International Journal of Science Academic Research
Vol. 05, Issue 07, pp.7892-7901, July, 2024
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

ARTIFICIAL INTELLIGENCE-DRIVEN CONSTRUCTION: LEVERAGING MACHINE LEARNING FOR PREDICTIVE MODELLING FOR COMMODITIES ESTIMATION

***Ahmed Safwat Aly Hussein Ewida**

Department of Electronics and Communication Engineering, Cairo University, Giza District, Giza Governorate 12613, Egypt

Received 09th May 2024; **Accepted** 05th June 2024; **Published online** 30th July 2024

Abstract

The integration of digital technologies like Artificial Intelligence (AI) and Machine Learning (ML), in construction practices can enhance productivity and reduce costs. ML is crucial in estimating construction costs, enhancing safety and reliability in projects like buildings, bridges, roads, airports, canals, and railroads. The price of commodities in construction includes all sorts of energy commodities and raw materials. This research analysed data from oil and gas and petrochemical projects in Egypt, Saudi Arabia, Qatar, UAE, and Oman from 2005-2017, focusing on daily reports and 460,172 records on piping erection, manpower, equipment, heat index, and delays. This research used predictive ML modelling to analyse daily piping erection rates in Egypt and Qatar under different conditions. Egypt's rate is high, but Qatar's implementation of HSE restrictions and high heat index significantly reduces it. The findings emphasised the importance of regulatory compliance, environmental conditions, and working conditions in optimizing production output in construction projects. The deep learning ANN model was found to be the most effective in predicting piping erection per day, outperforming other ensemble methods and suggesting better capture of the dataset's complexity and non-linearity.

Keywords: Artificial Intelligence; Machine Learning; Predictive Modelling; Commodities Cost Estimation; Construction; Piping Erection.

INTRODUCTION

The integration of digital technologies like Artificial Intelligence (AI), Big Data, Machine Learning (ML), and IoT in construction practices can enhance productivity and reduce costs [1-3]. The construction design process is generally divided into four stages, including pre-design, schematic design, design development and working drawing [4-6]. Precisely, the costs of construction are predicted during the 'Pre-Design' stage to assess project feasibility. However, during the 'Pre-Design' or 'Schematic Design' stages, the 'Unit Cost of Construction' is calculated for the existing building types annually. It must be computed with a margin of approximately 10%, which is lower than that of the 'Detailed Cost Estimate' [5]. ML plays a vital role in construction cost estimation for improving the safety and reliability of the project since the financial costs of commodities represent key criteria for constructing, designing, and maintaining buildings, bridges, roads, airports, canals, railroads, and others[7, 8]. Cost estimation is known as a quantitative estimation of the resource costs for all the processing parts based on the premise on which a company schedules its production. In customised mass production, various parts are needed to be estimated accurately and quickly. The production costs for mechanical part processing are always a concern for enterprises to survive the extreme competition [9]. Therefore, the conceptual costs' precise estimation is handled by cost engineers, project managers and decision-makers using ML models like Artificial Neural Networks (ANNs), Ordinary Least Square (OLS) and Support Vector Machines (SVM) for cost estimation [7, 10]. The price of commodities in construction includes all sorts of energy commodities and raw materials[11]. These costs are predicted using ML-based forecasting models [12].

Corresponding Author:** ***Ahmed Safwat Aly Hussein Ewida
Department of Electronics and Communication Engineering, Cairo University, Giza District, Giza Governorate 12613, Egypt.

In the construction sector, the Construction Cost Index (CCI) is the most important indicator to monitor cost trends, comparing goods and services for better acquisition. Therefore, use of ML for prediction modelling is efficient for better accuracy of estimation cost, bidding operation and construction investments [13, 14]. Past studies have examined the construction industry's massive data to address challenges like supply chain management, sustainability issues, project performance management, reduced productivity, and profitability using AI-driven prediction modelling. However, a gap was observed in examining the accuracy and application of various ML prediction models for predicting piping spool erection daily production in the oil and gas construction sector. Therefore, the current research aimed to develop a prediction model for piping spool erection daily production and the factors affecting productivity in industrial and oil and gas projects using historical data. The model can be applied to various construction commodities like concrete, steel fixing, shuttering, piping fabrication, equipment installation, steel structure fabrication/erection, facade systems, painting, and plastering. The research examined various ML and deep learning models, including ANNs, Linear Regression (LR), Random Forest (RF), Cat Boost Regressor (CR), and Light Gradient Boosting Model (LGBM) Regressor, to determine the best prediction model for daily piping spools erection production and identify features influencing production in oil and gas construction sites of Egypt, Saudi Arabia, Qatar, UAE, and Oman. The goal is to create a scientific dashboard using ML models for construction project commodity estimation.

LITERATURE REVIEW

In the study by Sammour *et al.* (2023), the application of ML techniques to predict Jordan's need for residential buildings is assessed. A study of the literature was conducted, features were chosen using stepwise backward elimination, and the ML predictions were compared with the actual residuals and the coefficient of prediction. The demand models were developed using nine different economic indicators. When compared to ANNs (0.727), Elastic-Net had the best accuracy (0.838), in addition to Eureqa (0.715) and Extra Trees (0.703). According to the best-performing model estimate, the estimated demand for residential buildings in Jordan's first quarter in 2023 will rise by 11.5% over the same period in 2022 [15]. Economic variables and indexes (EV&Is), as well as personnel, equipment, materials, and techniques, all have an impact on construction costs. The research by Rafiei and Adeli(2023) research proposed an innovative methodology for estimating construction costs that takes into account EV&Is and makes use of cutting-edge ML algorithms. The model takes into account the EV&Is factors that impact building, including the physical and financial (P&F) characteristics of real estate units. It consists of a softmax layer (DBM-SoftMax), support vector machine (SVM) or three-layer back-propagation neural network (BPNN), and an unsupervised deep Boltzmann machine (DBM) learning technique. The study demonstrates that combining traditional BPNN and SVM can enhance their effectiveness and accuracy by considering EV&I components at varying time delays. The study found that the proposed model had significantly lower cost error estimates compared to SVM-only and BPNN-only models, validating the model on 372 three- to nine-story structures[16].

