ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2222-7059 (Print)
ISSN: 2222-7067 (Online)

Advances in Industrial Engineering and Management (AIEM)

DOI: http://doi.org/10.7508 /aiem.01.2024.45.54

AIEM ADVANCES IN INDUSTRIAL ENGINEERING AND MANAGEMENT

CrossMark

RESEARCH ARTICLE

A SYSTEMATIC LITERATURE REVIEW OF BIG DATA ANALYTICS TECHNIQUES APPLICATIONS IN CONSTRUCTION COST FORECASTING

Jesse Amadosi Emmanuel[a*], Olajide Olamilokun[b]

[a]Department of Quantity Surveying, Baze University Abuja.
[b]Department of Quantity Surveying, Kaduna State University, Kaduna.
**Corresponding Authors Email: talk2amadosi@yahoo.com*

ARTICLE DETAILS

Article History:

Received 23 August 2023
Revised 26 September 2023
Accepted 19 October 2023
Available online 25 October 2023

ABSTRACT

The capacity of the construction industry to generates massive data – popularly termed big data – has become a mainstream subject. Consequently, the implementation of modern analytical techniques on these big data is envisioned to improve the accuracy of cost forecasted at the project conceptual phase. Yet, with growing trend in analytical techniques, cost forecasted for most project is still fraught with inaccuracies. To address this problem, this study examines how the adoption of big data analytics techniques can be leveraged upon to improve the accuracy of cost forecasted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol. From a total of 205 articles retrieved from EBSCOhost, Google Scholar, Science Direct, and Dimension databases and ASCE library, 79 articles published between 1976 and 2023 were considered for the study. From the analysis of the extant literature, four main components are outlined as results; first the review establish that predictive analytics is predominantly employed to forecast cost among other approaches like descriptive and prescriptive analytics. Secondly, the nature of cost drivers employed for forecasting raises concern on the need to combine unstructured and structured data to forecast cost. Thirdly, with regards to methodology adopted for cost estimation, the result of the analysis indicate that the quantitative methodology has been applied majorly to forecast cost in light of the intervention of modern analytical tools. Lastly, the review also highlights where data management comes in within the forecasting process harmonised from literature. The review ends with the recommendation that both structured and unstructured data be effectively managed on the basis of completed projects using robust analytical techniques.

KEYWORDS

Big data analytics, techniques, cost forecasting, accuracy, data management

1. INTRODUCTION

Data analytics and cost forecasting are two aspects of artificial intelligence and cost management that can be conflated to position the construction industry for improved service delivery. These two phenomena are considered mainstream development in the construction industry currently and hold strong potential for its future (Ashworth & Perera, 2015). To highlight the importance of data analytics in construction management, studies like that of (Bilal et al., 2016; Elfaki et al., 2014; Ram et al., 2019) underscore the applicability of data mining and analytics to identify inter alia the causes of cost overruns, improve forecast accuracy. Moreso, forecasting cost for construction is one of the most important aspects when dealing with client expenditure management (Adafin *et al.*, 2017). Forecasting in the industry became necessary due to the ample time lag between conception of a need to build and realisation of a facility. The outcome of any forecasting procedure is influenced by many dynamics (known and unknown); however, the goal is to generate a forecast with an acceptable range of accuracy (Rayyan, 2017).

The primary goal of forecasting is to provide a rationale for keeping project cost within reasonable limit and avoid cost overrun (Ashworth & Perera, 2015). In line with the Royal Institute of British Architect (RIBA) plan of work (RIBA) published in 1964, the Royal Institute of Chartered Surveyor's (RICS) New Rules of Measurement (NRM)1 provide a framework for establishing a realistic cost limit for a proposed project at the different stages of design evolution. At the early stage, the aim is to create a building design within the scope defined by the owner's requirements and within the cost target defined in the earlier stages. This goal makes cost forecasting a tool of control for the design in terms of cost.

In construction, a known measure of accuracy is the deviation from the acceptable tender figure obtained from a tenderer bidding for project. The concept of accurate estimate has been a subject of discourse over the last 50 years. For instance, following the evaluation of some civil engineering bills of quantities, Barnes (1970) reported a wide variation in the forecast of contract or net figures. The study by Morrison, (1984) provide a sound reference on the concept of accuracy. With reference to these early studies, one would expect the accuracy of cost forecasting to have improved given advances in science and technology. However, as studies show, the construction industry is still grappling with inaccuracies in cost forecast (Abdel-Monem et al., 2022; Arif et al., 2015; Awosina et al., 2006; Hatamleh et al., 2018).

Given the increasing volume of data generated at a high velocity from various sources, big data analytics has emerged to support the construction industry with the capacity to make evidenced based decisions. With big data, insight can be gained more easily, quickly and more clearly (Ngo et al., 2020). Bilal et al., (2016) asserts that the construction industry's capacity to generate large-scale data is not in doubt. With robust analytics, big data conducts advanced analytics such as

Quick Response Code	Access this article online	
	Website: www.aiem.com.my	**DOI:** 10.26480/aiem.01.2024.45.54

Cite The Article: Jesse Amadosi Emmanuel, Olajide Olamilokun (2024). A Systematic Literature Review of Big Data Analytics Techniques Applications in Construction Cost Forecasting. *Advances in Industrial Engineering and Management, 13(1): 45-54.*

predictive analytics which are capable of identifying hidden patterns and trends to accurately forecast large scale projects (Sivarajah et al., 2017). Their strength is in the ability to deal with different number of possible outcomes by swiftly modelling several situations, working with premature data and drawing inferences based on the learning acquired from previous project data. Big data analytics possess the ability to develop construction-specific benchmarks to reduce the incidences of optimism bias and strategic misrepresentation which are inherent with human role.

An appreciable number of research have been conducted within construction cost forecasting domain. However, as systematic reviews are increasingly gaining acceptance as a starting point in the development of evidence-based knowledge (Siddaway et al., 2019), only a corresponding sparse number have been devoted to accuracy of construction cost forecast. A review of eight construction cost estimation studies highlights three concerns where research is yet to address: 1) Effective means of data selection and manipulation to improve accuracy of forecast; 2) lack of consensus on the tools to that best address cost forecasting; 3) Over reliance on model development and non-consideration of the series of events that need to be updated so as to stem the tide of cost overrun.

In the last 10 years, only a limited number of articles are found to have carried out systematic review on construction cost estimation. Of this number, 2 focused on transport infrastructure (Barakchi et al., 2017; Membah & Asa, 2015) while the other 3 focused on construction project. The review by Tayefeh-Hashemi et al., (2020) studied mainly machine learning techniques employed in cost prediction. Dosumu et al., (2022) in their study, developed a framework aimed at evaluating effective estimation processes using Nigeria as a case study. Fazil et al., (2021) in their bid to advance knowledge on cost estimation performance, addressed the nature of factors influencing cost estimation and how these factors have exerted their influence. Elfaki et al., (2014) and (Elmousalami, 2020)'s review was conducted with a view to providing guidelines to improve the cost estimating implementation and the modelling process from a research perspective. Despite these efforts, practical implications remain elusive for practice; for instance, the demonstration of interest in a systematic forecasting process and emphasis on skilled professional to handle forecasting task.

According to Jahan et al. (2016), different tools to help develop a research question exist, depending on the type of question. Among the ones identified, the population, intervention, comparator, and outcomes (PICO); seem to align with this review subjects (Jahan et al., 2016). Where *construction projects* represent the sample Population, *analytical tools* denote the Intervention, *data* signify the Comparator and *forecasted cost* represent the Outcome. As a way of making sense of the available vast scholarly information and highlighting gaps between what is known and what needs to be known, this systematic review seeks to address the following question with respect to construction cost forecasting:

1. What are the different types of big data analytics approaches theorized/employed for construction cost forecasting?
2. What is the nature of data used for forecasting?
3. What methodologies have been adopted to forecast cost?
4. Where does data management fit in within the forecasting process?

