Acta logistica - International Scientific Journal about Logistics
Volume: 10 2023 Issue: 2 Pages: 191-198 ISSN 1339-5629

A new model for cost estimation construction project using Hybrid importance regression ensemble method
Ali Najah Abdulkadhim Alkhuadhan, Sepanta Naimi

https://doi.org/10.22306/al.v10i2.372 *Received: 24 Jan. 2023; Revised: 25 Feb. 2023; Accepted: 13 May 2023*

A new model for cost estimation construction project using Hybrid importance regression ensemble method

Ali Najah Abdulkadhim Alkhuadhan
Altinbaş University, Civil Engineering Department, Dilmenler Caddesi Mahmutbey Yerleşkesi No:26, 34218 Istanbul, Turkey, ali1997najah@gmail.com (corresponding author)
Sepanta Naimi
Altinbaş University, Civil Engineering Department, Dilmenler Caddesi Mahmutbey Yerleşkesi No:26, 34218 Istanbul, Turkey, ORCID 0000-0001-8641-7090, sepanta.naimi@altinbas.edu.tr

Keywords: cost factor, Importance Regression Ensemble Method, K-Nearest Neighbor Method, construction.
Abstract: Cost estimating entails gathering and evaluating historical data, as well as using quantitative models, methodologies, tools, and databases to forecast the cost of a program in the future. At the early stages of the building design process, the cost is considered one of the most important elements in making decisions. During the design phase of a project, cost estimating is quite essential. To compelet a construction project successfully, it is critical to design a usable model and method for cost estimation in construction projects. For the above reason, this study has developed a hybrid method to conduct an accurate cost estimation in construction projects in Iraq. This study also conducted a rigorous survey to find five main influential factors with thirty-six sub-factors in the Iraqi market. It was evaluated through previous studies, questionnaires, and surveys of twenty projects to build a matrix factors database for construction projects. This work gathered the construction cost factors from relevant research and expert views. In the second step, the researcher ranked the factors within the Importance Regression Ensemble Method then the K-Nearest Neighbor Method was applied to specify the effect of the near-effective factors on the cost. The outcome of this study will be helpful to construction professionals in estimating a construction project effectively.

1 Introduction

Estimating building costs is a crucial step in the construction process. The estimation is also very important logistics support of a construction project to meet the demand of the requirment of the project. It represents the total sum needed to cover the expense of everything that must be acquired in order to carry out the project. Preliminary estimates of costs based on conceptual designs are accurate for these two reasons. To begin, the estimates are only as strong as the sloppy, incomplete information used to create them. Second, the building's characteristics and its price tag are more amenable to change early in the design phase than they are later. These considerations motivate studies of development models based on different methods and techniques that might improve conceptual cost estimations [1]. The price of building work must account for direct expenses, indirect costs, and profit. In order to determine how much a job will cost, contractors might utilize unit pricing. The contractor's significant expertise is required for this technique. The unit price must account for all of the aforementioned expenses. It is common practice to calculate direct costs (labor, materials, and equipment), indirect costs (labor, materials, and equipment), and profit margins separately when making an estimate of total costs [2]. According to EL-Sawalhi (2013), the most prevalent types of cost estimations are [3].

Estimate models

In cost estimating, three models are defined based on the level of depth of the output information, time and resource availability, and the stage of the construction life cycle in which they are primarily developed, as shown below [4]:

1. Models based on mathematical underpinnings have unquestionably progressed as computing techniques have improved. In this regard, different estimate models based on probability, fuzzy set theory [5], or regression estimating models exist in the construction sector to assist estimators in selecting the cost model based on historical data for a certain cost estimating application [6].
2. Activity-Based Costing Models: This seems to be the most realistic model since all overhead expenditures can be attributed to the specific activities that need them. To be more specific, the construction sector may benefit from the ABC technique by defining and categorizing all the activities and operations necessary to generate each construction work unit, and then examining, in each instance, all the aspects that impact its cost. When construction firms have this information ahead of time, they are better able to manage their financial investment on the job site, regulate expenses, and do cost-benefit analyses at various stages of project development [7,8].
3. Input-output analysis is the basis of process-based cost (PBC) models. The primary output of the construction business is the physical structure that was built, while the necessary inputs are the materials that went into its creation. PBC models are able to synthesize the parts of