According to Jung *et al.* (2018), decision-making is improved throughout the planning and design phases of smart educational facilities by forecasting construction costs with more accuracy. When project managers are unable to precisely forecast construction costs in the early phases of the project, they take greater risks and make more logical decisions. Several AI-based models were employed to successfully anticipate building costs throughout the planning and design phase despite the restricted variables and absence of performance data. The study discovered that over fitting is a common issue with deep learning and artificial neural networks, which presents a challenge for real-world applications. The study recommends employing Deep Belief Network (DBN) and Deep Neural Network (DNN) models for more precise predictions for construction cost prediction in the planning and design stage[17]. Jo and Yun (2021) analysed the appropriate impact factors in order to increase the accuracy of the conceptual cost estimate prediction model. The purpose is to provide a more precise assessment. Therefore it suggested integrating a number of quantitative effect elements that may be calculated at an initial stage. Regression analysis was used to analyse the accuracy of several situations; the best combination of impact factors resulted in an accuracy improvement of 0.2-4.7%. This improvement develops from a more precise and effective project budget as a consequence of the elimination of superfluous impact variables. The study emphasises that it is crucial to take these things into account when planning and making decisions for projects [18]. According to Ning *et al.* (2020), accurately estimating production costs are essential for increasing product competitiveness in the era of mass customisation. Common cost estimating algorithms include feature-based, process-based, non-parametric, and activity-based techniques, which compare every step in the production process, including tooling, labour, transportation, packaging, manufacturing specifications, and processing equipment. Regression-model cost estimating techniques, however, have difficulty with complicated mapping connections because of their fluctuating processing interactions, structures, and functions. Deep learning techniques, such as voxel data approaches and two- and three-dimensional (3D) convolutional neural network (CNN) training pictures, automatically discover these correlations, as suggested in this research [9]. The study by Ma et al.(2023) examined the employment of several models, including time series and ML approaches, in areas including the cost and quality of building materials, construction expenses, product consumption, housing prices, and bid award amounts. The hedonic pricing model was used to anticipate house prices, while hybrid models were employed to predict building costs and CCI. Findings revealed that the ANN model is an effective ML model for the quick and simple estimation of commodity costs based on physical features and intricate interdependencies between factors. Its benefits include addressing nonlinear interactions between cost-related factors, modelling interdependencies in input data, and handling incomplete data sets more effectively than regression. Nevertheless, because of its greater complexity and decreased degree of freedom, it has drawbacks, including under fitting and over fitting [12]. The research by Al Janabi(2022) studied the perspectives of portfolio managers to examine the liquidity-adjusted risk modelling for the market risk parameters. A large commodity portfolio uses these to obtain coherent and efficient portfolios using reinforcement ML to handle risk-return characteristics subjected to meaningful financial and operational constraints under adverse and stressed marketing conditions. Results showed that the approach liquidity-adjusted value-at-risk (LVaR) is an effective framework for commodities price estimation for asset allocation, risk reduction, and coherent markets[19].

METHODOLOGY

Data Collection

In this research, data was gathered from oil and gas and petrochemical projects in Egypt, Saudi Arabia, Qatar, UAE, and Oman. The data was collected from 2005 to 2017, with an average of 35 projects, with Egypt having 14, Saudi Arabia having ten, Qatar having three, UAE having four, and Oman having four projects. The data was collected from daily reports, which encompass the daily production of piping erection, manpower status, equipment status, heat index/temperature, area of concern, and reasons for delay. Figure 1 below explains the stages of data collection and model evaluation.

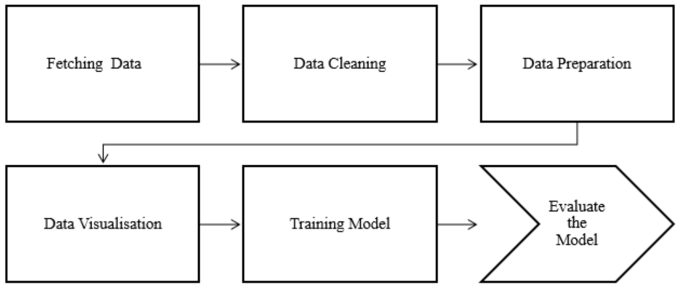

Figure 1. Problem Solution Process

Model Features

The selected features are listed in Table 1 below (23 Features).The shape of our data is 15,868 rows (inputs) and 29 columns (features), with a total of 460,172 records.

Table 1. Model Features

Features	Categories	Type
Country	• Egypt • Saudi Arabia • Qatar • UAE • Oman	Object
Health, Safety and Environment Executive (HSE)Restrictions	• Yes (If HSE and security restrictions are found severe in Gulf countries and oil and gas life areas in Egypt) • NO	Object
Temperature/Heat index	• High in June, July and August • Low Otherwise	Object
Political Issues	• Yes, in the financial crisis in 2008 and the Arab Spring (a series of anti-government protests) in 2011, many projects were impacted • No Otherwise	Object
Material of Pipes	• Carbon Steel • Stainless Steel • Low Temp • Duplex	Object
Pipes Diameter	• Low (< 10 D.I, Medium (between 10 and 22 D.I) • Large (< 22 D.I)	Object
Availability of Material	• Yes and No	Object
Holidays	• Yes and No	Object
Fabricated Spools Availability	• Low and High	Object
Distance between the spools fabrication Workshop and site(L.M)	• Low in case the distance is within the site boundary • High otherwise	Object
Crews Nationality	• Arab and others	Object
Crew Experience	• Low (< 5 years) • High (>= 5 years)	Object
Crew Supervision	• Normal (Supervision/ Direct Labor = 10% to 15%) • Low (Supervision/ Direct Labor <10% • High (Supervision/ Direct Labor > 15%)	Object
Drawings Availability	• Low and High	Object
Work Front Availability	• Yes and No	Object
Working at Heights	• Yes (If the piping erection is on pipe racks with a height of more than 4.60 meters) • No	Object
Number of Pipe Fitter		Float64
Number of Grinder		Float64
Number of Pipe Welders (CS)		Float64
Number of Pipe Welder (Argon)		Float64
Number of Riggers		Float64
Number of Cranes		Float64
No of Inspectors for NDT Test		Float64

