such, a data-driven approach is proposed to categorise common risks of NEC projects and to analyse risk rating of risk categories by combining the use of text mining analysis and decision tree analysis. A practical case study in Hong Kong is used to illustrate the method of application. The top four common types of NEC project risks, which are ground and utilities, design information, structures, and workmanship, were identified, quantified, and analysed. The new approach helps NEC project planners to identify, quantify, and analyse NEC project risks time-efficiently.

Keywords: risk identification, risk quantification, risk analysis, risk register, risk category, risk rating, decision tree, text mining, New Engineering Contract.

Introduction

There is a need for project risk management to extirpate any negative impact associated with construction project delivery regarding time and budget (i.e. to deliver the project on time and within budget). Using the traditional project procurement method, relationships between contracting parties (i.e. developers, architects, engineers, surveyors, and contractors) are adversarial. Project risks are intentionally transferred to other parties (Charoenngam, Yeh 1999; Mead 2007). As a result, most construction projects are behind schedule and over-budget. In 2001, the Construction Industry Review Committee (CIRC) emphasised that a partnering approach should be encouraged for project procurement in Hong Kong. The "New Engineering Contract (NEC)" is a family of partnering contracts that promote a collaborative working environment based on the spirit of mutual trust and co-operation by defining legal clauses in ordinary language (ICE 2013). It facilitates risk management by stipulating the parties to meet and seek mutually beneficial solutions in order to overcome the identified project risks. Subject to the fact that the NEC contract has a proven track record

for procuring and delivering excellent infrastructure and building projects in overseas countries such as the United Kingdom, France, Sweden, South Africa, New Zealand, and Australia (Mooney, C., Mooney, E. 2014; Chang 2014; Loh, Ofori 2000), since 2009, the Hong Kong Government has adopted the NEC contract as one of the feasible means for tendering public works.

Project risk is defined as an uncertain event or condition which has either a negative or positive impact on a project (Cretu *et al.* 2011; PMI 2013; Cagliano *et al.* 2014). The negative impact of a project risk item can be quantified by risk rating. The rating is determined by measuring the associated likelihood of occurrence and level of consequence (AbouRizk 2009; Chan *et al.* 2011; Sharma, Swain 2011; Li *et al.* 2016). As shown in Table 1, a sample risk matrix can be used to determine the risk rating. The scale of likelihood is standardised as almost certain, likely, possible, unlikely, and rare; the scale of consequence is standardised as serious, major, moderate, minor, and low; and the scale of risk rating is standardised as extreme, high, medium, low, and negligible.

*Corresponding author. E-mail: *francis.siu@polyu.edu.hk*

Table 1. Risk matrix for risk rating

Risk rating		Likelihood				
		Almost certain	Likely	Possible	Unlikely	Rare
Consequence	Serious	Extreme	Extreme	Extreme	High	Medium
	Major	Extreme	Extreme	High	Medium	Medium
	Moderate	Extreme	High	Medium	Medium	Low
	Minor	High	Medium	Medium	Low	Low
	Low	Medium	Medium	Low	Low	Low

Table 2. NEC contract clauses for risk management

Stage	Risk management process	NEC contract clause number and description
1	Risk identification	Clause 16.1 – Risk early warning Clause 16.3, 16.4 – Risk register
2	Risk quantification and analysis	Clause 16.2, 16.3 – Risk reduction meeting Clause 16.3, 16.4 – Risk register
3	Risk mitigation	Clause 16.2, 16.3 – Risk reduction meeting
4	Risk monitoring and control	Clause 16.3, 16.4 – Risk register

Risk management enables parties to cope with project uncertainties by taking reasonable steps to avoid, reduce, and control project risks (Wood, Ellis 2003; Sharma, Swain 2011). It involves four key stages: risk identification, risk quantification and analysis, risk mitigation, and risk monitoring and control (AbouRizk 2009; Cretu *et al.* 2011; Cagliano *et al.* 2014). The aim is to increase the probability and impact of positive events and decrease the probability and impact of negative events. With an effective risk management system, a construction company enjoys better decision making, greater productivity, and financial saving (PMI 2013). The NEC contract helps the parties to establish a risk management system for managing project risks. Table 2 summarises relevant NEC clauses for risk management (ICE 2013). Clause 16.1 establishes an early warning mechanism for project risks in a NEC project. A risk register shall be developed and managed for risk identification, quantification and analysis. Notably, the risk register is a document used to record project risk items, which are brainstormed and characterised by textual description, likelihood, and consequence (Mubin, S., Mubin, G. 2008). In general, the format of the risk matrix is identical to the one shown in Table 1 (Zhao *et al.* 2010; Mahamid 2011; Mahamid *et al.* 2015; Alhajri, Alshaibani 2018). Clause 16.2 and 16.3 establish the meetings for risk reduction. It is mandatory for both parties to attend the risk reduction meeting to decide the actions to be taken to mitigate the identified project risks. As such, risk management is proactively encouraged by open engagement and collaboration between parties.

However, NEC project risks are brainstormed, written, and analysed based on planner experience. The meeting time for risk identification is spent on brainstorming the risk items for past projects. It is also time-consuming to apply the existing experience-based approach, questionnaire-based approach, simulation approach, and data-mining approach (as discussed in the literature) to identify, quantify, and analyse the critical project risks. Given the data of NEC project risks archived in the NEC risk register as stipulated in NEC contract clauses, we were motivated to explore a novel approach to benchmark NEC project risks by fully utilising the past records of NEC risk registers to facilitate "risk identification" and "risk quantification and analysis" process in a time-efficient way.

The research aim was to propose a novel data-driven approach for facilitating risk identification and risk quantification and analysis in the context of NEC projects. Risk identification is performed by mining the textual descriptions given in NEC risk registers. Risk quantification and analysis is performed by building the decision tree of risk rating where the tree is branched based on risk likelihood and consequence. A software platform, IBM SPSS modeller (IBM 2018), was chosen to automate the proposed approach. As a result, common types of NEC project risks were identified, and the NEC risk rating of risk categories benchmarked. Thanks to the proposed approach, the time for brainstorming the project risk items during the risk management meeting can be significantly reduced. Also, the developed decision trees help the project planners to manage the project risk items from the perspective of its risk categories, instead of managing the project risk items one by one.

In the following sections, a literature review is firstly given on the advancement of tools and techniques used for construction risk management in connection with the experience-based approach, questionnaire-based approach, simulation approach, and data-mining approach. Then, the data-driven approach is proposed. A practical case study is given to illustrate the benchmarked results. Due to the limited number of NEC projects in Hong Kong, a

total number of seven NEC risk registers for both on-going and completed public works in Hong Kong was collected and used. Discussion is given to deliberate the NEC project risks. Results are validated by interviewing NEC professionals in Hong Kong. Lastly, conclusions are drawn informing the opportunities for future research.

1. Literature review

This section provides an overview of the advancement of tools and techniques used for construction risk management. The approaches are centred on experience-based, questionnaire-based, simulation, and data-mining approaches.

During the 1980s, studies devoted to the examination of risk management in the context of construction management were limited. Research communities emphasised the needs of risk analysis to avoid project cost overrun. The project risk items were identified solely based on personal experience of relevant past projects. For instance, Chapman and Cooper (1983) introduced a risk engineering approach for determining the impact of probability of activity delays on project duration. Cooper *et al.* (1985) proposed a risk breakdown structure for categorising project risks. Hayes *et al.* (1987) proposed the procedures of risk identification, analysis, and response to manage project risks. Franke (1987) advocated a mathematic model to evaluate project contingency based on risk likelihood and risk value. Kangari (1988) proposed an integrated knowledge system to manage project risks based on the applications of knowledge-bases, data-bases, and fuzzy sets. Kangari and Riggs (1989) introduced a fuzzy set theory to linguistically analyse project risks.