the construction system, therefore they look at the production processes obtained from the scheduling of building works as dynamic cost-generating elements that change resources into outcomes. As a result, building budgets and schedules may be linked together thereafter. These models were made to aid in the estimation process by making tacit knowledge of the building's processes plain to those who acted on the plans [1,9]. Estimating construction costs is both an art and a science. The contractor must be able to comprehend and conceptualize a facility project and design the strategy for constructing the facility to be an accurate estimator. The finest estimators are also skilled at analyzing past building costs and predicting the factors that will increase or decrease future costs [2]. Building estimated costs, or "estimates," are often formed throughout a project. To determine whether a project is economically feasible, the owner may create extremely preliminary feasibility estimates. To ensure the project is designed to the owner's budget, a designer or construction manager may build a sequence of progressively detailed approximations during the design process. A general servicer or trade contractor will decide to generate estimates to limit their bid or low values for a project. And many estimates might be developed industrially to assess the impact of different design decisions or create a cost estimate for a design modification throughout the construction process. Cost projections for construction projects are overly detailed and have varying degrees of accuracy. The RS Means levels of detail can be used as a guide when developing part-level definitions [3].

4. Averaged Square Base Estimate

These approximations are based on average statistical norms for a building's cost per square foot or unit cost. Although this correctness can vary, they are expected to be accurate within +/- 20%. This estimation can be made with limited information, such as the number of carriages for a parking garage as well as an approximation of the square footage for an office building. This calculation can be made during the planning stage and shouldn't take more than ten minutes [4].

5. Demonstrated Square Foot Estimate

These estimations use existing model structures to construct a structure representative of the upcoming building. Although this method is projected to be accurate to within +/- 15%, the accuracy can vary greatly depending on how strong the assumptions are. The Modeled Square Foot approximation method requires the contractor to have a general notion of the structure's footprint, be familiar with the structural system, and be aware of the facade plan. This estimating method typically takes an hour to complete and can be utilized during the Schematic Design stage [10].

6. Assemblies Estimate

A project's meetings are identified, valued, and quantified to create an assembly estimate. They are anticipated to be within 10% of the actual value. The contractor must be able to recognize a system-equal design and perform quantity departures for multiple schemes to estimate a project. Usually, this can happen throughout the project development period. He typically needs one day to create this kind of approximation, spending most of that time conducting quantity takeoffs (quantifying the amount of each item) [6].

7. Unit Value Estimate

In a component price estimate, every item restricted to the project is defined, and the materials are priced after the precise construction techniques that will be utilized to a specified concept. Depending on the intricacy of the project, this method can be accurate to within +/- 5% of the cost. This approximation typically occurs close to or after the construction documents phase since it requires a relatively detailed design. A thorough estimate of every item in a structure can take up to three weeks, with a lot of effort spent on execution quantity takeoffs [6].

Based on a review of the relevant literature, a thorough list of 27-factor indicators was compiled, and a questionnaire was constructed to solicit feedback from Iraqi construction project managers. The factors are outlined in Table 1.

Table 1 The major factors regading the cost estimation in construction project

Main Factor	Sub-Factor	Coding	Main Factor	Sub-Factor	Coding
Labor-related factors (F1)	Frequent exposure of workers to accidents	SF1	Equipment related factors (F4)	Lack of equipment on site	SF24
	Poor use of resources	SF2		Delay in the delivery of equipment	SF25
	Poor use of technology	SF3		Poor equipment productivity	SF26
	Poor understanding of working methods	SF4		Poor ability to operate equipment	SF27
	Low labor productivity	SF5		Frequent equipment malfunctions	SF28
	unskilled labour	SF6		Poor maintenance of equipment	SF29
	Skilled labor shortage	SF7			F5
Material related factors (F2)		F2	Management related factors (F5)	Poor construction planning	SF30
	Poor availability of materials in the market	SF8		weakness in decision making	SF31
	Delayed delivery of materials to the site	SF9		Low management skills	SF32
	weakness in Material quality	SF10		Too much owner interference	SF33
	uprise material prices	SF11		Additional business change of owner	SF34
	poor Material management	SF12		delay in payment	SF35
	Poor storage of materials	SF13		bureaucracy	SF36
Contractor related factors (F3)		F3			
	Poor contractor experience	SF14			
	Contractor productivity is not acceptable	SF15			
	Slow to make a decision	SF16			
	Poor management on site	SF17			
	Contractor's financial problems	SF18			
	Delays/mistakes due to subcontractors	SF19			
	Poor communication and coordination	SF20			