After studying the row data, some computed statistics are shown below:

- The Maximum Piping Erection per Day is 32945 D.I/Day.
- The Minimum Piping Erection per Day is 466 D.I/Day.
- The Mean Piping Erection per Day is 6938 D.I/Day.
- The 75% of the records have piping Erection 9216 D.I/Day.

Data Pre-processing

Data Cleaning

1. Gather and prepare training data.
2. Using web scraping to collect the official holidays in Egypt, Qatar, Saudi Arabia, UAE, and Oman for the period from 2005 to 2017.
3. Remove duplicates and outliers, and deal with missing data. Normalize categorical data by normalizing the float and integer values.
4. Using Natural Language Processing Techniques to collect the area of concern from daily reports, which impact the production rate.

Data Preparation

1. Visualize data to help detect relevant relationships between variables.
2. Split into training and evaluation sets as below:

Training Model

The goal of training is to make a prediction correctly as often as possible; the model becomes better as it is trained to more data.

Evaluate the Model

Using some metric or combination of metrics to measure the performance of the model.

1. Shuffling the data and selecting a 15/85 ratio for the test/train data set.
2. Hyper-parameter tuning is a cornerstone for Model efficiency and performance improvement.
3. Using test set data to predict the output.

Evaluation Metrics

The regression problem is the issue that necessitates the root mean squared error (RMSE). It is a metric used to measure the average magnitude of errors in regression models. It measures deviations from the actual value and is used to determine if a feature is improving the model's prediction. A RMSE value of zero indicates a perfect fit, while a lower RMSE indicates better predictions. RMSE can be used with various features to determine if a feature is improving the model's predictions[20].The formula for RMSE is as follows:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}{N}}$$

RESULTS

Data Visualisation and Analysis

The Figure 3-pair plot displays a grid of axes displaying pipe fitter, Argon Welder, and CS Welders by country. The pair plots created a grid of axes such that each numeric variable in the data will be shared on the y-axis across a single row and on the x-axis across a single column. The diagonal axes are treated differently, drawing a plot to show the univariate distribution of the data for the variable in that column.

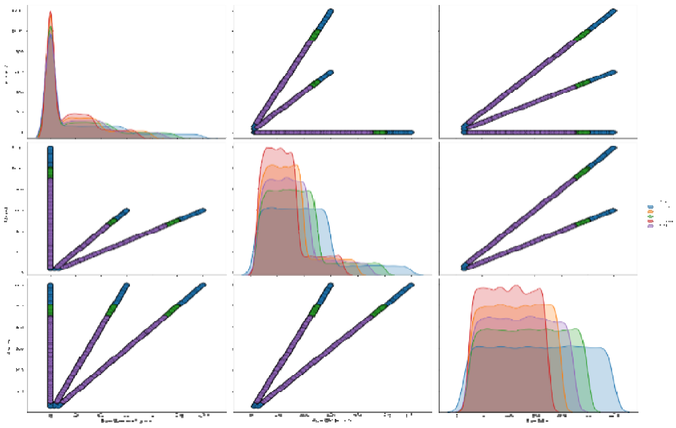

Figure 2. Pair Plot 1

A pair plot illustrating the relationship between Cranes, Riggers, and Grinders per Country is depicted in Figure 3 (pair plot 2).

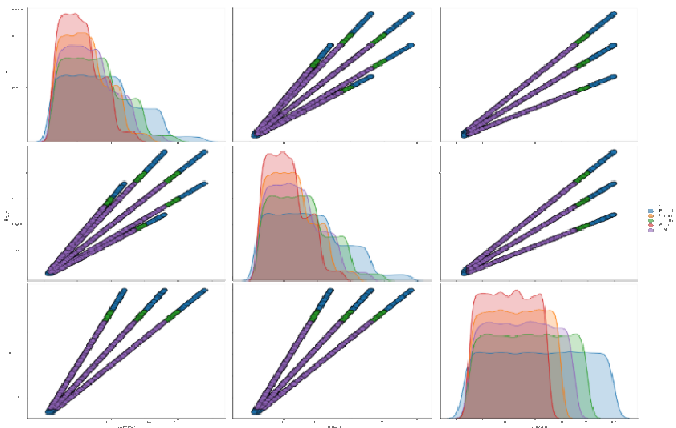

Figure 3. Pair Plot 2

Figure 4-pair plot3 exhibits a pair plot analysing the correlation between Erection and Daily Rate per Country.