As construction projects became complex and large-scale with more uncertainties, construction risk management became a hot research topic in the 1990s. The researchers relied on a questionnaire-based approach to collect data identifying project risks. For example, Akintoye and MacLeod (1997) carried out a questionnaire survey to examine industry perceptions of project risks. The authors found that project risk data was limited for practical application of risk analysis. Bing and Tiong (1999) carried out a questionnaire survey to identify project risk factors (i.e. internal, project-specific, and external factors). He identified the critical risk factors as the financial aspects of joint venture, government policies, economic conditions, and project relationships. Wang *et al.* (2004) conducted a questionnaire survey to identify eleven critical risk factors for construction industries in developing countries. The authors identified the critical risk factors as approval and permit, change in law, justice reinforcement, local partner's creditworthiness, political instability, cost overrun, corruption, inflation and interest rates, government policies, government influence on dispute, and termination of joint venture. Tang *et al.* (2007) determined five critical project risks in the Chinese construction industry. The project risks were identified as poor work quality, premature facility failure, poor safety, inadequate or incorrect design,

and financial risk. Choudhry and Iqbal (2013) adopted a similar approach to evaluate risk management practice in Pakistan. The critical project risks were identified as financial factors, economic factors, quality, premature facility failure, lack of planning and management, change in design and work scope, corruption, claim and dispute, inadequate or incorrect design, and quantity variation. Sarvari *et al.* (2014) identified critical project risks in Malaysia. They identified the critical risk factors as third-party tort liability, interest rate volatility, construction cost overrun, and change in law. Some researchers attempted to abstract the status quo of construction risk management in specific countries using the questionnaire-based approach. For example, Ahmed *et al.* (1999) examined the perception of risk management in Hong Kong from the perspective of contractors and owners. Results showed that the Contractor devoted more resources to managing the project risks than the Owners. Kim and Bajaj (2000) investigated the practice of risk management in South Korea. They found that project planners managed the project risks based on their experience instead of using a formal risk management system. In China, Tang *et al.* (2007) found that the application of risk management systems was inadequate. In Singapore, Hwang *et al.* (2014) discovered that the implementation level of risk management systems was low.

With the technological advance of computing power, the researchers focused on performing risk analysis using computers. For instance, Hull (1990) introduced a risk analysis approach to estimate project cost factoring in project risks. Al-Bahar and Crandall (1990) introduced a construction risk management system to identify, analyse, and evaluate project risks by using Monte Carlo simulation. Mustafa and Al-Bahar (1991) presented an analytic hierarchy process to prioritise project risks. Wirba *et al.* (1996) introduced a fuzzy set technique to examine the dependency of project risks linguistically. Tavares *et al.* (1998) proposed a simulation model for project planners to manage the project risks of time and cost overrun by considering the randomness of activity duration and cost. Tah and Carr (2000) developed a fuzzy model based on the relationships between project risks and project performance for qualitative risk assessment. Baccarini and Archer (2001) prioritised project risk items based on risk rating (by multiplying its risk likelihood and consequence). As such, limited project resources can be allocated to manage the project risk items with high risk rating. Jannadi and Almishari (2003) developed a computer-based risk assessor model to calculate an activity risk score in consideration of risk severity, exposure, and probability. Cagno *et al.* (2007) proposed a three-dimensional model to categorise project risks by risk sources, affected activities, and risk owners. Han *et al.* (2008) introduced a web-based system for managing project risks throughout project stages.

Recently, with the growth of big datasets collected by sensing technologies, data mining became a hot research topic in construction. For example, Kim *et al.* (2008) presented a knowledge discovery in databases (KDD) framework to analyse massive construction datasets and

deployed data mining techniques to identify the main causes of construction delays. Carrillo *et al.* (2011) found that the performance of finance, time, quality, and health and safety of future projects could be improved by performing text analysis, link analysis, and dimensional analysis using the data collected from multiple projects. Cheng *et al.* (2012) developed a risk early warning system based on KDD to improve the efficiency of bidding works and sharing information. Yildiz *et al.* (2014) proposed a knowledge-based risk mapping tool to assess risk-related variables, identify risk path leading to cost overrun, and store risk-related knowledge. Ding *et al.* (2016) proposed an ontology-based methodology for managing construction risk knowledge embedded in building information models. Serpell *et al.* (2017) proposed a knowledge-based risk management support system for assisting the contractors to manage project risks.

In a nutshell, although construction risk management has been researched over the past decades, research endeavours mainly focused on the development of the experience-based approach to brainstorming project risk items, the questionnaire-based approach to identify critical project risks and benchmark industry perceptions of project risks, the simulation approach to simulate (time and cost) impact of project risks, and the data-mining approach to mine project risks based on available datasets. It is time-consuming to apply these approaches because substantial effort is required in collecting the relevant data to perform analysis. For example, time and effort for brainstorming project risk items (experience-based approach), collecting feedback (questionnaire-based approach), building simulation models (simulation approach), and collecting relevant project risk data (data-mining approach).

Nevertheless, project risks must be archived in NEC risk registers as stipulated in the NEC contract if the NEC project delivery method is used. Limited research has been conducted on the exploration of the techniques employed to benchmark and categorise NEC project risks by fully utilising the records of NEC risk registers of past projects. Therefore, the authors were motivated to categorise and benchmark NEC project risks by the applications of text mining analysis and decision tree analysis. Proposed data-driven approach enables NEC project planners to automate "risk identification" and "risk quantification and analysis" process for construction risk management.

2. The proposed approach for "risk identification" and "risk quantification and analysis"

This section proposes a novel data-driven approach for use on NEC projects to automate "risk identification" and "risk quantification and analysis" process. Figure 1 shows an overview of the proposed approach. The input is the data recorded in NEC risk registers. The processes are to develop a consolidated NEC risk register, perform text mining analysis, develop a categorised NEC risk register, and perform decision tree analysis. The outputs are the identified NEC risk categories and the benchmarked NEC risk rating. The approach consists of four steps. Steps 1 and 2 perform text mining analysis for risk identification, and Steps 3 and 4 perform decision tree analysis for risk quantification and analysis.

2.1. Step 1: Develop a consolidated risk register

NEC risk registers identify project risks by detailing textual descriptions of project risk items. A consolidated NEC
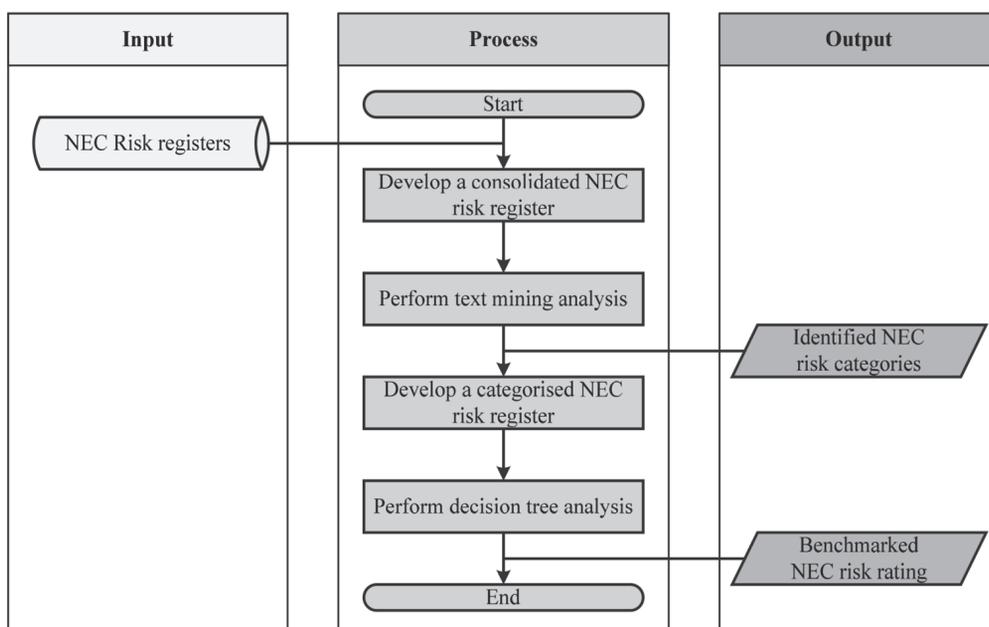
Figure 1. Proposed data-driven approach for risk identification, quantification, and analysis