1.1 Related Previous Studies

In recent years, the intersection of Big Data analytics and construction cost forecasting has emerged as a dynamic and transformative field. A systematic literature review reveals a diverse landscape of innovative techniques applied to harness the potential of vast datasets in predicting and managing construction costs. For example, the study focuses on using deep learning and BIM properties to predict construction costs in the schematic design phase by (Park & Yun, 2023). Another study involves a parametric method and BIM-based cost estimation model for predicting construction costs by (Yang et al., 2022). Yun (2022) analyses the performance of neural networks in construction cost prediction using multiple input variables. Ye (2021) presented an algorithm that combines particle swarm optimization and neural networks for cost forecasting. The remaining articles also explore different approaches such as artificial neural networks and genetic algorithms in construction cost prediction. Overall, these studies aim to develop accurate and efficient models for estimating construction costs based on various input variables and techniques (Ye, 2021).

The literature also underscores the practical applications of these Big Data analytics techniques in real-world construction scenarios (Zhang & Fang, 2019). The author focuses on feature extraction as a novel method for predicting costs. They consider factors such as city level, ground and underground floors, building acreage, seismic level, and various content units like concrete and electrical wiring. They employ techniques like random forest, support vector regression, and PCA. Similarly, Chandanshive & Kambekar (2019) explores the use of artificial neural networks for estimating construction costs. Their factors include ground and typical floor area, number of floors, parking area, elevator quantity, and the number of householders. Other studies which deploy current models and techniques such as genetic algorithm and least squares support vector machine, multilayer perceptron (MLP) and radial basis function (RBF) for construction cost estimation include (Bala et al., 2014; Bayram & Al-Jibouri, 2016; Elbeltagi et al., 2014; Xu et al., 2015). Notably, some papers propose hybrid techniques that combine multiple methods. This clearly indicates a growing interest in developing more sophisticated and accurate models in this field.

2. METHODOLOGY

There exist a number of guidelines outlining how to report systematic reviews and meta-analyses. For instance, the approach suggested by (Kitchenham, 2004) put forward a review process consisting 10 steps classified into three main stages namely; planning, conducting, and reporting the review. However, for this study, the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) checklist is adopted. Although the PRISMA was designed mainly for studies that evaluate the effects of health interventions, Moher et al., (2009) argues that its check lists items are applicable to other areas. The evidence that support its wide applicability across different research areas also indicate that it has gained inroad into construction related studies as demonstrated in the review by (Dosumu et al., 2022).

To develop a basis for research articles eligibility, three pre-defined criteria were set. First, the article has to be published in construction related journal. Consequently, articles which focus on cost estimation in other fields like nuclear, space, process or the manufacturing industry were not included. Secondly, only papers whose article were published in English language and are available in full-length for download is considered. In other words, publications, such as trade/labour union newsletters, research notes, editors' comments, readers' comment, and book reviews were excluded. The last criterion applied only to Scopus database; the selected articles were only limited to document type labelled (AR) meaning articles in final print. On the other hand, "Articles-in-Press (AiP), i.e., pre-published versions of accepted articles were not considered. The reason for this is because AiP do not contain cited references which this review considers a vital source of information too. However, alerts were activated to receive notification once an AiP is published as an article. Specifically, the alerts that has to do with "DOCTYPE(AR) [article]" was activated.

2.1 Information Sources and search strategy

Before search commenced fully, effort was made at identifying relevant databases and academic libraries; the result identified a total of 5 information sources; three conventional databases, one AI-driven database and one academic library. The conventional databases include Google Scholar (GS), EBSCOhost and Elsevier's Science direct. The American Society of Civil Engineers (ASCE) library and DIMENSIONS database were the two others resources used. An initial search revealed that studies use terminologies like estimation, prediction, forecasting to convey the process of providing an estimate of probable cost of construction which in this case is represented by the term forecast. In determining keywords, two sets of search configuration which include the following terms was used alongside the common Boolean operators. The first include the terms; ("Big data" OR "Big data analytics" OR "Big data analytics techniques" AND "construction industry cost forecasting" OR "construction cost prediction" OR "construction cost estimation") whereas the second comprise of ("Machine learning" AND "construction cost forecasting" OR "construction cost prediction" OR construction cost estimation"). To optimise search result, the peculiarities of different search engine was taken into consideration such that the positioning of search keywords were altered in the way the search engine could better interpret it to enhance its result. A system used to track strategy and result is shown in table 1. The "Advanced Search" option was engaged in almost all the databases. As indicated in table 1, the search process in most cases entail 2 – 3 search steps before satisfactory outcome are achieved. To avoid the generation of irrelevant papers, achieve a comprehensive quick enquiry, search keywords were delimited to the "Title" only search option; this was carried out in the ASCE library. On the science direct database, the query string "TITLE ABS KEY" which denote title, abstract and keywords was used. Since the selected databases serves as repository for literature cutting across difference disciplines and geographical locations, it is envisaged that the process in its entirety will have avoided any bias geographically.

The selection of primary studies was conducted in two stages, first, through a careful analysis of the titles and abstracts. This was carried out as result were filtered on the individual databases, the selected articles were exported to Microsoft excel document after this exercise. The second selection was made by fully reviewing the studies. In the first filter, papers were excluded when their characteristics were clearly against the selection criteria. In the second filter, a study was selected only when it fulfilled all the selection criteria. The need to set eligibility criteria guided the process of screening to select the relevant papers. The preselection narrowed the list of papers from 205 down to 84. Conducting another round of full review cut down the numbers to 68 papers. A backward and forward snowballing process was performed on the 68 articles following the suggestions by (Wohlin, 2014) . With this process, 11 additional studies were identified, finalising with 79 papers in total.

Table 1: Search strategy, process and result collection

Database	Keyword search	SR	ER
EBSCOhost	“Big data” OR “Big data analytics” OR “Big data analytics techniques” AND “construction industry cost forecasting” OR “construction cost prediction” OR construction cost estimation” “Machine learning” AND “construction cost forecasting” OR “construction cost prediction” OR construction cost estimation”	33	16
Google Scholar	**STEP 1**- DB1“Construction cost forecasting” OR “construction cost prediction” DB2 "big data" OR "big data analytics" DB3 construction industry = **109,000 results** *(Too large)* **Step 2** - DB1“Construction cost forecasting” OR “construction cost prediction” DB2 "big data" OR "big data analytics" DB3 construction industry DB4 without words “firm failure = **67,600 results** *(Too large)* **Step 3** - DB1“Construction cost forecasting” OR “construction cost prediction” DB2 "big data" OR "big data analytics" DB3 “construction industry” = **65 result** *(Outcome Okay)*	65	29
Science Direct	**STEP 1** - “Big data” OR “Big data analytics” OR “Big data analytics techniques” AND “construction industry cost forecasting” OR “construction cost prediction” OR construction cost estimation” = **7 results** *(Not satisfactory)* **STEP 2** – “Big data” AND “construction industry cost forecasting” OR “construction cost prediction” OR construction cost estimation” = 11 **results** *(Not satisfactory)* **STEP 3** – “Machine learning” AND “construction cost forecasting” OR “construction cost prediction” OR construction cost estimation” **= 19 result** *(Okay)*	19	15
ASCE library	*DB1 (Title)* – “Big data” OR “Big data analytics” OR “Big data analytics techniques” AND “construction industry cost forecasting” OR “construction cost prediction” OR construction cost estimation” *DB2 (Title*) “Machine learning” AND “construction cost forecasting” OR “construction cost prediction” OR construction cost estimation”	63	19
Dimensions Database	***STEP 1 – B***ig data” AND “Construction cost forecasting” (*Full data* search) =**15 result *(Not satisfactory)*** ***STEP 2-*** “Construction cost forecasting” AND “big data analytics” (*Full data* search) =**7 result *(Not satisfactory)*** ***STEP 3***“Construction cost prediction” AND “big data analytics” (*Full data* search) =**24 result** *(Okay)*	25	5
	Total	**205**	**84**