2 Methodology of present work

The primary objective of the research is to discover which components of construction projects are the most crucial to address. A skilled group of specialists must perform the rating. The Delphi technique offers the researcher a flexible and adjustable data collection and evaluation instrument. Some of the reasons why the Delphi approach is utilized are as follows [11,12]: The word "method to evaluation" refers to comprehending each focused component in depth. Input is required to find the most cost-effective approaches for factor management and to evaluate the variables that must be managed. It takes both the causes and effects of the factors into consideration. Combining outcomes and probability while analyzing and investigating causes is a usual practice. Factor assessments could be qualitative, quantitative, or semi-quantitative [13]. When there are disparities in the data quality or the data sources, quantitative analysis is not always the best solution. Under these conditions, it is possible to do a thorough qualitative analysis with the same level of specificity. Regardless of the approach employed, the documentation for the study should include a description of the data quality and the data sources used in the investigation. In addition, you must include a description and explanation of the system and the problem definition, as well as identified factor sources and factors, initiation events, etc. According to the findings of this study, the significance of cost-estimating challenges is evaluated in a practical and effective manner [14,15]. Using the Firefly approach and enhancing it with a scanning methodology, elewe et al. [16,17] built a new set of techniques for managing such complicated research problems, which were considered cutting-edge at the time. Using the MC-GPSO methodology, Talib et al. [18,19] and bin Hasnan et al. [20,21]created a method for tackling huge problems. [22] and [23] individually. In [24], Majeed et al. made an assimilated ANN for defect pattern recognition, which was later reported by [25,26]. A common principle in factor management asserts that the party with the most relevant abilities and qualifications will execute the most effective factor management. As a result, specific responsibilities might be outsourced to a single organization, ensuring that the enterprise is safeguarded or that the repercussions of the relevant conditions are avoided [27]. In the first stage of this work, which consisted of identifying the study area, the cost estimation criteria were determined using the Delphi method based on pertinent literature and expert opinion. In the second stage, the researcher determined the Importance Regression Ensemble Method of factors for each region. In the third step, the K-Nearest Neighbor Method was developed utilizing MATLAB software, which examines the given data to identify and evaluate factors and to establish the factor rating. On the basis of the Importance Regression Ensemble Method and K-Nearest Neighbor Method findings, the influencing variables of building projects were identified and tested.

3 Result and discussion

3.1 Importance regression ensemble results

Importance Regression Ensemble analysis ranks indicators based on participant replies that determine the majority of significant criteria (Figure 1 - Figure 5). The relative relevance of the criterion was determined using a relative index analysis. Following are tables containing the relative index study's ranking results for each region. As a result of these rankings, twenty hazards were identified as having a high level of importance in assessing the implications of cost estimation on building projects.

The Importance Regression Ensemble was constructed to uncover each challenge's of the cost estimation of the construction project. These factors were ranked using the Importance Regression Ensemble values that were produced.