Table 3. Sample of consolidated risk register

Risk ID	Project ID	Risk item descriptions	Consequence of risk	Likelihood of risk
1	1	\<Textual description of risk item 1\>	\<Consequence\>	\<Likelihood\>
2	1	\<Textual description of risk item 2\>	\<Consequence\>	\<Likelihood\>
3	2	\<Textual description of risk item 3\>	\<Consequence\>	\<Likelihood\>
…	…	…	…	…

Table 4. Sample of categorised risk register

Risk ID	Risk item descriptions	\<Category 1\>	\<Category 2\>	…
1	\<Textual description of risk item 1\>	\<True/False\>	\<True/False\>	…
2	\<Textual description of risk item 2\>	\<True/False\>	\<True/False\>	…
3	\<Textual description of risk item 3\>	\<True/False\>	\<True/False\>	…
…	…	…	…	…

risk register shall be developed by combining the NEC risk registers from different NEC projects. Notably, in this consolidated risk register, a standardised scale of risk likelihood and consequence (e.g. Table 1) shall be used to value the risk likelihood and consequence respectively for each project risk item. Table 3 shows the recommended structure of the consolidated risk register. For example, if the consequence of risk and likelihood of risk of the "1st" NEC risk item "Changes of works information" in Project "1" are "Minor" and "Possible" respectively, the cell values associated with Columns "Risk ID", "Project ID", "Risk item descriptions", "Consequence of risk", and "Likelihood of risk" in Row 1 of the consolidated risk register are labelled as "1", "1", "Changes of works information", "Minor", "Possible", respectively.

2.2. Step 2: Perform text mining analysis

The consolidated NEC risk register is used to categorise common types of NEC project risks. Natural language processing (NLP), which is a linguistic-based text mining technique, is deployed to extract and categorise key concepts from unstructured text data. The NLP technique analyses words, phrases, syntax, and structure of a text or paragraph based on human languages using computer power. The technique extracts the concepts and groups the concepts into classes. The NLP rules are concept root deviation, concept inclusion, semantic network, and co-occurrence. Concept root deviation examines the morphological features of the concepts, groups the concepts into categories by searching the suffixes of each word in the concept, and terminated based on language-specific rules. Concept inclusion builds the categories by adopting lexical series algorithms to identify the concepts which are included in the other concepts. Semantic networks build the categories using a built-in network of word relationships. Co-occurrence identifies and groups the concepts which frequently appear together and are strongly related among the records. To improve the results of text mining,

a combination of the aforementioned rules shall be used. By using the NLP technique, the concept is produced for better understanding of the textual descriptions of NEC project risks. As a result, NEC risk categories can be extracted and identified.

2.3. Step 3: Develop a categorised risk register

A categorised NEC risk register shall be developed by populating the identified NEC risk categories back to each project risk item. Thus, each project item is characterised by a particular NEC risk category. Table 4 illustrates an example of a categorised risk register. For example, if the project risk item "Changes of works information during the course of the works" is characterised by the NEC risk categories "Variation" and "Design information", the categories "Variation" and "Design information" are labelled as "True" while the others as "False" for this project risk item.

2.4. Step 4: Perform decision tree analysis

Based on the value of risk likelihood and risk consequence recorded for each project risk item in the risk registers, the risk rating of the project risk item can be determined using risk matrix. Records of project risk items are extracted for each identified risk categories. Decision tree analysis is used to generate a flowchart-like structure for a particular risk category in which each tree node presents how severe is the risk rating in connection with the scales of risk likelihood and risk consequence. Tree branching is based on either risk likelihood or risk consequence. The higher level of branching is proposed to be the one with greater importance to project risk rating. Chi-squared automatic interaction detection (CHAID) is selected to identify the optimal splits with more than two splits at any tree level (Kass 1980; IBM 2011). When performing CHAID decision tree analysis, the categories with fewer numbers of records are first merged with the most similar categories

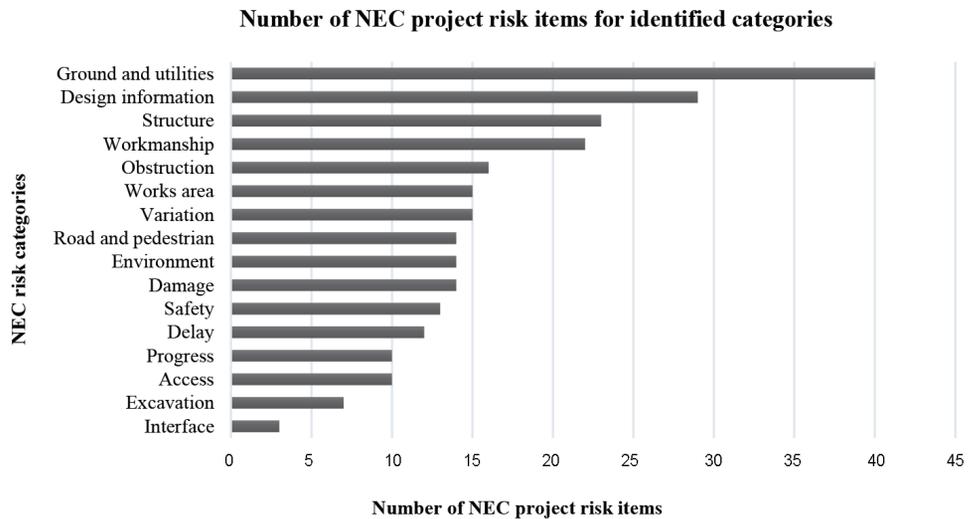**Number of NEC project risk items for identified categories**

Figure 2. Number of NEC project risk items for identified NEC risk categories

for predictors and the records are then split based on the merged categories. This iteration process is repeated until the stopping rule is satisfied (e.g. minimum records in parent branch, minimum records in child branch). As such, a decision tree for a particular risk category is built by setting the target field as risk rating and setting the predictors as risk likelihood and risk consequence.

3. Benchmarked results of NEC project risks in Hong Kong

A total number of seven risk registers for NEC projects in Hong Kong was used to illustrate the practical application of the proposed four-step data-driven approach.

- Step 1: A NEC consolidated risk register was developed in Microsoft Excel format given the seven risk registers by consolidating the textual descriptions of project risk items. A consolidated risk register with 240 data records is shown in Appendix A.
- Step 2: To extract the concepts from unstructured textual descriptions of project risk items, the software platform IBM SPSS modeller (2018), was used. The modelling steps with SPSS modeller are detailed in Appendix B. Totally, 16 NEC risk categories were identified.
- Step 3: A NEC categorised risk register was developed in Microsoft Excel format as shown in Appendix C. Vertical axis and horizontal axis of categorised risk register represent project risk items and project risk categories respectively. A cell value of "True" infers that the project risk item is characterized by a particular project risk category. A cell value of "False" implies that the project risk item is not characterized by a particular project risk category.
- Step 4: A risk rating for each NEC project risk item was determined using a risk matrix (Table 1). In general, risk rating was at medium risk level. Table 5 illustrates the number and percentage of records

Table 5. Risk rating of identified project risk items in Hong Kong NEC projects

Risk rating	Number of records	Percentage
Extreme	12	5%
High	53	22%
Medium	138	58%
Low	37	15%

with respect to risk rating level. IBM SPSS modeller (2018) was employed to develop the decision trees. The modelling steps when using IBM SPSS modeller are detailed in Appendix D. Four decision trees were developed for the top four critical NEC risk categories.