2.2 Study and Data Extraction

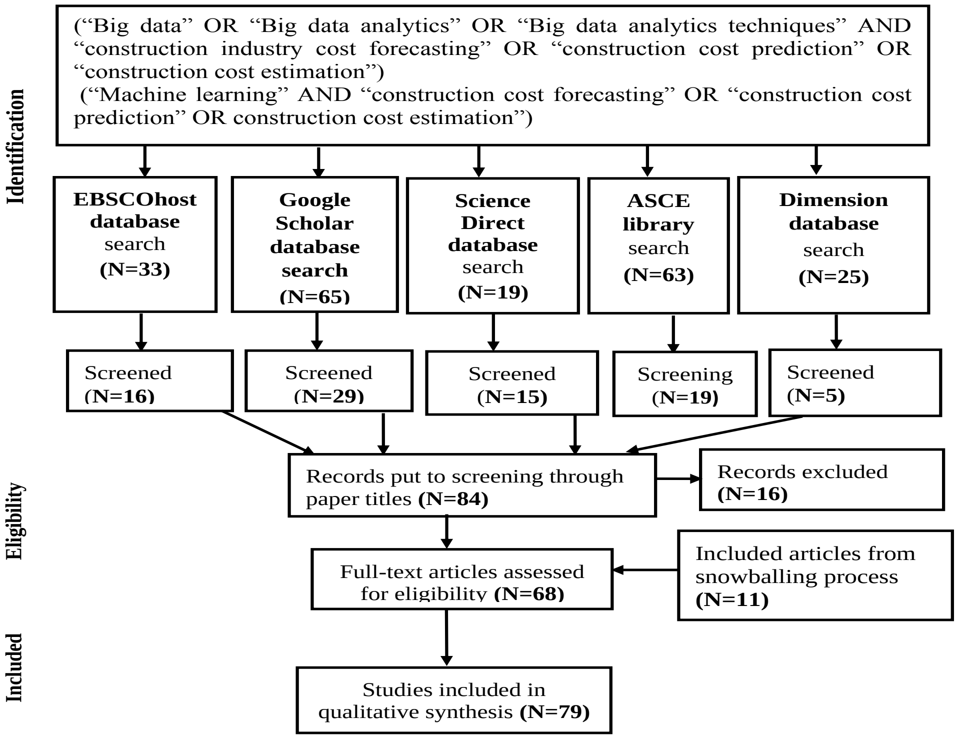

Figure 1: Flow chart illustrating the study selection process

This section describes in details the selection process leading to the extraction of data from one of the information sources; American Society of Civil Engineers (ASCE) in this case. Beginning with the "Advanced Search" option, the search terms which was broken into two was typed into separate query box; the first containing ("Big data" OR Big data analytics" OR Big data analytics technics"; and the second query box containing ("Machine learning" AND "construction cost forecasting" OR "construction cost prediction" OR "construction cost estimation" using the "Title" option as the focus for the search for each query box. Once this was executed, the filter generated 63 results. However, owing to earlier set criteria which exclude grey materials like book chapters and proceeding papers, a refined process filtered 22 research articles termed "technical papers". By publication data, the article spans between 1962 and 2023. The list of papers resulting from the search spread across the following publication title: Journal of Construction Engineering and Management (17); Journal of Infrastructure Systems (2); Journal of Computing in Civil Engineering (1); Journal of Management in Engineering (1) and Journal of Pipeline Systems Engineering and Practice (1).

Information extracted from the 79 articles reviewed were summarised into a Microsoft Excel (MS Excel) literature matrix under the following heading: This is with a view to simplify the synthesis process.

3. RESULTS AND DISCUSSION

This section captures and present a synthesis and discussion of data extracted from the 79 studies used for this systematic literature review. The first subsection provides an overview of the bibliographical features of the publications followed by the result of investigation under the four research questions set out in the study.

3.1 Studies Description

From the 79 studies selected for this review, 7 papers (8.8%) containing the term big data in their titles were considered relevant as they are directly linked with cost forecasting subject, the bulk of the studies almost 90% related to forecasting as implemented through different methods, approaches and techniques. The journal with the highest number of publications is attributable to *Journal of Construction Engineering and Management* with 14 studies followed by *Automation in Construction* which has 7 studies. Going by time horizon, the studies dates from 1970 to 2022, the 1970 studies is identified as the thorough evaluation of building contractors' estimate accuracy carried out by (Barnes, 1970). This perhaps give a sense of the earliest usage of the term "estimate accuracy" with regards to construction cost management before researcher like (Bowen & Edwards, 1985; McCaffer, 1976; Morrison, 1984). Besides the basic information relating to paper title, authors and year, the selected 79 articles were then profiled according to following sub-categories: Research problem identified by each study, theme – cost forecasting theme suited, Input variable employed for forecasting, Methodology adopted for forecasting, Technique; forecasting technique used and Journal title (See table 2). The choice of articles and technical papers selected were based on what researchers have done with respect to big data analytics as applicable to forecasting cost, estimate's accuracy, under/over-estimation. Table 2 presents excerpt of the literature and the different aspect the analysis is based on.

Table 2: Outline of the literature review matrix

	Title	Author	Year	Problem	Theme	input variable	Methodology	Technique	Journal title
1	The Case-Based Reasoning Model of Cost Estimation at The Preliminary Stage of a Construction Project	Krzysztof Zima	2015	Absence of a knowledge-based system for cost estimation at early stage of construction	AI	Elemental unit prices listed in tendered bids.	Quantitative	Mono-technique	Procedia Engineering
2	Model for Forecasting Highway Construction Cost	Zohar Herbsman	1986	Lack of suitable tool to evaluate future road cost	Parametric	Market inflationary trends, bidding volume	Quantitative	Mono-technique	Transportation research record
3	Predicting Construction Cost Using Multiple Regression Techniques	David J Lowe, Margaret W Emsley and Anthony Harding	2006	Absence of time-tested methodology to predict cost of UK buildings	Parametric	41 variables classified as Project strategic (5), site related (4), design related (32),	Quantitative	Mono-technique Multiple Regression Analysis	Journal of Construction Engineering and Management
4	Predicting accuracy of early cost estimates using factor analysis and multivariate regression	Steven M. Trost and Garold D. Oberlender	2003	Non-existent methodology that optimally assess factor affecting the estimate accuracy	Parametric	estimated cost, actual cost plus 45 variables	Quantitative	Hybrid-technique Factor analysis and multivariate regression analysis.	Journal of Construction Engineering and Management
5	Project cost prediction model using principal component regression for public building projects in Nigeria	B.O. Ganiyu, and I.K. Zubairu	2010	Rarity of factors influencing cost overrun in public projects	Parametric	NA	Quantitative	Mono-technique_ PCA	Journal of Building Performance

3.2 What are the different types of big data analytics approaches theorized/employed for construction cost forecasting?

The indication that big data analytics has a rich application for forecasting cost is demonstrated in the following studies; driven by the need for an early and accurate cost advice, Lowe et al., (2006) predicted construction cost using multiple regression techniques. By means of front-end prediction and back-end correction, Wang et al., (2023) adopts regression analysis with ANN to predict cost of building project following the identification of significant cost influencing variables. Compelled by non-existent methodology which optimally assess factor affecting the estimate accuracy, Trost & Oberlender (2003) conducted a study to forecast the accuracy of early cost estimates using factor analysis and multivariate regression. Bilal et al., (2016b) describes regression as a technique which is concerned with predicting the numerical value of a target variable based on input variables Regression analysis have served a unique purpose especially in modelling the relationship between a dependent variable y and one explanatory variable x.