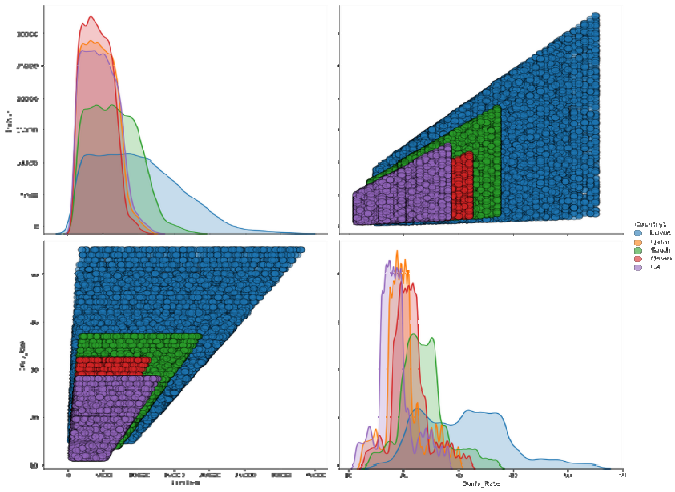

Figure 4. Pair Plot 3

According to the Figure 4-pair plot presented above, the production of erections in Egypt which exceeds 20,000 D.I./day, accounts for around 5% of Egypt's total record, while the production of erections over 10,000 D.I./day accounts for approximately 45%. The average is 9,775 D.I./day, with maximums of 32,945 D.I./day and minimums of 465 D.I./day.
Similar results were obtained regarding the daily rate of erection. With an average per crew of 31 DI/day, a high of 55 DI/day, and a low of 15 DI/day, Egypt has a large standard deviation. More than 10,000 D.I./day of erections are produced in the Gulf region, accounting for around 13% of all Gulf country records, while more than 5,000 D.I./day is produced at a rate of approximately 89%. The average is 5,352 D.I./Day, with a high of 18,463 D.I./Day and a low of 610 D.I./Day. Similarly, Gulf nations have improved standards. The average variance per crew is 22 D.I./day, with minimum and maximum values of 11.5 and 37 D.I./day, respectively.

The average Erection production per year for each country is shown in Figure5:

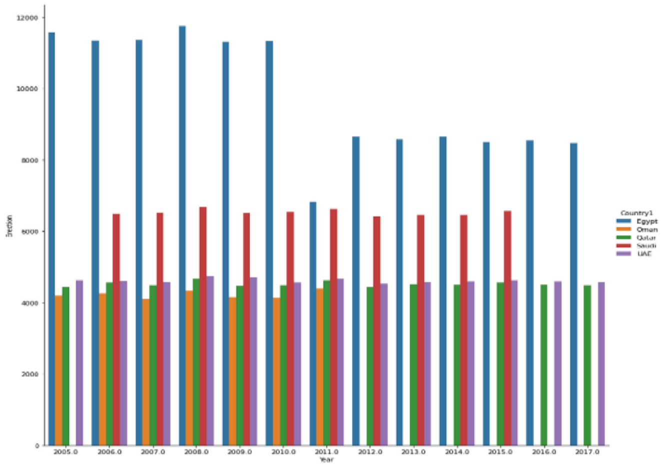

Figure 5. Average Erection/Year

There has been a drop in production in piping erection in Egypt starting from 2011 till 2017, although the average number of projects is almost the same within that period. Figure 6 below shows the average daily production rate per crew for each country.

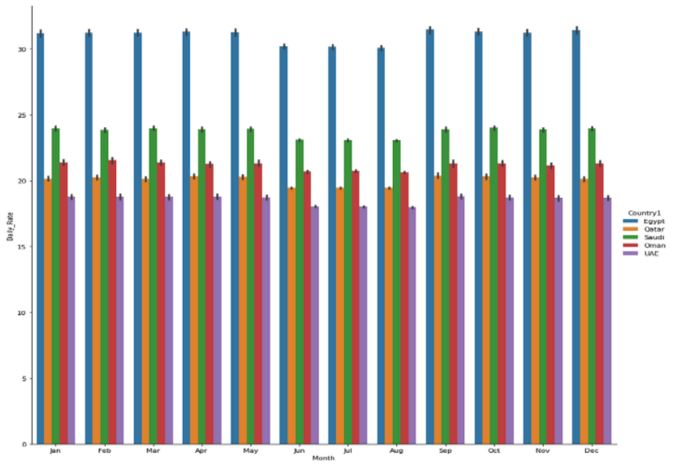

Figure 6. Average Daily Erection Rate Per Crew Per Country

The average daily production rate per crew in Egypt is around 31 D.I./Crew; in Oman, it is 21 D.I./Crew; in Saudi Arabia, it is 23.5 D.I./Crew; in Qatar, it is 20 D.I./Crew; and in the UAE it is 18.5 D.I./Crew, as shown in Figure 7.

In every country, crew productivity has decreased throughout the summer. The average daily production rate per crew annually for each nation is displayed in Figure 7 below.

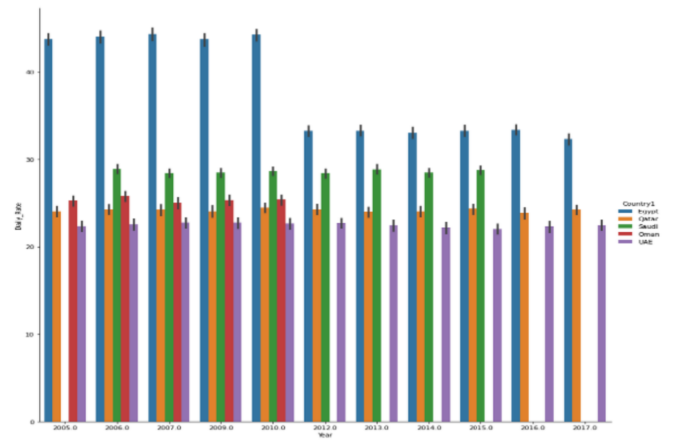

Figure 7. Average Daily Erection Rate Per Crew Per Year Per Country

The crew productivity has dropped in Egypt starting from 2011 till 2017. The average number of pipe fitters per year for each country is shown in Figure 8 below.