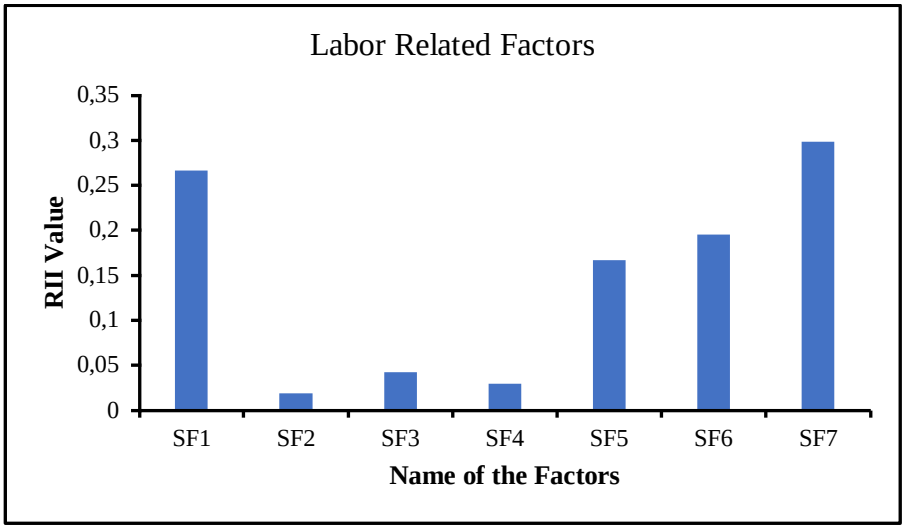

Figure 1 Importance Regression Ensemble of labor related factors respond scoring

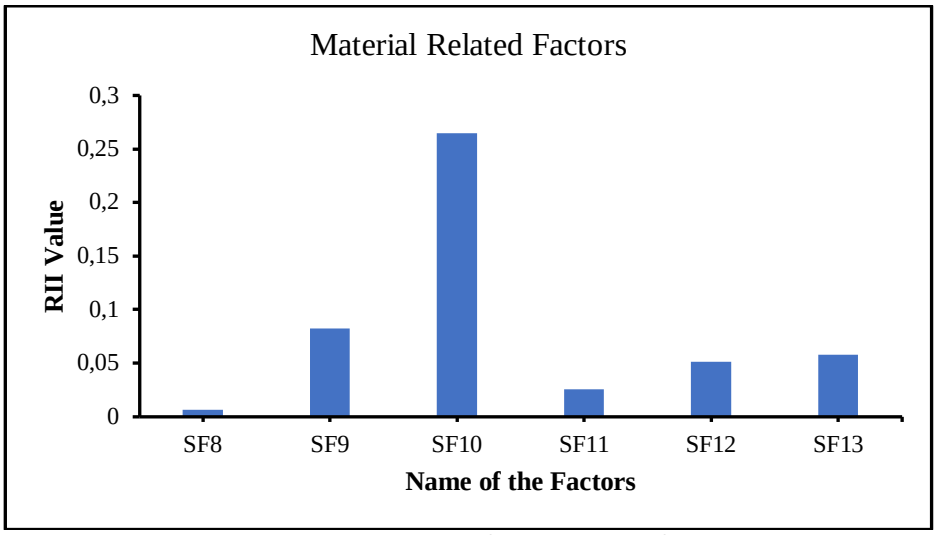

Figure 2 Importance Regression Ensemble of material related factors respond scoring

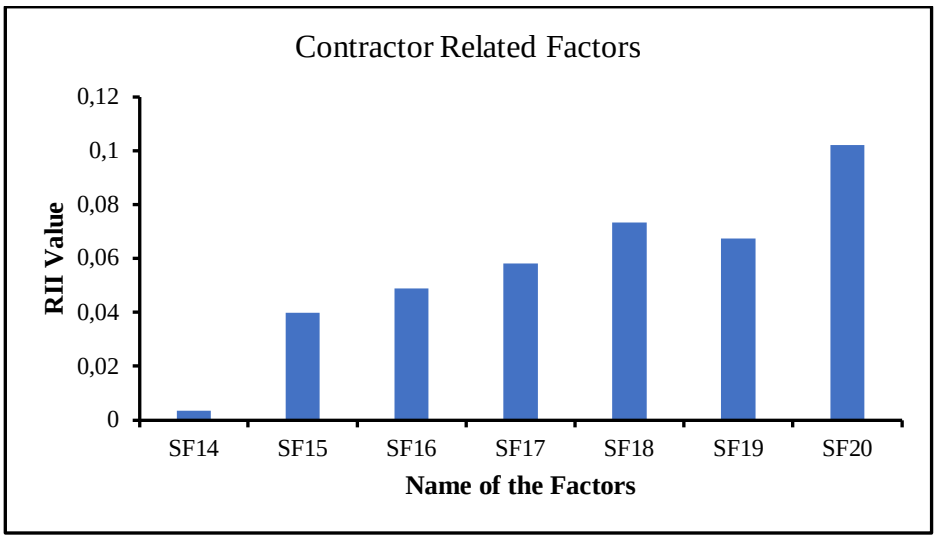

Figure 3 Importance Regression Ensemble of material related factors respond scoring