3.1. Results of common NEC risk categories

Figure 2 shows the number of NEC project risk items for 16 benchmarked risk categories in Hong Kong NEC projects. It was found that the highest number of NEC project risk items is associated with the ground and utilities category. It reveals that NEC project risks related to ground and utilities are extremely likely to occur. Apart from this, the number of NEC project risk items associated with design information, structure, and workmanship categories were the second, third, and fourth highest, respectively. This implies that NEC project risks, in these risk categories, are highly likely to occur. The following sub-sections discuss the results of decision tree analysis for these four categories.

3.2. The relative importance of risk likelihood and risk consequence for NEC risk categories

Table 6 shows the relative importance of risk likelihood and risk consequence towards risk rating for the four critical risk categories. For the ground and utilities category,

Risk rating

Extreme	2%
High	55%
Medium	33%
Low	10%

(n=49)

Likelihood

Likely

Extreme	0%
High	95%
Medium	5%
Low	0%

(n=19)

Possible

Extreme	0%
High	39%
Medium	61%
Low	0%

(n=23)

Rare/unlikely

Extreme	0%
High	0%
Medium	17%
Low	83%

(n=6)

Very unlikely

Extreme	100%
High	0%
Medium	0%
Low	0%

(n=1)

Consequence

Minor

Extreme	0%
High	0%
Medium	100%
Low	0%

(n=1)

Moderate

Extreme	0%
High	100%
Medium	0%
Low	0%

(n=18)

Major

Extreme	0%
High	100%
Medium	0%
Low	0%

(n=9)

Minor/moderate

Extreme	0%
High	0%
Medium	100%
Low	0%

(n=14)

Major

Extreme	0%
High	0%
Medium	100%
Low	0%

(n=1)

Minor

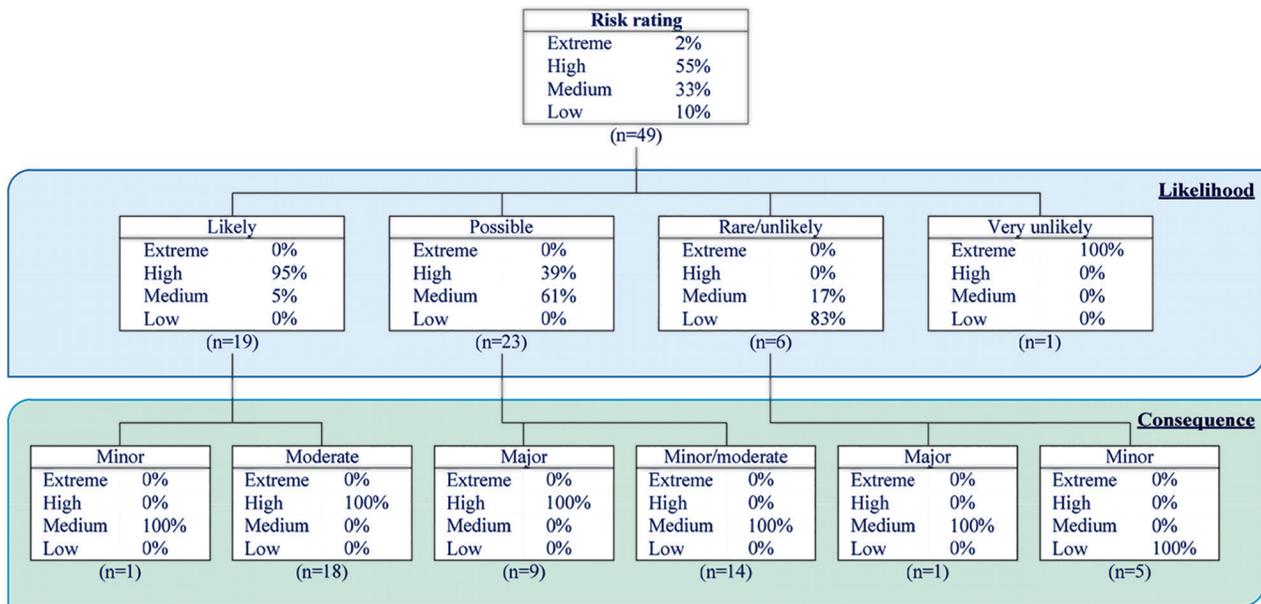|	
Extreme	0%
High	0%
Medium	0%
Low	100%

(n=5)

Figure 3. Decision tree model for ground and utilities

Table 6. Relative importance of risk likelihood and consequence towards risk rating for four critical NEC risk categories

Risk categories	Risk likelihood	Risk consequence
Ground and utilities	0.78	0.22
Design information	0.60	0.40
Structures	0.82	0.18
Workmanship	0.42	0.58

risk likelihood is more important than risk consequence towards risk rating. When developing the decision tree for ground and utilities, risk likelihood is used as the criterion for branching at the first tree level while risk consequence is used as the criterion for branching at the second tree level. For the design information category, the relative importance of risk likelihood is slightly greater than risk consequence towards risk rating in the ratio 3:2. For the structures category, the relative importance of risk likelihood is significantly greater than risk consequence towards risk rating. For the workmanship category, the relative importance of risk consequence is greater than risk likelihood towards risk rating.

3.3. Decision tree analysis for "ground and utilities"

Figure 3 presents the decision tree model for the ground and utilities category (where *n* represents the total number of NEC project risk items found in NEC risk registers). Overall, 55% of project risks are at high-risk level, and 33% of risks are at medium risk level. 47% of project risks may possibly to occur, and 39% of project risks are likely to occur. 37% of project risks with likely occurrence have moderate consequence, 29% of project risks with possible occurrence have minor/moderate consequence, and 18%

of project risks with possible occurrence have major consequence.

3.4. Decision tree analysis for "design information"

Figure 4 demonstrates the decision tree model for the design information category. Overall, 69% of project risks are at medium risk level, and 28% are at high-risk level. 79% of project risks are possible/unlikely to occur, and 21% of project risks are likely to occur. Since the consequence of the project risks with possible/unlikely occurrence is identical, no further split is required. 17% of project risks with likely occurrence have moderate consequence.

3.5. Decision tree analysis for "structures"

Figure 5 shows the decision tree model for the structures category. Overall, 61% of project risks are at medium risk level, and 22% are at high-risk level. 83% of project risks are possible/unlikely to occur, and 17% of project risks are likely to occur. Since the consequence of the project risks with possible/unlikely occurrence is identical, no further split is required. 13% of project risks with likely occurrence have moderate consequence.

3.6. Decision tree analysis for "workmanship"

Figure 6 illustrates the decision tree model for the workmanship category. Overall, 63% of project risks are at medium risk level, and 18% are at high risk level. 86% of project risks are of major/moderate consequence, and 14% of project risks are of minor consequence. Since the likelihood of the project risks with minor consequence is identical, no further split is required. 73% of project risks with possible/unlikely occurrence have major/moderate consequence.
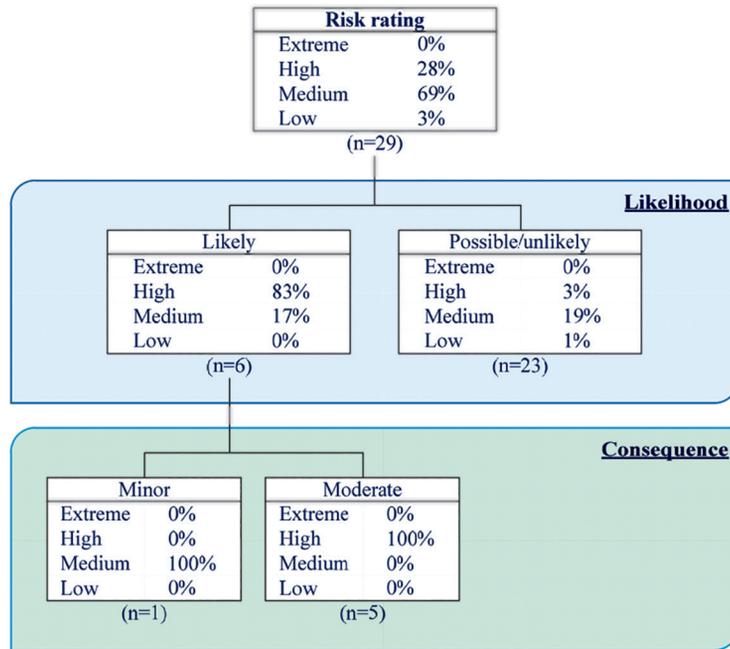
Risk rating	
Extreme	0%
High	28%
Medium	69%
Low	3%
(n=29)	