However, as Trost and Oberlender noted, multivariate regression can be of no use in a large and complex data structure and dataset with reasonable multicollinearity. This likely explains why author like Wang et al. earlier cited implemented regression alongside the neural network. Ganiyu & Zubairu, (2010) argue that developing predictive cost model for public building projects using principal components regression is necessary for the purposes of reducing large number of variables required for the estimation. With observed limitation associated with regression technique, artificial neural network (ANN) was introduced in cost forecasting as a means of managing large datasets and tackling restrictions imposed by non-linear relationships and assumptions (Elhag & Boussabaine, 1998; Kim et al., 2004; Wilmot & Mei, 2005). Specifically, ANN is applied to an early cost estimation of road tunnel construction (Petroutsatou et al., 2012). Classified as a non-parametric cost estimation technique, Juszczyk (2017) contends that ANN is a veritable tool for forecasting cost at the conceptual stage. ANN is described to possess advantages not limited its ability to establish relationship between cost and influencing parameters without making assumption or prior investigation of rules, work within a shorter time. Other researches that execute cost estimation by means of ANN include (Elhag & Boussabaine, 1998; Emsley et al., 2002; Matel et al., 2022; Wang et al., 2021). It is clear that researchers have employed different techniques to develop tools with a view to improving the accuracy of cost forecasting. Chronologically, these studies can be classified into two segments. For the period between 1990 – 2000, out of the retrieved articles devoted to forecasting techniques, regression techniques (Kim et al., 2004; Lowe et al., 2006; Trost & Oberlender, 2003) and artificial neural network (ANN) (Attalla & Hegazy, 2003; Elhag & Boussabaine, 1998; Emsley et al., 2002) were the most featured tools representing between 40% and 25% respectively. There is also a sense that these two have singularly been used as mono-technique. Others include case-based reasoning (CBR), support vector machine (SVM) and a combination of genetic algorithm (GA) and ANN (G. Kim et al., 2004; An et al., 2007; G. H. Kim et al., 2004).

The need to combine more than one technique presuppose some dissatisfaction with the existing technique. The year following 2011 saw the use of hybrid technique for estimating cost. For example, neural network's capability of modelling complex relations between factors and costs was exploited alongside a bootstrap technique that enabled a pragmatic method for quantification of the prediction variability (Sonmez, 2011). A study conducted in Taiwan saw the use of hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation (Cheng et al., 2013). With growing interest in forecasting studies and the quest to adopt approaches that complement themselves for improved performance, successive attempts saw the sustained application of either Regression technique or ANN with other techniques e.g. Case-Based Reasoning and Multiple Regression Analysis (Ji et al., 2012); integration of neural networks with bootstrap prediction intervals (Sonmez, 2011); and in some instance the combined use of ANN and regression techniques (Cirilovic et al., 2014; Wang et al., 2023). Other techniques include a machine learning simulation which presents a novel construction cost prediction model that incorporates hybrid natural and light gradient boosting (Chakraborty et al., 2020). The authors noted that the hybridisation gives room to not only enhance the accuracy of the predictions but also quantify the prediction interval and uncertainties.

The last categories of techniques found in literature within the forecasting domain include ensemble technique. The ensemble methods are clever techniques which combine multiple learning algorithms to improve general performance (Elmousalami, 2020). From the review, four studies is identified to have employed ensemble methods. Mohamed & Moselhi (2022) opted for grid search optimisation alongside ANN, random forest and support vector machine to develop conceptual estimation of construction duration and cost of public highway projects. The work focused on predicting construction cost using adaptive boosting and artificial neural networks by (Feng & Zou, 2023). The study's novelty lies in the surrogate approach and the addition of the beetle antennae search algorithm and enhancement of the genetic algorithm. Whereas the former aided a significant increased the search efficiency of the network, the latter generally increased the population fitness and mitigated the drawback of the genetic algorithm which was prone to local convergence. Others include the study who developed an Integrative data intelligence model for construction cost estimation by (Ali et al., 2022). The authors demonstrated that enhanced prediction is achieved using an advanced input selector algorithm of extreme gradient boosting (XGBoosting).

As big data analytics techniques continue on an increasing trajectory, several methods have been proposed to forecast cost. However, on the basis of the methodologies employed, this review establish that approaches employed in forecasting cost can be situated within two distinct approaches namely parametric method and artificial intelligence. This stance is supported by (Elmousalami, 2020; Juszczyk, 2017). With parametric estimating, large dataset can be processed to forecast more accurately, particularly when the underlying data is scalable. According to Project Management Body of Knowledge (PMBOK, 2021), parametric estimating is a technique that uses a statistical relationship between relevant historical data and other variables to predict a cost estimate for project work. The parametric cost estimation models are used to express a dependent variable (cost) in terms of independent variables (parameters). With robust analytical capabilities, big data conducts different advanced analytics; Evidence from this review clearly indicate that cost forecasting in the construction industry relies heavily on predictive analytics system to tap into the opportunities of big data as an effective strategy for improved decision-making and overall performance. Other forms of analytics include, inferential analytics, prescriptive analytics and descriptive analytics (Ohlhorst, 2012)

Q2 What is the nature of the construction data?

In providing answer to the question, this section begins with a discussion of the nature of the construction industry, examines the important features of the data used for forecasting in past studies, highlight other aspects such as sources of data employed for forecasting, quality and quantity of data used as input variable.

Ofori (2015) argued that for several decades, what the construction industry truly depicts has been vaguely captured. An understanding of the nature of the industry in term of specific requirement is important to understanding the nature of construction data (Ofori, 2015). The construction industry has largely operated by some rule-of-thumb rather than scientific evidence derived from solid data analytics. With the current data revolution, the construction industry is not exempted from the spate of data upsurge currently been witnessed in many industries. The industry currently deals with substantial amount of data which stems from the interactions carried out in the various phase from strategic definition to operation (Ashworth & Perera, 2015; Bilal et al., 2016b). Similarly, cost custodians within the industry are expected to provide reliable cost advise which itself spring from cost information. However, little attention is given to the nature of data & its management which the information themselves depends on (Boisot & Canals, 2007).

Data which serve as input variable is considered a vital element for the model's performance. From the 79 studies considered for review, only 18 studies (representing about 27%) clearly indicated details about the data used. While some authors related the forecasting with specific form of construction say road project others related cost estimation to building projects (Mahamid, 2011; Mohamed & Moselhi, 2022; Wilmot & Mei, 2005). Shin (2005) who utilised boosting regression trees for preliminary cost estimation of building projects identified budget, school level, land class number, building area, gross floor area, storey, basement floor storey, floor height as input variable; and others identified cost of materials, labour, equipment, type of contract, contracting environment (Shin, 2005; Wilmot & Mei, 2005). Zima (2015) adopts case-based reasoning model to estimation cost at a preliminary project stage, owing to the unique feature of the selected estimating model, elemental unit prices listed in tendered bids of past project served as data for input variable (Zima, 2015).

Data used by these studies to model attributes is in tandem with Shmueli & Koppi, (2011)'s position that data should, ideally, be extracted from a population of similar characteristics to achieve more accurate predictions. The category of buildings identified in the review were predominantly multistorey primarily for residential educational purpose. While the source of data used in some studies was not identifiable, 10 studies indicated the source of data used as input variable. These data originate mostly from general contractors, public databases (e.g., Building Cost Information Service, Bureau of Economic Analysis (BEA) of the Department of Commerce) public and private organisations. From the analysis, contracting firms and databases were the most commonly used data sources. At the early stage of development, building type, gross floor area, outline building shape, number of storeys, and such information may be available (Ashworth & Perera, 2015). From this review, floor area ranked highest among the studies that indicated the use of data as input variable (Ali et al., 2022; U. Park et al., 2022; Shin, 2015; Ugur, 2017). This is because information about the superficial area is easily identified at the early stage, though the accuracy of this method has been considered low (Miranda et al., 2022), it is still considered a reliable means of producing cost estimate based on updated historical data (Benge, 2014).