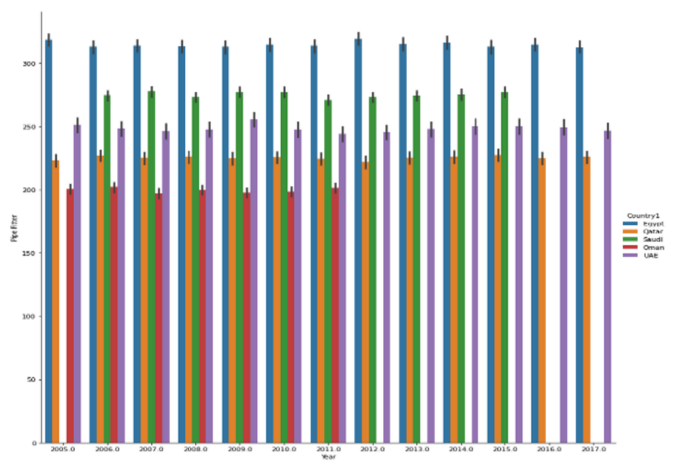

Figure 8. Average Pipe Fitters per Year for Each Country

The average number of cranes per year for each country is represented in Figure 9 below.

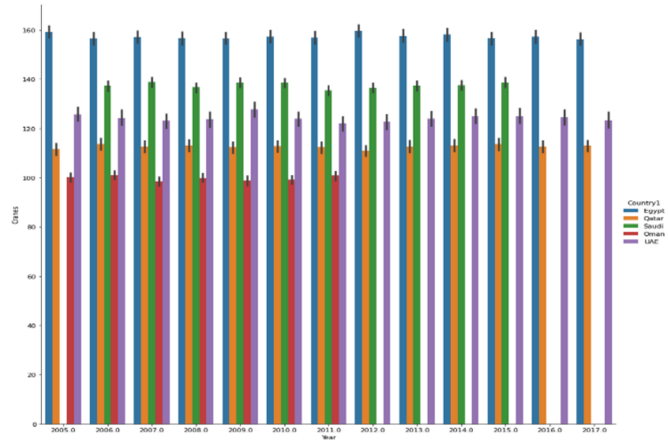

Figure 9. Average Number of Cranes Per Year Per Country

The average quantity of Pipe Welders (C.S) annually for each country is shown in Figure 10 below.

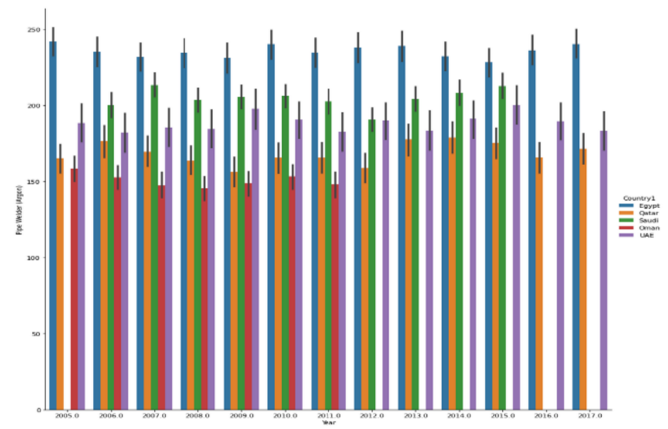

Figure 10. Average Numberof Welders Per Year Per Country

Model Features Importance

The impact of the model features was demonstrated on our target in Figure 11 Heat map. A heat map is a 2D graphical representation of data, depicting individual matrix values as colours, and revealing relationships between different data features [21].

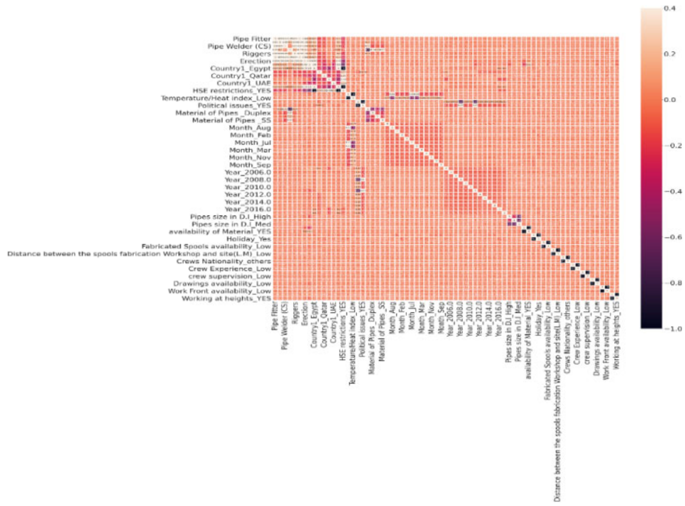

Figure 11. Heat Map

Features that impact the erection production are positively listed in descending order, as presented below in Figure 12.The skilled manpower, especially pipe fitters and equipment like cranes, the location like Egypt, the absence of HSE restrictions, the Low Heat index, the absence of holidays, and the availability of material and drawings all impact the erection production rate positively.