Likelihood

Likely	
Extreme	0%
High	83%
Medium	17%
Low	0%
(n=6)	

Possible/unlikely	
Extreme	0%
High	3%
Medium	19%
Low	1%
(n=23)	

Consequence

Minor	
Extreme	0%
High	0%
Medium	100%
Low	0%
(n=1)	

Moderate	
Extreme	0%
High	100%
Medium	0%
Low	0%
(n=5)	

Figure 4. Decision tree model for design information

Risk rating	
Extreme	4%
High	22%
Medium	61%
Low	13%
(n=23)	

Likelihood

Likely	
Extreme	25%
High	75%
Medium	0%
Low	0%
(n=4)	

Possible/unlikely	
Extreme	0%
High	11%
Medium	74%
Low	16%
(n=19)	

Consequence

Minor	
Extreme	100%
High	0%
Medium	0%
Low	0%
(n=1)	

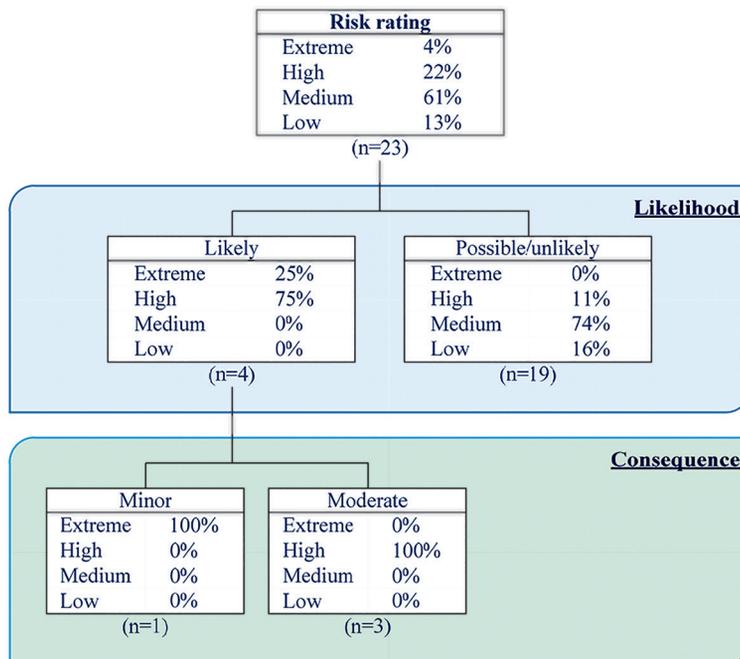| Moderate	
Extreme	0%
High	100%
Medium	0%
Low	0%
(n=3)	

Figure 5. Decision tree model for structures

4. Discussions

4.1. Explanation of benchmarked NEC project risks in Hong Kong

The NEC project risks commonly occurring in Hong Kong NEC projects were categorised as ground and utilities, design information, structures, and workmanship. In general, project risks in the ground and utilities category are identified at the medium to high-risk levels. Since the underground system and utilities in Hong Kong were developed by public and private utility undertakers (e.g. Towngas, China light and power), it is challenging to identify the exact locations of existing utilities within and nearby site areas using out-of-date information and as-built records. Utility design is complex. The structural elements may clash with the existing utility elements such that modification of utility design is required. To mitigate project risk, the Contractor should carry out a utility survey at the initial construction stage to identify any existing utilities within the site area. Building information models may also help to detect clashes between the existing and designed construction products.
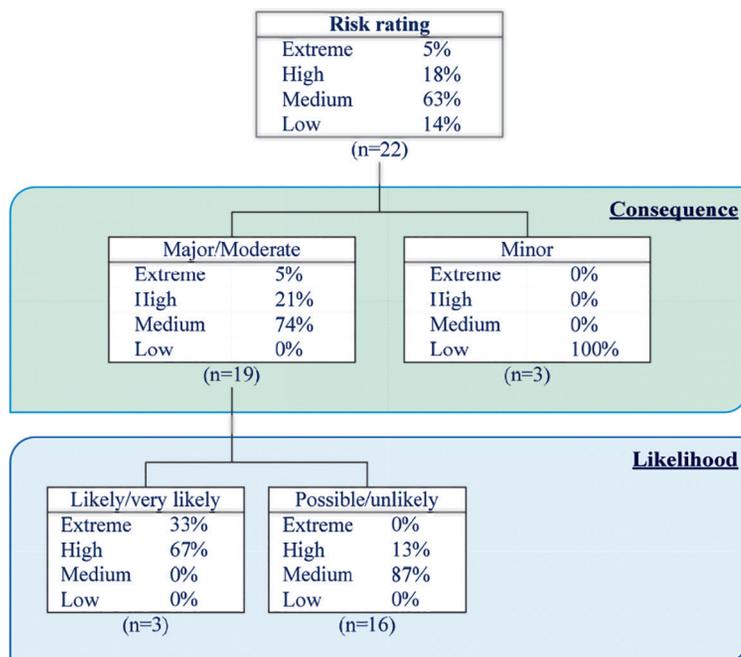
Risk rating	
Extreme	5%
High	18%
Medium	63%
Low	14%
(n=22)	

Consequence

Major/Moderate	
Extreme	5%
High	21%
Medium	74%
Low	0%
(n=19)	

Minor	
Extreme	0%
High	0%
Medium	0%
Low	100%
(n=3)	

Likelihood

Likely/very likely	
Extreme	33%
High	67%
Medium	0%
Low	0%
(n=3)	

Possible/unlikely	
Extreme	0%
High	13%
Medium	87%
Low	0%
(n=16)	

Figure 6. Decision tree model for workmanship

Project risks in the design information category are identified on average at medium risk level. In practice, a construction product is designed by professionals (e.g. architects, building engineers, surveyors, and consultants). Design information should be prepared with reference to the design and specification (e.g. laws and regulations, guidance notes, British Standards). However, modification of design information may incur during the construction stage owing to changes in site condition, law regulation and the owner's requirement. Although project risks in this category are difficult to eliminate, the Contractor should carefully review the product design and provide suggestions to improve constructability during the design stage.

Project risks in the structures category are identified at medium risk level. In Hong Kong, most construction sites are located near existing buildings and infrastructure. Damage to existing structures or subsidence may incur during the construction stage. The Contractor may spend extra time and cost to carry out remedial works. Besides, the Contractor is required to coordinate and interface with other contractors in order not to damage or obstruct others' works. However, damage and inconsistent structures are often found in the interfacing area. To mitigate the project risks, the Contractor should regularly monitor soil settlement during the construction stage, provide sufficient protection to existing structures to avoid any damage, and provide effective communication channels between the Contractor and relevant parties for coordinating the works in the interfacing area.

Project risks in the workmanship category are identified on average to be of medium risk level. In practice, the Contractor requires the submission of method statements, drawings, and risk assessments before carrying out relevant works. The Engineer may approve or reject the Contractor's submission based on the quality of workmanship. However, multi-layer subletting is common practice in the Hong Kong construction industry. The main-contractor may face difficulties in ensuring all sub-contractors are qualified and experienced in delivering quality work. To mitigate project risks, the Contractor should avoid the multi-layer subletting practice and provide sufficient work supervision during construction.

4.2. Validation by interviewing NEC professionals in Hong Kong

Face-to-face interviews were conducted with NEC professionals in Hong Kong to validate the research findings. A total number of three NEC professionals in Hong Kong were involved to provide expert opinion. All professionals have more than 20 years of professional experience. In particular, two are consultants who have provided contract administration services to the government. They delivered two NEC projects in Hong Kong. These were the pioneer projects in adopting the NEC contract in the public sector. Another professional is a contractor who tendered for NEC projects. He delivered three NEC projects in Hong Kong. Interviewees were asked to comment on risk allocation between parties under the NEC contract.