In describing how organisations can create and implement effective data strategies, Corea, (2019) highlight different types of data namely data at rest, data in motion and data in use. In the era of big data, diversified structured and unstructured data are important features of the construction industry. However, as shown in table 2, the data used for prediction are best described as quantitative variables; these represent the structured data as argued by (Alaka et al., 2018). Emphasis on relational database management has been given to evaluating the structured data yet adequate evidence from past studies support the claim that the unstructured data accounts for 80% of data generated in the industry ; Boyd & Crawford, 2012; Khan et al., 2014). The review uncovers the need to employ both structured and unstructured data to enhance reliable forecast. This has been demonstrated in studies by (Alaka, 2017; Delgado et al., 2020)

Table 3: Data used in the reviewed studies

References	Data (input variable)
Wilmot and Cheng (2003)	Labour, material, equipment, bill item quantity, contract duration, project location, quarter in which contract was let, annual bid volume, bid volume variance, number of plan changes per year, changes in practice, standards/specification
Shin (2015)	Budget, school level, land class number, building area, gross floor area, storey, basement floor storey, floor height
Wilmot and Mei (2005)	Cost of materials, labour, equipment, type of contract, contracting environment.
Zima (2015)	Elemental unit prices
Mohammed & Moselhi (2022)	Facility type, project scope, highway type, length, width, location, technical complexity, payment & procurement method
Park et al., (2022)	Gross floor area, building area, building height, numbers of floors, number of basement floors, numbers of parking space
Wang et al., (2021)	House type, foundation type, Nos of floors, Exterior wall decoration, Doors & window type, floor finish, project location
Ali et al., (2022)	GFA, Total floor area, floor number, elevator number, footing type, Inflation, duration
Emsley et al., (2010)	Past projects final sum and 41 variables clustered into 3 namely project strategic variables, site & design related variables.
Matel et al., (2019)	Scale of work, project phase, project duration, scope of work, type of work, client representative's experience, scope definition, project team size, multidisciplinary project nature, client type, market type, attitude to design changes, project manager' experience, pre-contract design, contact type, intensity
Alshamrani (2017)	Structure type, building area, numbers of floors, floor height, floor area
Ugur (2017)	Total apartments, maximum high, floor space, front area and front blank surface
Wang, et al., (2023)	Total floor area, height and number of floors

What methodologies have been adopted in previous researches to forecast cost?

The analysis literature carried out through a literature matrix indicate that cost forecasting comes under three themes namely parametric, artificial intelligence and a combination of both. First, 11 studies developed a parametric cost estimation model. The parametric cost estimation models are used to express a dependent variable (cost) in terms of independent variables (parameters). Some of the studies which execute parametric model in estimating cost include (Alshamrani, 2020; El-Assaly et al., 2006; Hegazy & Ayed, 1998; Sha et al., 2023). Secondly, a total of 32 studies employs artificial intelligence (AI) techniques for cost modelling. According to Elmousalami, (2020), AI-based techniques have a strong ability to develop applicable and accurate cost predictive models. The most common identified within the reviewed studies include artificial neural networks, fuzzy logic, regression models, random forest, case-based reasoning, support vector machine, LGBoosting, AdaBoosting, XGBoosting and genetic algorithm (GA). The third relates to studies which combine the parametric and AI technique for cost estimation; the 2 studies with this focus include that of (Elmousalami, 2020; Alshboul et al., 2022). The former, being a review paper provides a synthesis of applicable parametric and AI modelling technique, however the latter is a researched articles which developed workable models for determining the impact of external influences on green construction projects' costs.

The move away from subjective and potentially inaccurate forecast to the use of parametric models and AI forecasting techniques signal the increase in quantitative approaches. Cost estimation classified on the basis of methodology indicate that 55 studies (representing 69.6%) out of the 79 studies employ the quantitative approach. This classification sit well with a previous study which examined the challenges of nonparametric cost estimating approaches alongside the use of AI tools (Juszczyk, 2017). However, Alaka, (2017) argue that developing an efficient prediction model will require the consideration of both quantitative and qualitative variables. This study has revealed that quantitative methodology is more common than quantitative methods (Alaka, 2017).

Where does data management fit in within the forecasting process?

The requirement to produce a forecast at an early stage constrains the quantity surveyor to work within insufficient data provided by the designers (Lu et al., 2019). The analysis of the articles put together for this review indicate a unanimous position that the forecasting exercise is a process concluded within stage 0 (strategic definition) and stage 1 (preparation of brief) of the RIBA work stages (Benge, 2014). The development of forecasting models involves the identification of data sources, collection and preparation of data relevant to event being forecasted, data processing, data tunning, and implementation of big data analytics techniques. Lu et al., (2019) noted that the purpose of collecting data is to extract useful information and create knowledge so as to support informed decisions The concept of data subsists around terminologies like "information", "knowledge" which goes on to provide a robust framework for effective data management.

The discussion on these concept is evident in literature like (Lu et al., 2019; Corea, 2019). However, this review subscribes to Boisot & Canals, (2007)'s description which views data as facts originating in discernible differences in physical state of the world, registered through stimuli. Significant regularities in this data then constitutes information. This implies that the information gained from data, depends on the agent extracting it - more precisely: his expectations, or hypotheses (Boisot & Canals, 2007). This set of hypotheses held by an agent can then be referred to as knowledge and is constantly modified by the arrival of information. By extension, managing a construction project, certainly including its cost management, exploits available information and knowledge to make a web of decisions across the processes of architecture, engineering, and construction. It is against this setting that big data is guaranteed to play an important role.

Having considered how big data management supports decisions, this concluding aspect considers where big data situate within a forecasting process. The process by which a reliable initial forecast of a proposed facility cost involves both the formulation and communication of advice that will influence the client to make a decision on whether to progress or not with the project itself. The formulation phase requires the professional to select an appropriate cost model, which is then manipulated and a judgement is made on the output of the model in terms of the reliability or otherwise of its product. Big data has a crucial role to play in the forecasting process.

In a bid to set standardised estimating procedure, public entities and industry organisations like Royal Institute of Chartered Surveyors (RICS), Infrastructure and Project Authority both in the UK and the Government Accountability Office (GAO), Department of Energy (DOE) in the USA document cost estimating guides (Dept. of Energy, 2009; RICS, 2009; Smallwood, 2021). In pursuing the standardisation of cost-information management, the GAO published the "GAO Cost Estimating and Assessment Guide". This guide provides a 12-step process that acts as a guideline for the proper development of a cost estimate, and one that will withstand GAO's scrutiny. The guides developed by the RICS new rules of measurement provide important improvement and contain sets of rules to estimate construction projects' costs. More recently, in its effort to bring about a revolutionary step-change in how it delivers major projects. The Infrastructure and Project Authority issued a 8 steps cost estimating process in its report – Cost Estimating Guidance – to the Cabinet Office and HM Treasury (Smallwood, 2021). Although none of these have been universally adopted, each detail series of steps in an estimating process. From table 3, a harmonised 10 steps estimation process extracted from literature is classified into three major aspects namely; definition & planning (steps 1-3), data management (steps 4-7) and presentation/reporting (steps 8-10). Within the harmonised process, step 4 involves *gathering data and discernible fact.* This where data management comes in. The first three steps give a clear indication about the need to understand the forecast purpose, put together people who ensures timely delivery of a credible forecast and establish principles (*purpose, people, principles*). Data and evidence supporting a reliable estimate is expected to include the quantitative and qualitative variables. For quantitative variables, items such as land and property, administration, market trends, direct and indirect cost are necessary for consideration. As data availability is expected to improve with progresses in project, the aspect of data quality and integrity is to been taken as a fundamental aspect. As Benge, (2014) illustrates, data and information gathering stage provide basis for collecting historical data required for future cost estimate. The analysis and application of quantitative techniques will lead to an array of decision-making based on available information and knowledge. Big data, with its characteristics of volume, variety, and velocity, provides an opportunity to help make informed decision (Benge, 2014).