Pipe Fitter	0.868539
Cranes	0.868539
Grinder	0.773901
Riggers	0.772349
Pipe Welder (CS)	0.699907
Daily_Rate	0.574518
Country1_Egypt	0.450903
HSE restrictions_NO	0.450903
Pipe Welder (Argon)	0.319736
availability of Material_YES	0.097399
Year_2005.0	0.035286
Year_2008.0	0.032485
Pipes size in D.I_Med	0.030312
Political issues_NO	0.028038
Holiday_Yes	0.026085
Working at heights_NO	0.023672
Fabricated Spools availability_High	0.023546
Temperature/Heat index_Low	0.023347
Year_2006.0	0.017812
Year_2009.0	0.016451
Drawings availability_High	0.016303
Year_2007.0	0.016302
Year_2010.0	0.015916
Material of Pipes _CS	0.015312

Figure 12. Features Impact

Model Training and Testing

Linear Regression

Several models are used to train the data, and the outcomes are compared to the Benchmark Model. Since it is quick and easy to apply, the linear regression model serves as a standard by which to evaluate the performance of the other models. The researchers examined the RMSE as a statistic to compare the output of various models. Probably one of the most significant and popular regression approaches is linear regression. It is one of the most basic techniques for regression. Its simplicity in interpreting findings is one of its primary benefits. A linear connection between y and $\mathbf{x}$ is examined when doing linear regression of a dependent variable y on the set of independent variables $\mathbf{x} = (x_1, \ldots, x_r)$, where $\boldsymbol{U}$ is the number of predictors: $y = \boldsymbol{\lambda}_0 + \boldsymbol{\theta}_1 x_1 + \cdots + \beta_\mho x_\mho + \varepsilon$. The regression equation is this one. The regression coefficients are β_0, β_1, …, $\beta_\mho$, and the random error is ε[22].If the connection between the outcome variable and the covariates is known to be linear, then linear regression fits well. It moves the emphasis from pre-processing and data analysis to statistical modeling [23].

The RMSE after implementing the Linear Regressor is 1896.

Elastic Regressor

It is a Linear regression with combined L1 and L2 priors as regulariser a * L1 + b * L2, where: alpha = a + b and l1_ratio = a / (a + b)

The RMSE for Elastic Regressor is 1985.

Ridge Regressor

The ridge is a linear least square with l2 regularisation. This model solves a regression model where the loss function is the linear least squares function, and the l2-norm gives regularisation.

The RMSE for Ridge Regressor is 1896.

Lasso Regressor

It is a linear model trained with L1 prior as a regulariser. Technically, the Lasso model is optimising the same objective function as the Elastic Net with l1_ratio = 1.0 (no L2 penalty).
The RMSE for Lasso Regressor is 1896.
Features importance for the Linear, Ridge, Lasso and Elastic Regressors are represented in Figure 13below.The highest features are skilled labours like, Pipefitters, Grinders and Welders.

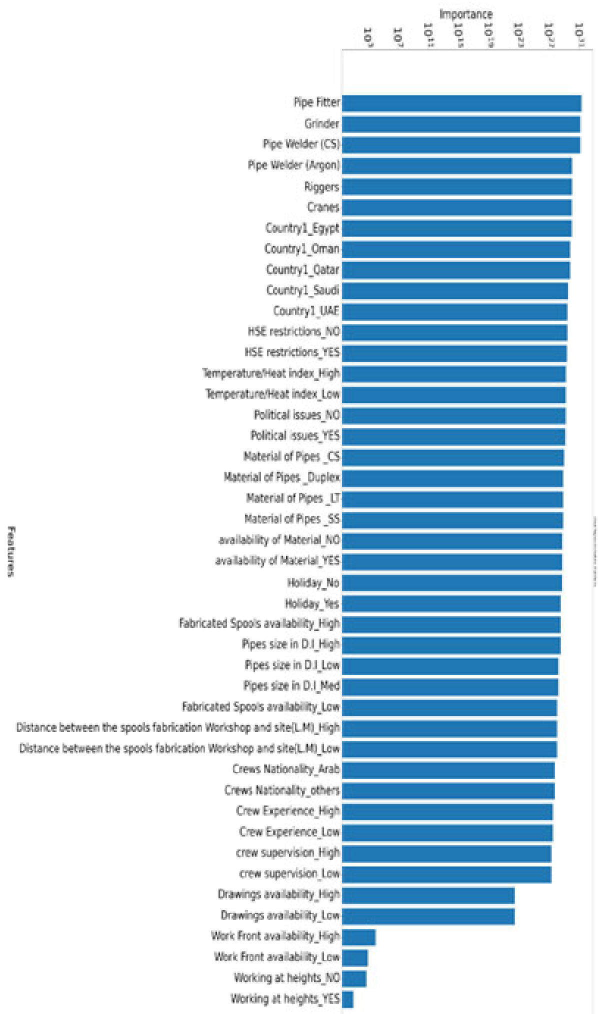

Figure 13. Features Importance for Linear, Ridge, Lasso and Elastic Regressors

Artificial Neutral Network (ANN)

ANN architecture consists of three main components: the input layer, which feeds training observations and specifies predictor variables through neurons; the hidden layers, intermediate layers between the input and output layers, learning about data relationships; and the output layer, where the final output is extracted from the previous two layers, with one neuron in the output for regression problems[24]. Figure 14 below depicts the ANN implementation.

The RMSE for ANN Model is: 1330

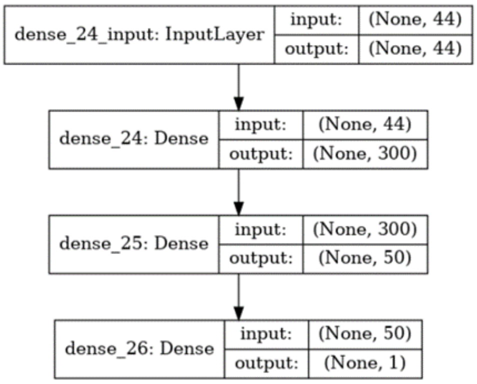

Figure 14. Artificial Neural NetworkModel Implementation

The highest features for ANN are skilled labours like, Pipefitters, Grinders and Welders (See Figure 15).