All interviewees agreed that risk responsibility is fairly allocated under the NEC contract. Traditionally, the Employer allocates excessive project risks to the Contractor with the traditional form of contract. This harms the relationships between parties such that the parties become uncooperative. This situation is improved for NEC projects. During the tendering process, both the Employer and the Contractor are allowed to specify potential risks in

Contract data Part I and Part II respectively. From the Contractor's perspective, compensation events can be stated in the NEC contract with wider coverage in contrast to the one set in the traditional contract. The Contractor's project risk is thus potentially reduced. Nevertheless, the interviewees confirmed that the risk items associated with 16 identified categories of NEC project risks, were discussed one by one during the contract drafting stage and the risk reduction meeting.

All interviewees believed that the NEC contract facilitates proactive risk management for delivering construction projects. In particular, the responsibility of risk management is shared among the parties. Early warning notification set out in the NEC contract forces the parties to play a proactive role in risk management. If the Contractor fails to serve an early warning notification, sanction would be imposed. Presently, the Employer and the Contractor often rely on brainstorming technique to identify project risks. The proposed data-driven approach helps the parties to automate the "risk identification" and "risk quantification and analysis" process. The interviewees emphasised that traditional experience-based brainstorming approach is time-consuming. Around 80% of the meeting time for risk management is consumed in identifying the risk items for past projects. The proposed data-driven approach outperforms the traditional one in terms of reducing effort for collecting data and reducing time to identify the project risk items. Rather than brainstorming the project risk items, around 20% of the meeting time is spent on verifying the identified project risks which are applicable to the current project. In addition, the interviewees emphasised that the decision tree helps planners to manage the project risks during the risk reduction meeting. The decision tree can help to automatically categorise, quantify, and analyse the project risks. As such, the planners shall manage the project risks from the perspective of project risk categories instead of managing the project risk items one by one.

Conclusions

This research study proposes a new data-driven approach for facilitating the "risk identification" and "risk quantification and analysis" process in delivering NEC construction projects. Driven by NEC risk registers, this novel approach identifies common NEC project risk categories using text mining analysis, and quantifies and analyses project risk ratings with regard to risk likelihood and consequence using decision tree analysis.

The results of a practical case study found that the NEC risk categories of ground and utilities, design information, structures, and workmanship are commonly incurred in Hong Kong NEC projects. Risk levels for ground and utilities, design information, structures, and workmanship categories are identified at medium to high, medium, medium, and medium, respectively. The benchmarks are helpful for NEC project planners in taking precautionary measures to mitigate, monitor, and control NEC

project risks. Although this research relates to NEC projects in Hong Kong, the proposed approach will apply to NEC projects in other countries.

The benchmarks were validated by interviewing experienced NEC professionals. The professionals commented that the NEC contract helps to change the relationships between contracting parties from adversarial to cooperative. The proposed data-driven approach assists risk management to manage NEC project risks proactively. This approach allows NEC project planners to automatically categorise common NEC project risks and characterise risk rating of NEC project risk categories with regard to risk likelihood and consequence. As such, the time for identifying, quantifying, and analysing the NEC project risks can be significantly reduced.

As for future studies, the proposed method can be extended to cater for different formats of NEC risk register to benchmark project risks. For example, if there are two or more criteria (apart from risk likelihood and risk consequence) to evaluate risk rating, the proposed method would develop the decision tree with two or more levels. In addition, the NEC risk registers can be analysed on the basis of building trades. For example, NEC project risks associated with electrical and mechanical installation can be identified, quantified, and analysed. This would help electrical and mechanical sub-contractors to manage the project risks and complete the project on time and within budget. Moreover, with the increasing interest in big data analysis, a future study of the development of the text analysis package for the construction industry is suggested. Since technical terms are used in the construction industry, the existing text mining package may fail to mine textual descriptions. Such development will help research communities and project planners to mine big data with less human interference in construction.

References

AbouRizk, S. 2009. *Risk analysis for construction projects: A practical guide for engineers and project managers*. University of Alberta, Edmonton, Alberta, Canada.

Ahmed, S.; Ahmad, R.; Darshi De Saram, D. 1999. Risk management trends in the Hong Kong construction industry: A comparison of contractors and owners perceptions, *Engineering, Construction and Architectural Management* 6(3): 225–234. https://doi.org/10.1108/eb021114

Akintoye, A. S.; MacLeod, M. J. 1997. Risk analysis and management in construction, *International Journal of Project Management* 15(1): 31–38. https://doi.org/10.1016/S0263-7863(96)00035-X

Alhajri, A.; Alshaibani, A. 2018. Causes of construction delay in petrochemical projects in KSA, *Energies* 11(7): 1652. https://doi.org/10.3390/en11071652

Al-Bahar, J.; Crandall, K. 1990. Systematic risk management approach for construction projects, *Journal of Construction Engineering and Management* 116(3): 533–546. https://doi.org/10.1061/(ASCE)0733-9364(1990)116:3(533)

Baccarini, D.; Archer, R. 2001. The risk ranking of projects: A methodology, *International Journal of Project Management* 19(3): 139–145. https://doi.org/10.1016/S0263-7863(99)00074-5

Bing, L.; Tiong, R. L. 1999. Risk management model for international construction joint ventures, *Journal of Construction Engineering and Management* 125(5): 377–384. https://doi.org/10.1061/(ASCE)0733-9364(1999)125:5(377)

Cagliano, A.; Grimaldi, S.; Rafele, C. 2014. Choosing project risk management techniques. A theoretical framework, *Journal of Risk Research* 18(2): 1–17.

Carrillo, P.; Harding, J.; Choudhary, A. 2011. Knowledge discovery from post-project reviews, *Construction Management and Economics* 29(7): 713–723. https://doi.org/10.1080/01446193.2011.588953

Chan, D. W. M.; Chan, A. P. C.; Lam, P. T. I.; Yeung, J. F. Y.; Chan, J. H. L. 2011. Risk ranking and analysis in target cost contracts: Empirical evidence from the construction industry, *International Journal of Project Management* 29(6): 751–763. https://doi.org/10.1016/j.ijproman.2010.08.003

Chapman, C. B.; Cooper, D. F. 1983. Risk engineering: basic controlled interval and memory models, *Journal of the Operational Research Society* 34(1): 51–60. https://doi.org/10.1057/jors.1983.7

Charoenngam, C.; Yeh, C. Y. 1999. Contractual risk and liability sharing in hydropower construction, *International Journal of Project Management* 17(1): 29–37. https://doi.org/10.1016/S0263-7863(97)00064-1

Chang, C. Y. 2014. Principal-agent model of risk allocation in construction contracts and its critique, *Journal of Construction Engineering and Management* 140(1): 04013032. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000779

Cheng, T; Wang, Y.; Sun, Y. 2012. Development and application of tender evaluation decision-making and risk early warning system for water projects based on KDD, *Advances in Engineering Software* 48: 58–69. https://doi.org/10.1016/j.advengsoft.2012.02.003

Choudhry, R. M.; Iqbal, K. 2013. Identification of risk management system in construction industry in Pakistan, *Journal of Management in Engineering* 29(1): 42–49. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000122

Cooper, D. F.; Macdonald, D. H.; Chapman, C. B. 1985. Risk analysis of a construction cost estimate, *International Journal of Project Management* 3(3): 141–149. https://doi.org/10.1016/0263-7863(85)90065-1

Cretu, O.; Stewart, R.; Berends, T. 2011. *Risk management for design and construction*. Hoboken, New Jersey, United States: John Wiley and Sons, Inc. https://doi.org/10.1002/9781118984017

Cagno, E.; Caron, F.; Mancini, M. 2007. A multi-dimensional analysis of major risks in complex projects, *Risk Management* 9(1): 1–18. https://doi.org/10.1057/palgrave.rm.8250014

Ding, L.; Zhong, B.; Wu, S.; Luo, H. 2016. Construction risk knowledge management in BIM using ontology and semantic web technology, *Safety Science* 87: 202–213. https://doi.org/10.1016/j.ssci.2016.04.008

Franke, A. 1987. Risk analysis in project management, *International Journal of Project Management* 5(1): 29–34. https://doi.org/10.1016/0263-7863(87)90007-X

Han, S. H.; Kim, D. Y.; Kim, H.; Jang, W. S. 2008. A web-based integrated system for international project risk management, *Automation in Construction* 17(3): 342–356. https://doi.org/10.1016/j.autcon.2007.05.012

Hayes, R. W.; Science and Engineering Research Council; University of Manchester Institute of Science and Technology; Project Management Group. 1987. *Risk management in engineering construction: Implications for project managers*. London: Thomas Telford.