Table 4: Steps involved in preparation of order of cost estimate

US practice	Developed in UK	Developed in UK	Excerpt from Literature
12 Steps Cost Estimating Guide Department of Energy (DOE). GAO, (2018)	13 Steps NRM 1 Cost Management Handbook - Order of cost estimate. Benge, 2014.	8 Steps Cost estimating guidance - HM Treasury. Smallwood (2021).	10 harmonised steps of order of cost estimate
Step 1: Define the Estimate's Purpose	Step 1: Basic information gathering	Step 1: Establish brief and engage the team	Step 1: Define estimate's purpose
Step 2: Develop an Estimating Plan	Step 2: Produce building work estimate	Step 2: Gather data and evidence	Step 2: Establish brief and engage the team
Step 3: Define the Program Characteristics	Step 3: Ascertain work cost estimate	Step 3: Select cost estimating methodology	Step 3: Establish estimating ground rules and assumptions
Step 4: Determine the Estimating Structure	Step 4: Produce project and design team fees estimate	Step 4: Calculate base cost estimate, uncertainty, risk	Step 4: Gather data & discernable facts
Step 5: Identify Ground Rules and Assumptions	Step 5: Produce development and project cost estimate	Step 5: Produce cost estimate report	Step 5: Determine cost estimating methodology
Step 6: Obtain Data	Step 6: Ascertain base cost estimate	Step 6: Review and assure	Step 6: Calculate base cost estimate
Step 7: Develop a Point Estimate, Compare to Independent Estimate	Step 7: Produce risk allowance estimate	Step 7: Project leadership sign-off	Step 7: Conduct sensitivity analysis; allowance for inflations, uncertainty, risk
Step 8: Conduct Sensitivity Analysis	Step 8: Ascertain cost limit exclusive of inflation	Step 8: Use the cost estimate to support decision- making	Step 8: Produce cost estimate report
Step 9: Conduct Risk & Uncertainty Analysis	Step 9: Produce inflation estimate		Step 9: Review, ascertain and validate estimate
Step 10: Document the Estimate	Step 10: Ascertain cost limit inclusive of inflation		Step 10: Authorize use of cost estimate
Step 11: Present Estimate to Management for Approval	Step 11: Produce VAT assessment		
Step 12: Update the Estimate to Reflect Actual Costs and Changes	Step 12: Produce order of cost estimate report		
	Step 13: Issue order of cost estimate report		

4. Conclusion

The purpose of this research was to investigate Big Data Analytics Techniques Applications in Construction Cost Forecasting domain. A systematic literature search found most of the search results. The accumulation of data has instigated the development of various techniques for extracting useful insight to support decision making. Over the last two decades, progress in modern analytical techniques have been seen to impact on business decisions. However, its uptake and potential to improve cost forecasting in the construction industry has not been sufficiently highlighted thus limiting the industry's capacity to optimise its functions. To encourage big data analytics utilisation, a systematic literature review is conducted in this study to investigate the application of big data analytics techniques to construction cost forecasting. From the research questions put forward, extensive analysis from literature point to the following *implication*; First, a variety of techniques have been examined to understand their evolution, predictive power. The different types of big data analytics approaches employed for construction cost forecasting have been classified into three distinctive types namely mono-technique, hybrid techniques and ensemble technique. On the basis of the methodologies employed, this review establish that approaches employed in forecasting cost can be situated within two distinct approaches namely parametric method and artificial intelligence. Secondly, the nature of the construction industry dictates the features of data necessary to support decisions. The initial analysis of the studies indicates that less than one-third of the articles indicate details about the data used. Further analysis disclose the exclusive use of structured data for all the model developed to estimate cost. Against this backdrop, this study however canvass for the inclusion of unstructured (qualitative) data variables in building cost forecasting models. To develop a valid forecasting model, both structured and the unstructured data should be considered. Thirdly, on the basis of methodology, the two broad classification which is seen to apply to the construction industry include parametric approach and artificial intelligence approach. The techniques under these approaches is classified in this study as mono-technique, hybrid technique and ensemble techniques. The classification of these techniques is directly be linked to chronological advancement in techniques employed to forecast. Lastly, the sources of data used in the study was also not uniform; however, their extraction represent the diversified nature of the construction industry. The study makes a distinction between structured and unstructured data and makes a case for the inclusion of unstructured data as vital consideration for a robust forecasting model.

References

Abdel-Monem, M., Alshaer, K., & El-Dash, K. 2022. Assessing Risk Factors Affecting the Accuracy of Conceptual Cost Estimation in the Middle East. Buildings, 12(7), 950. https://doi.org/10.3390/buildings12070950

Adafin, J., Rotimi, J. O. B., Wilkinson, S., & Windapo, A. O. 2017. Achieving Design-Stage Elemental Cost Planning Accuracy: Case Study of New Zealand. 11(1).

Alaka, H. A. (2017). 'Big data analytics' for construction firms insolvency prediction models. University of the West of England, Bristol.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Bilal, M., Ajayi, S. O., & Akinade, O. O. 2018. A framework for big data analytics approach to failure prediction of construction firms. Applied Computing and Informatics, 16(1/2), 207–222. https://doi.org/10.1016/j.aci.2018.04.003

Ali, Z. H., Burhan, A. M., Kassim, M., & Al-Khafaji, Z. 2022. Developing an Integrative Data Intelligence Model for Construction Cost Estimation. Complexity, 2022, 1–18. https://doi.org/10.1155/2022/4285328

Alshamrani, O. S. 2020. Initial cost forecasting model of mid-rise green office buildings. Journal of Asian Architecture and Building Engineering, 19(6), 613–625. https://doi.org/10.1080/13467581.2020.1778005

Alshboul, O., Shehadeh, A., Almasabha, G., Mamlook, R. E. A., & Almuflih, A. S. 2022. Evaluating the Impact of External Support on Green Building Construction Cost: A Hybrid Mathematical and Machine Learning Prediction Approach. Buildings, 12(8), 1256. https://doi.org/10.3390/buildings12081256

An, S. H., Park, U. Y., Kang, K. I., Cho, M. Y., & Cho, H. H. 2007. Application of support vector machines in assessing conceptual cost estimates. Journal of Computing in Civil Engineering, 21(4), 259–264. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:4(259)

Arif, F., Lodi, S. H., & Azhar, N. 2015. Factors influencing accuracy of construction project cost estimates in Pakistan: Perception and reality. International Journal of Construction Management, 15(1), 59–70. https://doi.org/10.1080/15623599.2015.1012141

Ashworth, A., & Perera, S. 2015. Cost studies of buildings (6th edition). Routledge, Taylor & Francis Group.

Attalla, M., & Hegazy, T. 2003. Predicting Cost Deviation in Reconstruction Projects: Artificial Neural Networks versus Regression. Journal of Construction Engineering and Management, 129(4), 405–411. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:4(405)

Awosina, A., Ndihokubwayo, R., & Fapohunda, J. 2006. Effects of inaccurate cost estimate on construction project stakeholder.