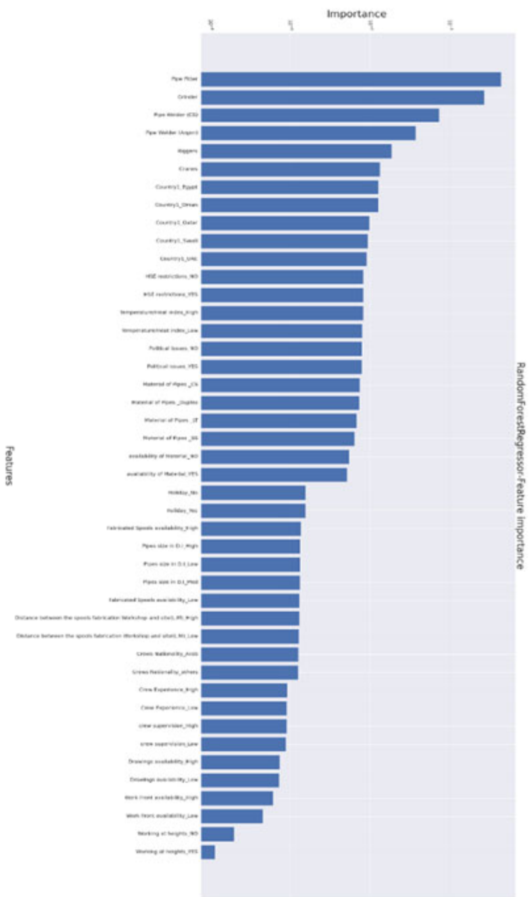

Figure 15: Features Importance for ANN

Random Forest (RM) Regressor

A random forest is a meta-estimator that fits many classification decision trees on various subsamples of the dataset and utilises averaging to increase predicted accuracy and reduce over-fitting. The original input sample size and the sub-sample size are always the same [25].
The RMSE for the RF Regressor Model is 1429

Cat Boost Regressor

Cat Boost (categorical data (the Cat) and it uses gradient boosting (the Boost)) regressor is an effective ML technique that is frequently used to solve a variety of commercial problems, including forecasting, fraud detection, and recommendation items. Additionally, it can produce exceptional outcomes with comparatively less data, in contrast to deep learning models that require a large quantity of data to be learned [26].

The Cat Boost regressor model's RMSE is 1341.

LGBM Regressor

LGBM is a distributed, high-performance gradient-boosting framework that is quick and efficient for a wide range of ML applications, including classification and ranking. It is based on the decision tree technique; therefore, it divides the tree with the greatest fit leaf-wise, as opposed to other boosting methods that divide the tree level-wise or depth-wise. Therefore, in LGBM, while growing on the same leaf, the leaf-wise approach can cut loss more than the level-wise technique and produces substantially superior accuracy—an outcome that is seldom possible with any of the boosting algorithms now in use [27].

The RMSE for the LGB regressor Model is 1330.

XGB (Extreme Gradient Boosting) Regressor Model

XGB is an implementation of gradient-boosted decision trees designed for speed and performance [28].

The RMSE for the XGB regressor Model is 1421.

Features importance for XGB, LGBM and CatBoostRegressors are shown in Figure16 below. The highest features are skilled labours like, Pipefitters, Grinders and Welders.

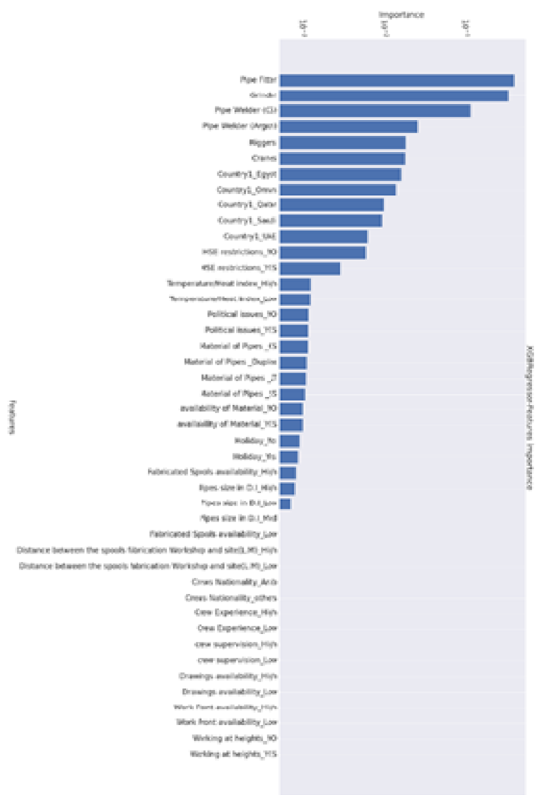

Figure 16. Features Importance for XGB, LGBM and Cat Boost Regressors

Model Evaluation

First, all of the models were merged,and the results using our metric, the RMSE, are shown in Table 2 below. Since we are working with a goal in thousands, the ANN model with the lowest RMSE (1330) is certainly the best. Second place comes to the Boosting Models (CatBoost, LGB, and XGB), whose RMSE (1341) is quite similar to that of the Deep Learning Model (ANN).

Table 2. Combined and Sorted RMSE

S.No.	Model	RMSE
1.	Deep Learning Model	1330.4
2.	CatBoostRegressor	1341.7
3.	LGB Regressor	1342.7
4.	XGB Regressor	1421.3
5.	Random Forrest Regressor	1429.7
6.	Lasso Regressor	1896.8
7.	Ridge Regressor	1896.9
8.	Linear Regressor	1896.9
9.	Elastic Regressor	1985.3

Furthermore, model implementation in the next section shows the outputof the model using new data.

Model Implementation

The model was implemented for different countries, and features are presented in this section. The term "Piping Erection/day"in this research isdescribed as the amount of effort that goes into erecting or installing pipe systems in a single day. This statistic assesses an individual's or a group's productivity or ability to install or set up pipe systems, frequently in industrial or construction operations. In this research, in projects requiring the construction of pipe systems, it measured the quantity or length of pipe installed within a given time frame, often per day, and is utilised as a performance or productivity indicator.