Hull, J. 1990. Application of risk analysis techniques in proposal assessment, *International Journal of Project Management* 8(3): 152–157. https://doi.org/10.1016/0263-7863(90)90016-5

Hwang, B. G.; Zhao, X.; Toh, L. P. 2014. Risk management in small construction projects in Singapore: status, barriers and impact, *International Journal of Project Management* 32(1): 116–124. https://doi.org/10.1016/j.ijproman.2013.01.007

International Business Machines (IBM). 2011. *CHAID and exhaustive CHAID algorithms*. New Orchard Road, Armonk, North Castle, New York, United States: International Business Machines Corporation.

International Business Machines (IBM). 2018. *IBM SPSS Modeler*. [Computer software]. New Orchard Road, Armonk, North Castle, New York, United States: International Business Machines Corporation.

Institution of Civil Engineers (ICE). 2013. *The New engineering contract*. 3rd ed. London: Thomas Telford.

Jannadi, O. A.; Almishari, S. 2003. Risk assessment in construction, *Journal of Construction Engineering and Management* 129(5): 492–500. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:5(492)

Kangari, R. 1988. Construction risk management, *Civil Engineering Systems* 5(3): 114–120. https://doi.org/10.1080/02630258808970516

Kangari, R.; Riggs, L. 1989. Construction risk assessment by linguistics, *IEEE Transactions on Engineering Management* 36(2): 126–131. https://doi.org/10.1109/17.18829

Kass, G. 1980. An exploratory technique for investigating large quantities of categorical data, *Journal of the Royal Statistical Society* 29(2): 119–127. https://doi.org/10.2307/2986296

Kim, S.; Bajaj, D. 2000. Risk management in construction: an approach for contractors in South Korea, *Cost Engineering* 42(1): 38–44.

Kim, H.; Soibelman, L.; Grobler, F. 2008. Factor selection for delay analysis using knowledge discovery in databases, *Automation in Construction* 17(5): 550–560. https://doi.org/10.1016/j.autcon.2007.10.001

Li, N.; Fang, D.; Sun, Y. 2016. Cognitive psychological approach for risk assessment in construction projects, *Journal of Management in Engineering* 32(2): 4015037. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000397

Loh, W. H.; Ofori, G. 2000. Effect of registration on performance of construction subcontractors in Singapore, *Engineering, Construction and Architectural Management* 7(1): 29–40. https://doi.org/10.1046/j.1365-232x.2000.00130.x

Mahamid, I. 2011. Risk matrix for factors affecting time delay in road construction projects: owners' perspective, *Engineering, Construction and Architectural Management* 18(6): 609–617. https://doi.org/10.1108/09699981111180917

Mahamid, I.; Al-ghonamy, A.; Aichouni, M. 2015. Risk matrix for delay causes in construction projects in Saudi Arabia, *Research Journal of Applied Sciences, Engineering and Technology* 9(8): 665–670. https://doi.org/10.19026/rjaset.9.1452

Mead, P. 2007. Current trends in risk allocation in construction projects and their implications for industry participants, *Construction Law Journal* 23(1): 23–45.

Mooney, C.; Mooney, E. 2014. Risk allocation in construction contracts – Irish public works, *Proceedings of the Institution of Civil Engineers* 167(2): 68–74. https://doi.org/10.1680/mpal.12.00025

Mubin, S.; Mubin, G. 2008. Risk analysis for construction and operation of gas pipeline projects in Pakistan, *Pakistan Journal of Engineering and Applied Sciences* 2: 22–37.

Mustafa, M.; Al-Bahar, J. 1991. Project risk assessment using the analytic hierarchy process, *IEEE Transactions on Engineering Management* 38(1): 46–52. https://doi.org/10.1109/17.65759

Project Management Institute (PMI). 2013. *A guide to the project management body of knowledge* (*PMBOK® guide*). 5th ed. Newtown Square, Pennsylvania, United States: Project Management Institute.

Sarvari, H.; Valipour, A.; Yahaya, N.; Md Noor, N. 2014. Risk identification and assessment in Malaysian public-private partnership projects, in *The Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM)* and *The Sixth International Symposium on Uncertainty, Modeling, and Analysis* (*ISUMA*), 13–16 July 2014, Liverpool, United Kingdom. https://doi.org/10.1061/9780784413609.045

Serpell, A.; Ferrada, X.; Rubio, N. L. 2017. Fostering the effective usage of risk management in construction, *Journal of Civil Engineering and Management* 23(7): 858–867. https://doi.org/10.3846/13923730.2017.1321578

Sharma, S.; Swain, N. 2011. Risk management in construction projects, *Asia Pacific Business Review* 7(3): 107–120. https://doi.org/10.1177/097324701100700310

Tah, J.; Carr, V. 2000. A proposal for construction project risk assessment using fuzzy logic, *Construction Management and Economics* 18(4): 491–500. https://doi.org/10.1080/01446190050024905

Tang, W.; Qiang, M.; Duffield, C.; Young, D. 2007. Risk management in the Chinese construction industry, *Journal of Construction Engineering and Management* 133(12): 944–956. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:12(944)

Tavares, L. V.; Ferreira, J. A. A.; Coelho, J. S. 1998. On the optimal management of project risk, *European Journal of Operational Research* 107(2): 451–469. https://doi.org/10.1016/S0377-2217(97)00344-5

Wang, S.; Dulaimi, M.; Aguria, M. 2004. Risk management framework for construction projects in developing countries, *Construction Management and Economics* 22(3): 237–252. https://doi.org/10.1080/0144619032000124689

Wirba, E.; Tah, J.; Howes, R. 1996. Risk interdependencies and natural language computations, *Engineering, Construction and Architectural Management* 3(4): 251–269. https://doi.org/10.1108/eb021034

Wood, G.; Ellis, R. 2003. Risk management practices of leading UK cost consultants, *Engineering, Construction and Architectural Management* 10(4): 254–262. https://doi.org/10.1108/09699980310489960

Yildiz, A. E.; Dikmen, I.; Birgonul, M. T.; Ercoskun, K.; Alten, S. 2014. A knowledge-based risk mapping tool for cost estimation of international construction projects, *Automation in Construction* 43: 144–155. https://doi.org/10.1016/j.autcon.2014.03.010

Zhao, J.; Xin, C.; Yang, R. 2010. Risk assessment of mine shaft construction projects based on risk matrix, in *3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, 26–28 November 2010, Kunming, China, 4: 527–530. https://doi.org/10.1109/ICIII.2010.607

Appendix A. Consolidated risk register

Table A.1. Consolidated risk register for practical case study

Risk ID	Project ID	Risk item description	Consequence of risk	Likelihood of risk
1	A	Changes of works information during the course of the works	Minor	Possible
2	A	A degree of variation in actual ground conditions including utilities and subsoil information currently available	Moderate	Likely
3	A	Accidents caused by improper design or construction of temporary works design	Major	Unlikely
4	A	Additional watermains to be incorporated under the Contract	Moderate	Very Likely
5	A	Defects due to default of the Contractor	Major	Unlikely
6	A	Unavailability of construction materials	Major	Unlikely
7	A	Contractor unable to complete the works in accordance with the time specified in this contract	Major	Possible
8	A	Unwilling to take over the completed works by maintenance parties	Major	Unlikely
9	A	Alignment of temporary road	Major	Possible
10	A	Alignment of gullies at certain locations	Minor	Possible
11	A	Non-compliance with safety and environmental requirements	Moderate	Unlikely
12	A	Failure to maintain the safety of public adjacent to the working areas	Major	Rare
13	A	Serious traffic congestion caused by temporary road closure and hence unable to continue with the works until a revised TTA is implemented	Minor	Unlikely