Bala, K., Ahmad Bustani, S., & Shehu Waziri, B. 2014. A computer-based cost prediction model for institutional building projects in Nigeria: An artificial neural network approach. Journal of Engineering, Design and Technology, 12(4), 519–530. https://doi.org/10.1108/JEDT-06-2012-0026

Barakchi, M., Torp, O., & Belay, A. M. 2017. Cost Estimation Methods for Transport Infrastructure: A Systematic Literature Review. Procedia Engineering, 196, 270–277. https://doi.org/10.1016/j.proeng.2017.07.199

Barnes, N. 1970. Technical note on civil engineering bills of quantities. An interim report on CIRIA research project. 98. Proceedings of the Institution of Civil Engineers, 45(1), 131–134. https://doi.org/10.1680/iicep.1970.7209

Bayram, S., & Al-Jibouri, S. 2016. Efficacy of Estimation Methods in Forecasting Building Projects' Costs. Journal of Construction Engineering and Management, 142(11), 05016012. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001183

Benge, D. P. 2014. NRM 1 Cost Management Handbook. Routledge, Taylor & Francis Group.

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. 2016a. Big Data in the construction industry: A review of present status, opportunities, and future trends. Advanced Engineering Informatics, 30(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. 2016b. Big Data in the construction industry: A review of present status, opportunities, and future trends. Advanced Engineering Informatics, 30(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

Boisot, M. H., & Canals, A. 2007. Data, Information, and Knowledge: Have We Got It Right. In Exploration in Information Space: Knowledge, Agents, and Organization. Open University. Press.

Bowen, P. A., & Edwards, P. J. 1985. Cost modelling and price forecasting: Practice and theory in perspective. Construction Management and Economics, 3(3), 199–215. https://doi.org/10.1080/01446198500000015

Boyd, D., & Crawford, K. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly phenomenon. Information Communication, & Society, 15(5), 662–679.

Chakraborty, D., Elhegazy, H., Elzarka, H., & Gutierrez, L. 2020. A novel construction cost prediction model using hybrid natural and light gradient boosting. Advanced Engineering Informatics, 46, 101201. https://doi.org/10.1016/j.aei.2020.101201

Chandanshive, V., & Kambekar, A. 2019. Estimation of Building Construction Cost Using Artificial Neural Networks. Journal of Soft Computing in Civil Engineering, 3(1). https://doi.org/10.22115/scce.2019.173862.1098

Cheng, M.-Y., Hoang, N.-D., & Wu, Y.-W. 2013. Hybrid intelligence approach based on LS-SVM and Differential Evolution for construction cost index estimation: A Taiwan case study. Automation in Construction, 35, 306–313. https://doi.org/10.1016/j.autcon.2013.05.018

Cirilovic, J., Vajdic, N., Mladenovic, G., & Queiroz, C. 2014. Developing Cost Estimation Models for Road Rehabilitation and Reconstruction: Case Study of Projects in Europe and Central Asia. Journal of Construction Engineering and Management, 140(3), 04013065. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000817

Corea, F. 2019. An Introduction to Data: Everything You Need to Know about AI, Big Data and Data Science (Vol. 50). Springer.

Davila Delgado, J. M., Oyedele, L., Bilal, M., Ajayi, A., Akanbi, L., & Akinade, O. 2020. Big Data Analytics System for Costing Power Transmission Projects. Journal of Construction Engineering and Management, 146(1), 05019017. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001745

Dept. of Energy. 2009. GAO Cost Estimating and Assessment Guide (09-3SP).

Dosumu, B., Ejohwomu, O., Yunusa-Kaltungo, A., & Daramola, O. 2022. A Systematic Review on Development of a Framework for Construction Project Cost Estimation: A Case Study of Nigeria. 11.

El-Assaly, A., Ariaratnam, S. T., Ruwanpura, J., & Ng, H. 2006. Cost forecast model for sewer infrastructure. Proceedings of the Institution of Civil Engineers - Municipal Engineer, 159(3), 155–160. https://doi.org/10.1680/muen.2006.159.3.155

Elbeltagi, E., Hosny, O., Abdel-Razek, R., & El-Fitory, A. 2014. Conceptual Cost Estimate of Libyan Highway Projects Using Artificial Neural Network. 4(8).

Elfaki, A. O., Alatawi, S., & Abushandi, E. 2014. Using Intelligent Techniques in Construction Project Cost Estimation: 10-Year Survey. Advances in Civil Engineering, 2014, 1–11. https://doi.org/10.1155/2014/107926

Elhag, T. M. S., & Boussabaine, A. H. 1998. An artificial neural system for cost estimation of construction projects.

Elmousalami, H. H. 2020. Artificial Intelligence and Parametric Construction Cost Estimate Modeling: State-of-the-Art Review. Journal of Construction Engineering and Management, 146(1), 03119008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001678

Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A., & Hickson, A. 2002. Data modelling and the application of a neural network approach to the prediction of total construction costs. Construction Management and Economics, 20(6), 465–472. https://doi.org/10.1080/01446190210151050

Fazil, M. W., Lee, C. K., & Muhamad Tamyez, P. F. 2021. Cost estimation performance in the construction projects: A systematic review and future directions. International Journal of Industrial Management, 11, 217–234. https://doi.org/10.15282/ijim.11.1.2021.6131

Feng, W., & Zou, Y. 2023. Construction cost prediction based on adaptive boosting and artificial neural networks. Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction, 1–9. https://doi.org/10.1680/jsmic.22.00027

Ganiyu, B. O., & Zubairu, I. K. 2010. Project Cost Prediction Model Using Principal Component Regression For Public Building Projects In Nigeria. 1(1).

Hatamleh, M. T., Hiyassat, M., Sweis, G. J., & Sweis, R. J. 2018. Factors affecting the accuracy of cost estimate: Case of Jordan. Engineering, Construction and Architectural Management, 25(1), 113–131. https://doi.org/10.1108/ECAM-10-2016-0232

Hegazy, T., & Ayed, A. 1998. Neural Network Model for Parametric Cost Estimation of Highway Projects. Journal of Construction Engineering and Management, 124(3), 210–218. https://doi.org/10.1061/(ASCE)0733-9364(1998)124:3(210)

Jahan, N., Naveed, S., Zeshan, M., & Tahir, M. A. 2016. How to Conduct a Systematic Review: A Narrative Literature Review. Cureus. https://doi.org/10.7759/cureus.864

Ji, S.-H., Park, M., & Lee, H.-S. 2012. Case Adaptation Method of Case-Based Reasoning for Construction Cost Estimation in Korea. Journal of Construction Engineering and Management, 138(1), 43–52. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000409

Juszczyk, M. 2017. The Challenges of Nonparametric Cost Estimation of Construction Works with the use of Artificial Intelligence Tools. Procedia Engineering, 196, 415–422. https://doi.org/10.1016/j.proeng.2017.07.218

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., & Gani, A. 2014. Big Data: Survey, Technologies, Opportunities, and Challenges. The Scientific World Journal, 2014, 1–18. https://doi.org/10.1155/2014/712826

Kim, G., An, S., & Kang, K. 2004. Comparison of construction cost estimating models based on regression analysis , neural networks , and case-based reasoning. 39, 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Kim, G. H., Yoon, J. E., An, S. H., Cho, H. H., & Kang, K. I. 2004. Neural network model incorporating a genetic algorithm in estimating construction costs. Building and Environment, 39(11), 1333–1340. https://doi.org/10.1016/j.buildenv.2004.03.009

Kim, G.-H., An, S.-H., & Kang, K.-I. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Building and Environment, 39(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Kitchenham, B. 2004. Procedures for Performing Systematic Reviews.

Lowe, D. J., Emsley, M. W., & Harding, A. 2006. Predicting Construction Cost Using Multiple Regression Techniques. Journal of Construction Engineering and Management, 132(7), 750–758. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(750)

Lu, W., Lai, C. C., & Tse, T. 2019. BIM and Big Data for Construction Cost Management. Routledge, Taylor & Francis Group.