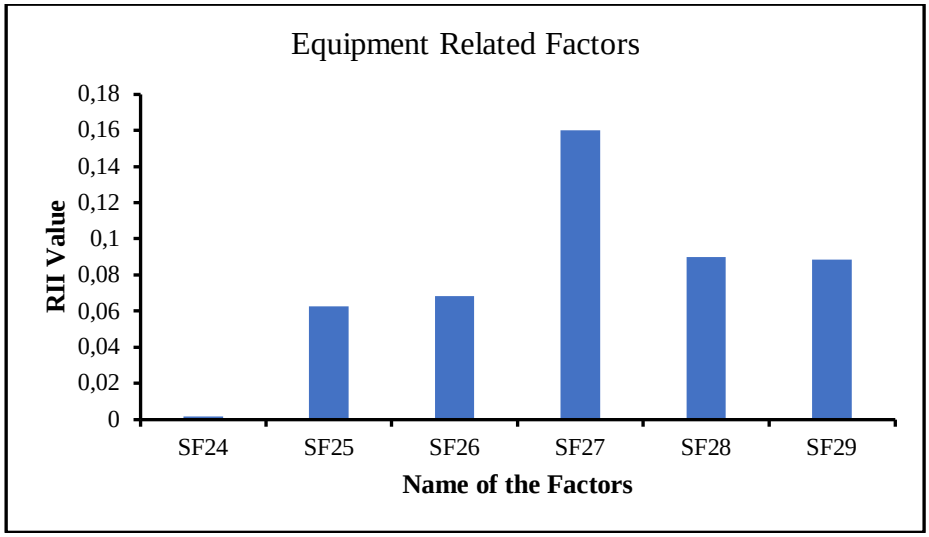

Figure 4 Importance Regression Ensemble of equipment related factors respond scoring

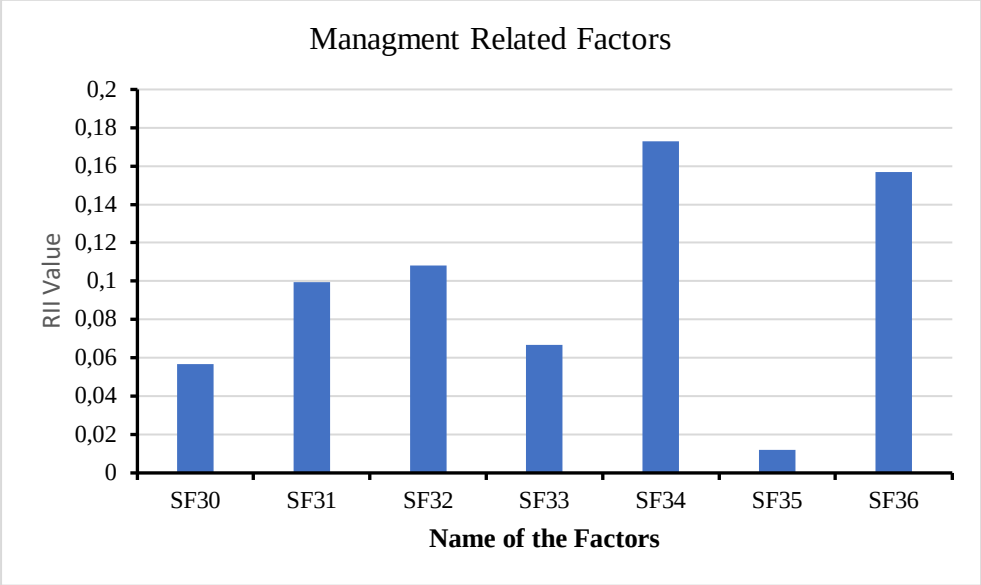

Figure 5 Importance Regression Ensemble of management related factors respond scoring

3.2 K-Nearest Neighbour Method Results

The K-Nearest Neighbour Method analysis results are reported in the following sections (Table 2). The K-Nearest Neighbour Method analysis results for each region are presented in the tables shown in the following sections. Based on these ranking results, it was found that the selected factors had considerable degrees of importance in the factor evaluation of building projects based on the effects of cost estimation. The K-Nearest Neighbour Method was applied for each building cost estimation component to identify factor components. These criteria were ranked based on the K-Nearest Neighbour Method's outcomes. It presents a significant cost estimation risk. In addition, the parent construction trash disposal is a major source of cost estimation complications. According to the research, these components were identified as an intriguing cost estimation contributing factor. Comparing the results that the top higher ranks, as shown in the table, are much higher. According to the experts, the results indicate that the Method has a high degree of reliability.