Risk ID	Project ID	Risk item description	Consequence of risk	Likelihood of risk
14	A	Damage to existing buildings	Minor	Unlikely
15	A	Delay in obtaining noise permit (CNP) from EPD	Moderate	Possible
16	A	Poorly developed & implemented health & procedures	Minor	Unlikely
17	A	Working in confined space	Major	Likely
18	A	Working in close vicinity to busy roads	Major	Likely
19	A	Injury / fatality to workers / public while construction works at height	Major	Rare
20	A	Inadequate site security	Minor	Rare
21	A	Serious public objection	Minor	Rare
22	A	Conflict in details and programme with interface projects	Minor	Rare
23	A	Unexpected restrictions for working in close vicinity to residential areas	Minor	Unlikely
24	A	Ground settlement and damage of existing structures / road surface	Major	Possible
25	A	Impact on local community and amenity	Minor	Unlikely
26	A	Disputes on the interfacing issues or access arrangement between parties involved	Moderate	Possible
27	A	Unable to obtain the agreement of proposed temporary traffic diversion schemes	Minor	Possible
28	A	Impacts on existing traffic caused by the temporary road closure	Minor	Rare
29	A	Delay in obtaining TTA approval	Moderate	Likely
30	A	Inadequate traffic / pedestrian signage	Minor	Likely
…	…	…	…	…
240	G	Slow progress in construction of works	Serious	Likely

Appendix B. IBM SPSS model for text mining

In IBM SPSS Modeler, data source node and text mining node are connected as shown in Figure B.1. The consolidated risk register is selected as the data source. Both linguistic and frequency techniques are deployed to export textual data and produce category model nuggets. The NEC project risk categories identified are shown in Table B.1.

Figure B.1. IBM SPSS model for text mining

Table B.1. Risk categories identified by IBM SPSS model

Category	Number of descriptors
Interface	3
Excavation	5
Access	8
Progress	4
Delay	5
Safety	5
Damage	4
Environment	13
Road and pedestrian	11
Variation	7
Works area	7
Obstruction	7
Workmanship	9
Structure	9
Design information	18
Ground and utilities	36

Appendix C. Categorised risk register

Table C.1. Categorised risk register for practical case study

Risk ID	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	F	F	F	F	F	F	F	F	T	T	F	F	F	F	F	F
2	T	F	F	F	F	F	F	F	T	T	F	F	F	F	F	F
3	F	F	F	F	F	F	T	F	F	T	F	F	F	F	F	F
4	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F
5	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F
6	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
7	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F
8	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
9	F	T	F	F	F	F	F	F	F	T	F	F	F	F	F	F
10	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F
11	F	F	F	F	F	T	T	F	F	F	F	F	F	F	F	F
12	F	F	F	F	T	F	T	F	F	F	F	F	F	F	F	F
13	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F
14	F	F	F	F	F	F	F	F	F	F	F	F	F	F	T	F
15	F	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F
16	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
17	F	F	F	F	T	F	F	F	F	F	F	F	F	F	F	F
18	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F
19	F	F	F	F	F	F	T	F	F	F	F	F	F	F	F	F
20	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
21	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
22	F	F	F	T	F	F	F	F	F	F	F	F	F	F	F	T
23	F	F	F	F	T	F	F	F	F	F	F	F	F	F	F	F
24	T	T	F	F	F	F	F	F	F	F	F	F	T	F	T	F
25	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
26	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F
27	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F
28	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F
29	F	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F
30	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
240	F	F	F	F	F	F	F	T	F	F	F	F	F	F	F	F

Note: Columns A – Ground and utilities, B – Road and pedestrian, C – Delay, D – Obstruction, E – Works area, F – Environment, G – Safety, H – Progress, I – Variation, J – Design information, K – Workmanship, L – Access, M – Structure, N – Excavation, O – Damage, P – Interface; Cell value T – True, F – False.

Appendix D. IBM SPSS model for decision tree analysis

Using a risk matrix, risk ratings of project risk items can be determined. A column of risk rating for project risk items is augmented to the categorised risk register. Decision tree is then developed for particular risk categories. Figure D.1 shows the IBM SPSS model for automating this process. Firstly, the data type of each column in the categorised risk register should be identified as shown in Table D.1. Type node is placed to characterise the data. Notably, the data type of project risk categories is measured as either "True" or "False". Then, select node is placed to extract relevant risk items for particular project risk categories. For instance, Figure D.2 shows the computer code embedded in the select node for retrieving project risk items of the "Delay" category. The code is written as "Category_delay='T'". Next, an auto-classifier node is placed to generate the bar chart showing the relative importance of risk likelihood and risk consequence towards the risk rating for a particular risk category. CHAID node is placed to generate decision tree using the CHAID algorithm. It is noteworthy that the significance for tree splitting is set as 0.05, the limiting value of maximum tree depth is set as 5, minimum record in the parent branch is set as 2%, and minimum record in the child branch is set as 1%.

Table D.1. Classification of field data types

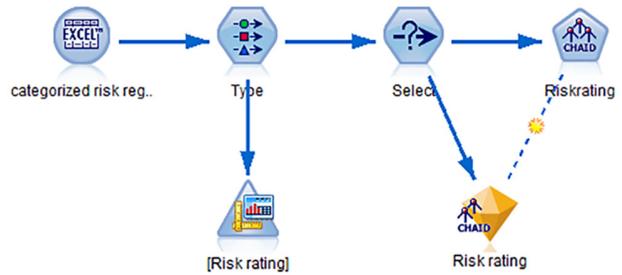| Field	Types of data field	Values
Risk ID	Continuous	1, 2, 3, …, 240
Project ID	Nominal	A, B, C, D, E, F, G
Risk item description	Nominal	<Risk descriptions> e.g. defects due to default of the contractor
Consequence of risk	Nominal	Major, moderate, minor, serious
Likelihood of risk	Nominal	Very likely, likely, possible, unlikely, rare
Risk rating	Nominal	Extreme, high, medium, low
Risk categories	Flag	T/F

Figure D.1. IBM SPSS model for decision tree analysis
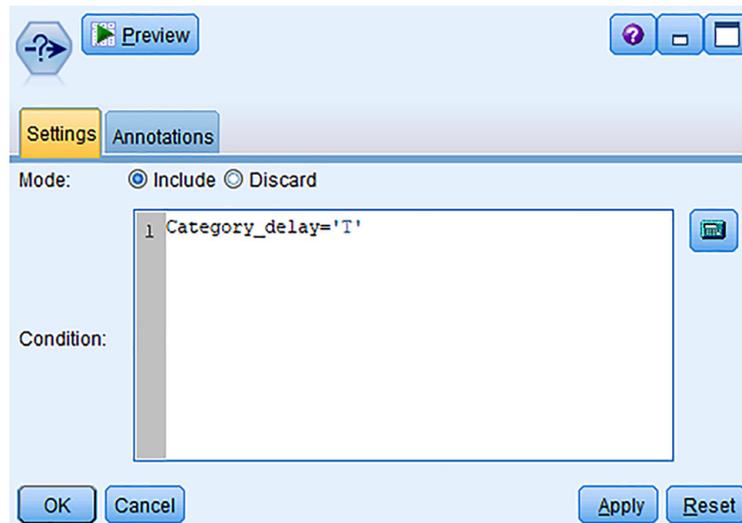
Figure D.2. Computer code embedded in select node