Mahamid, I. 2011. Early cost estimating for road construction projects using multiple regression techniques. Construction Economics and Building, 11(4), 87–101. https://doi.org/10.5130/AJCEB.v11i4.2195

Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H. 2022. An artificial neural network approach for cost estimation of engineering services. International Journal of Construction Management, 22(7), 1274–1287. https://doi.org/10.1080/15623599.2019.1692400

McCaffer, R. 1976. Contractors' bidding behaviour and tender price prediction [PhD Thesis]. Loughborough University of Technology.

Membah, J., & Asa, E. 2015. Estimating cost for transportation tunnel projects: A systematic literature review. International Journal of Construction Management, 15(3), 196–218. https://doi.org/10.1080/15623599.2015.1067345

Miranda, S. L. C., Castillo, E. D. R., Gonzalez, V., & Adafin, J. 2022. Predictive Analytics for Early-Stage Construction Costs Estimation. https://doi.org/10.3390/buildings12071043

Mohamed, B., & Moselhi, O. 2022. Conceptual estimation of construction duration and cost of public highway projects. Journal of Information Technology in Construction, 27, 595–618. https://doi.org/10.36680/j.itcon.2022.029

Moher, David, Liberati, Alessandro, Tetzlaff, Jennifer, Altman, G. Douglas, & The Prisma Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS, 6(7), 7. https://doi.org/doi:10.1371/ journal.pmed.1000097

Morrison, N. 1984. The accuracy of quantity surveyors' cost estimating. Construction Management and Economics, 2(1), 57–75. https://doi.org/10.1080/01446198400000006

Ngo, J., Hwang, B.-G., & Zhang, C. 2020. Factor-based big data and predictive analytics capability assessment tool for the construction industry. Automation in Construction, 110, 103042. https://doi.org/10.1016/j.autcon.2019.103042

Ofori, G. 2015. Nature of the construction industry, Its Needs and Its Development: A Review of Four Decades of Research. Journal of Construction in Developing Countries, 20(2).

Ohlhorst, F. 2012. Big Data Analytics—Turning Big Data into Big Money. John Wiley & Sons, Inc.

Park, D., & Yun, S. 2023. Construction Cost Prediction Using Deep Learning with BIM Properties in the Schematic Design Phase. Applied Sciences, 13(12), 7207. https://doi.org/10.3390/app13127207

Park, U., Kang, Y., Lee, H., & Yun, S. 2022. A Stacking Heterogeneous Ensemble Learning Method for the Prediction of Building Construction Project Costs. Applied Sciences, 12(19), 9729. https://doi.org/10.3390/app12199729

Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., & Pantouvakis, J. P. 2012. Early Cost Estimating of Road Tunnel Construction Using Neural Networks. Journal of Construction Engineering and Management, 138(6), 679–687. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000479

PMBOK, G. 2021. A Guide to the Project Management Body of Knowledge (7th ed.). Project Management Institute, Inc.

Ram, J., Afridi, N. K., & Khan, K. A. 2019. Adoption of Big Data analytics in construction: Development of a conceptual model. Built Environment Project and Asset Management, 9(4), 564–579. https://doi.org/10.1108/BEPAM-05-2018-0077

Rayyan, T. M. 2017. Cost Estimation of Building Construction Projects in Gaza Strip Using Support Vector Machines Model (SVM). Cost Estimation of Building Construction Projects in Gaza Strip Using Support Vector Machines Model (SVM).

RICS, R. I. of C. S. 2009. NRM 1: Order of cost estimating and cost planning for capital building works (First).

Sha, J., Dong, H., Xie, H., Yang, B., Shang, X., & Ling, Y. 2023. Construction Cost Prediction of Transmission Line Engineering Under the Background of Big Data. In I. Ahmad, J. Ye, & W. Liu (Eds.), The 2021 International Conference on Smart Technologies and Systems for Internet of Things (Vol. 122, pp. 461–468). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-3632-6_56

Shin, Y. 2015. Application of Boosting Regression Trees to Preliminary Cost Estimation in Building Construction Projects. Computational Intelligence and Neuroscience, 2015, 1–9. https://doi.org/10.1155/2015/149702

Shmueli, G., & Koppi, O. R. 2011. Predictive Analytics in Information Systems Research. MIS Quarterly, 35(3), 553–572.

Siddaway, A. P., Wood, A. M., & Hedges, L. V. 2019. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. Annual Review of Psychology, 70(1), 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. 2017. Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Smallwood, N. 2021. Cost Estimating Guidance: Best approach for infrastructture projects & programmes.

Sonmez, R. 2011. Range estimation of construction costs using neural networks with bootstrap prediction intervals. Expert Systems with Applications, 38(8), 9913–9917. https://doi.org/10.1016/j.eswa.2011.02.042

Tayefeh Hashemi, S., Ebadati, O. M., & Kaur, H. 2020. Cost estimation and prediction in construction projects: A systematic review on machine learning techniques. SN Applied Sciences, 2(10), 1703. https://doi.org/10.1007/s42452-020-03497-1

Trost, S. M., & Oberlender, G. D. 2003. Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression. Journal of Construction Engineering and Management, 129(2), 198–204. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:2(198)

Ugur, L. O. 2017. A Neuro-Adaptive Learning (NAL) Approach about Costs of Residential Buildings. Acta Physica Polonica A, 132(3), 585–587. https://doi.org/10.12693/APhysPolA.132.585

Wang, B., Yuan, J., & Ghafoor, K. Z. 2021. Research on Construction Cost Estimation based on Artificial Intelligence Technology. Scalable Computing: Practice and Experience, 22(2), 93–104. https://doi.org/10.12694/scpe.v22i2.1868

Wang, Y., Zuo, J., Pan, M., Tu, B., Chang, R.-D., Liu, S., Xiong, F., & Dong, N. 2023. Cost prediction of building projects using the novel hybrid RA-ANN model. Engineering, Construction and Architectural Management. https://doi.org/10.1108/ECAM-07-2022-0666

Wilmot, C. G., & Mei, B. 2005. Neural Network Modeling of Highway Construction Costs. Journal of Construction Engineering and Management, 131(7), 765–771. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:7(765)

Wohlin, C. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, 1–10. https://doi.org/10.1145/2601248.2601268

Xu, M., Xu, B., Zhou, L., & Wu, L. 2015. Construction Project Cost Prediction Based on Genetic Algorithm and Least Squares Support Vector Machine. Proceedings of the 5th International Conference on Civil Engineering and Transportation 2015. 5th International Conference on Civil Engineering and Transportation, Guangzhou, China. https://doi.org/10.2991/iccet-15.2015.190

Yang, S.-W., Moon, S.-W., Jang, H., Choo, S., & Kim, S.A. 2022. Parametric Method and Building Information Modeling-Based Cost Estimation Model for Construction Cost Prediction in Architectural Planning. Applied Sciences, 12(19), 9553. https://doi.org/10.3390/app12199553

Yun, S. 2022. Performance Analysis of Construction Cost Prediction Using Neural Network for Multioutput Regression. Applied Sciences, 12(19), 9592. https://doi.org/10.3390/app12199592

Zhang, Y., & Fang, S. 2019. RSVRs based on Feature Extraction: A Novel Method for Prediction of Construction Projects' Costs. KSCE Journal of Civil Engineering, 23(4), 1436–1441. https://doi.org/10.1007/s12205-019-0336-3

Zima, K. 2015. The Case-based Reasoning Model of Cost Estimation at the Preliminary Stage of a Construction Project. Procedia Engineering, 122, 57–64. https://doi.org/10.1016/j.proeng.2015.10.007