Table 2 K-Nearest Neighbor Method's outcomes

Project	Actual cost	Estimated cost	Error
building size 200m3	46393	47872	3%
building size 300m4	71373	72783	2%
building size 400m5	119336	103021	14%
building size 500m3	84899	84920	0%
building size 600m4	130613	130846	0%
building size 700m5	218386	220662	1%
building size 800m3	155365	162658	5%
building size 900m4	239023	222370	7%
building size 1000m3	399646	389708	2%

4 Conclusion

This study aims to establish a unique way of estimating project cost concerns in Iraq by creating a model that can aid parties involved in construction projects to recognize obstacles and factors in advance. These strategies and procedures were employed to achieve this objective. The current study created a hybrid strategy for identifying precise cost estimations in Iraqi construction projects. The study identifies five primary effective elements with thirty-six sub-factors in the Iraqi market. It evaluates them using prior studies, questionnaires, and a survey of twenty building projects to create a database of matrix factors for construction projects. This study aimed to determine the cost components based on relevant research and expert opinions. In the second phase, the researcher ranked the factors using the Importance Regression Ensemble Method and then utilized the K-Nearest Neighbor Method to determine the effect of the near-effective factors on the cost.

References

[1] HILDEBRAND, J.P.: *Estimating the life cycle cost of microgrid resilience*, Naval Postgraduate School Monterey Ca, 2020. https://apps.dtic.mil/sti/citations/AD1126927

[2] MISLICK, G.K., NUSSBAUM, D.A.: *Cost estimation: Methods and tools*, John Wiley & Sons, 2015. https://www.wiley.com/en-sg/Cost+Estimation:+Methods+and+Tools-p-9781118536131

[3] PEŠKO, I., MUČENSKI, V., ŠEŠLIJA, M., RADOVIĆ, N., VUJKOV, A., BIBIĆ, D., KRKLJEŠ, M.: Estimation of costs and durations of construction of urban roads using ANN and SVM, *Complexity*, Vol. 2017, pp. 1-13, 2017. https://doi.org/10.1155/2017/2450370

[4] XU, S., LIU, K., TANG, L.C.: *Cost estimation in building information models*, In: ICCREM 2013: Construction and Operation in the Context of Sustainability, 2013, pp. 555-566, 2013.

[5] LAKSONO, T.D.: The Analysis of the Factors Which Affect the Building Costs in Indonesia, *Civil Engineering Journal*, Vol. 3, No. 1, pp. 57-62, 2017. https://doi.org/10.28991/cej-2017-00000072

[6] GUREEV, K., AFTAKHOVA, U., GLADKIKH V.: *Tools development for estimated cost management of construction and assembly works*, In: IOP Conference Series: Materials Science and Engineering, Vol. 687, No. 4, pp. 44-55, 2019. https://doi.org/10.1088/1757-899X/687/4/044003

[7] QUESADO, P., SILVA, R.: Activity-based costing (ABC) and its implication for open innovation, *Journal of Open Innovation Technology, Market, and Complexity*, Vol. 7, No. 1, pp. 41-49, 2021. https://doi.org/10.3390/joitmc7010041

[8] BRICIU, S., CAPUSNEANU, S.: Effective cost analysis tools of the activity-based costing (ABC) method, *Annales Universitatis Apulensis: Series Oeconomica*, Vol. 12, No. 1, pp. 25-35, 2010.

[9] MONTES, M.V., FALCÓN, R.M, RAMÍREZ-DE-ARELLANO, A.: Estimating building construction costs by production processes, *The Open Construction & Building Technology Journal*, Vol. 8, No. 1, pp. 171-181, 2014.

[10] HOEK, E., PALMIERI, A.: *Geotechnical risks on large civil engineering projects*, In: Proceedings of the 8th Congres IAEG, 1998: International Association of Engineering Geologists Congress Vancouver, Canada, pp. 79-88, 1998.

[11] BLACK, M.: *Crossrail project: managing geotechnical risk on London's Elizabeth line*, In: Proceedings of the Institution of Civil Engineers-Civil Engineering, Vol. 170, No. 5, pp. 23-30, 2017.

[12] CORREIA, A.G., WINTER, M.G., PUPPALA, A.J: A review of sustainable approaches in transport infrastructure geotechnics, *Transportation Geotechnics*, Vol. 7, pp. 21-28, 2016. https://doi.org/10.1016/j.trgeo.2016.03.003

[13] YADOLLAHI, M., MIRGHASEMI, M., MOHAMAD ZIN, R., SINGH, B.: Architect critical challenges as a project manager in construction projects: a case study, *Advances in civil engineering*, Vol. 2014, Article ID 205310, 2014. https://doi.org/10.1155/2014/205310

[14] DE RISI, R., DE LUCA, KWON, F., SEXTOS O.S.: Scenario-based seismic risk assessment for buried transmission gas pipelines at regional scale, *Journal of Pipeline Systems Engineering and Practice*, Vol. 9, No. 4, pp. 1-12, 2018. https://doi.org/10.1061/%28ASCE%29PS.1949-1204.0000330

[15] NAIMI, S. CELIKAG, M.: *Problems of Reinforced Concrete Building Construction in North Cyprus*, In: 12th International Conference on Inspection, Appraisal, Repairs & Maintenance of Structures, pp. 1-8, 2010.

[16] NAIMI, S., KARIMI, M.A.: Pavement Management System Investigation in Case of Afghanistan, *Cumhuriyet Science Journal*, Vol. 40, No. 1, pp. 221-232, 2019. https://doi.org/10.17776/csj.471334

[17] NAİMİ, S., HRİZİ, H.: Risk Analysis of Slaving Floor in Construction Sites, *International Journal of Electronics, Mechanical and Mechatronics Engineering*, Vol. 9, No. 1, pp. 1637-1645, 2019.

[18] SISSAKIAN, V.K., ABDUL AHAD, A.D., HAMID, A.: Geological hazards in Iraq, classification and geographical distribution, *Iraqi Bulletin of Geology and mining*, Vol. 7, No. 1, pp. 1-28, 2011.

[19] LAM, T.Y., SIWINGWA, N.: Risk management and contingency sum of construction projects, *Journal of Financial Management of Property and Construction*, Vol. 22, pp. 237-251, 2017.

[20] HONG, Z.S., YIN, J., CUI, Y.J.: Compression behaviour of reconstituted soils at high initial water contents, *Géotechnique*, Vol. 60, No. 9, pp. 691-700, 2010. https://doi.org/10.1680/geot.09.P.059

[21] WEAK, B.: *Foundations on Weak and/or Compressible Soils*, 53:139 Foundation Engineering, The University of Iowa, pp. 1-5, 2007.

[22] OYEDIRAN, I., DUROJAIYE, H.:Variability in the geotechnical properties of some residual clay soils from southwestern Nigeria, *International Journal of Scientific & Engineering Research*, Vol. 2, No. 9, pp. 1-6, 2011. https://www.ijser.org/viewPaperDetail.aspx?SEP1140

[23] DE LA CRUZ LÓPEZ, M.P., CARTELLE BARROS, J.J., DEL CAÑO GOCHI, A., LARA COIRA, M.: New Approach for Managing Sustainability in Projects, *Sustainability*, Vol. 13, No. 13, pp. 1-27, 2021. https://doi.org/10.3390/su13137037

[24] SUNKE, N., SCHULTMANN, F.: Requirements for sustainable construction materials and components,

Lifecycle Design of Buildings, Systems and Materials, Vol. 24, No. 2, pp. 78-87 2019.

[25] ELEWE. A.M., HASNAN, K.B., NAWAWI, A.B.: Hybridized firefly algorithm for multi-objective radio frequency identification (RFID) network planning, *ARPN J Eng Appl Sci*, Vol. 12, No. 3, pp. 834-840, 2017.

[26] ABDULLAH, H., TALIB, N., ELEWE, TAHI, A.S.: *An efficient algorithm for largscale RFID network planning*, No. September, 2019, IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019.

[27] BIN HASNAN, K., ELEWE, A.M., BIN NAWAWI, A., TAHIR, S.: *Comparative evaluation of firefly algorithm and MC-GPSO for optimal RFID network planning*, In: 2017 8th international conference on information technology (ICIT), IEEE, pp. 70-74, 2017. https://doi.org/10.1109/ICITECH.2017.8079927

Review process
Single-blind peer review process.