Case 1

Case 1 presents data on Egypt's piping erection rate, incorporating factors like country, HSE restrictions, temperature, political issues, pipe material, workforce, material availability, work front availability and crew experience.

Table 3. Case 1: Country "Egypt" with Model Features

S.No.	Features	Input to The Model	Predicted Output (Piping Erection/day)
1.	Country	Egypt	9,752.6 D.I/day
2.	HSE restrictions	NO	
3.	Temperature/Heat index	Low	
4.	Political issues	NO	
5.	Material of Pipes	CS	
6.	Pipes Diameter	Med	
7.	Availability of Material	High	
8.	Holidays	No	
9.	Number of Pipe Fitter	250	
10.	Number of Grinder	350	
11.	Number of Pipe Welder (CS)	400	
12.	Number of Pipe Welder (Argon)	100	
13.	Number of Riggers	350	
14.	Number of Cranes	125	
15.	Fabricated Spools availability	High	
16.	No of Inspectors for NDT Test	MED	
17.	Distance between the spools fabrication Workshop and site (L.M)	Low	
18.	Crews Nationality	Arab	
19.	Crew Experience	High	
20.	crew supervision	High	
21.	Drawing's availability	High	
22.	Work Front availability	High	
23.	Working at heights	NO	

The model predicts a daily output of 9,752.6 D.I/day (Table 3), indicating a rate of approximately 9,752.6 units of piping erection per day in Egypt under these conditions. The data is used to predict the piping erection rate in Egypt.

Case 2

In Case 2, Qatar's HSE restriction is marked as "YES", affecting the predicted output for piping erection per day. The estimated rate of piping erection per day in Qatar is now approximately 8,262 units, compared to 9,752.6 units in Case 1, where the HSE restriction was marked as "NO" (Table 4). This suggests that the implementation of HSE restrictions in Qatar has influenced the projected daily rate of piping erection.

Table 4.Case 2: Country "Qatar" with Model Features

S.No.	Features	Input to The Model	Predicted Output (Piping Erection/day)
1.	Country	Qatar	8,262 D.I/day
2.	HSE restrictions	YES	
3.	Temperature/Heat index	Low	
4.	Political issues	NO	
5.	Material of Pipes	CS	
6.	Pipes Diameter	Med	
7.	Availability of Material	High	
8.	Holidays	No	
9.	Number of Pipe Fitter	250	
10.	Number of Grinder	350	
11.	Number of Pipe Welder (CS)	400	
12.	Number of Pipe Welder (Argon)	100	
13.	Number of Riggers	350	
14.	Number of Cranes	125	
15.	Fabricated Spools availability	High	
16.	No of Inspectors for NDT Test	MED	
17.	Distance between the spools fabrication Workshop and site(L.M)	Low	
18.	Crews Nationality	Arab	
19.	Crew Experience	High	
20.	crew supervision	High	
21.	Drawings availability	High	
22.	Work Front availability	High	
23.	Working at heights	NO	

Case 3

Case 3 involves Qatar, with identical features to Case 2, except for a modification in the Heat Index. The high Heat Index leads to a significant decrease in the predicted output for piping erection per day, reducing the daily rate to approximately 5,577.5 D.I/day(Table 5). This decline indicates a significant decrease in production output during the summer season or in high heat index conditions. Therefore, the adverse impact of a high heat index is observable in terms of reduced daily piping erection rates in Qatar.

Table 5. Case 3: Country 'Qatar' with Change in the Heat Index to High

Case 4

The case 4 scenario in Qatar involves modifying the "Working at heights" feature to "YES", indicating that the crew now involves work at heights. This results in a reduced daily rate of piping erection, estimated at approximately 5,415.5 D.I/day (Table 6). This decrease in production output suggests that working at heights has a more significant impact on daily production rates, emphasising the potential challenges associated with working at heights in piping erection projects.

The model predicts a significant drop in daily production rates due to the increased workload. There are more drops in piping production in case crew are working at heights.

S.No.	Features	Input to The Model	Predicted Output (Piping Erection/day)
1.	Country	Qatar	
2.	HSE restrictions	YES	
3.	Temperature/Heat index	High	
4.	Political issues	NO	
5.	Material of Pipes	CS	
6.	Pipes Diameter	Med	
7.	Availability of Material	High	
8.	Holidays	No	
9.	Number of Pipe Fitter	250	
10.	Number of Grinder	350	
11.	Number of Pipe Welder (CS)	400	5,577.5 D.I/day
12.	Number of Pipe Welder (Argon)	100	
13.	Number of Riggers	350	
14.	Number of Cranes	125	
15.	Fabricated Spools availability	High	
16.	No of Inspectors for NDT Test	MED	
17.	Distance between the spools fabrication Workshop and site(L.M)	Low	
18.	Crews Nationality	Arab	
19.	Crew Experience	High	
20.	crew supervision	High	
21.	Drawings availability	High	
22.	Work Front availability	High	
23.	Working at heights	NO	

Table 6. Case 4: Country 'Qatar' with Change in Working at Height to (YES)

S.No.	Features	Input to The Model	Predicted Output (Piping Erection/day)
1.	Country	Qatar	
2.	HSE restrictions	YES	
3.	Temperature/Heat index	High	
4.	Political issues	NO	
5.	Material of Pipes	CS	
6.	Pipes Diameter	Med	
7.	Availability of Material	High	
8.	Holidays	No	
9.	Number of Pipe Fitter	250	
10.	Number of Grinder	350	
11.	Number of Pipe Welder (CS)	400	
12.	Number of Pipe Welder (Argon)	100	5,415.5 D.I/day
13.	Number of Riggers	350	
14.	Number of Cranes	125	
15.	Fabricated Spools availability	High	
16.	No of Inspectors for NDT Test	MED	
17.	Distance between the spools fabrication Workshop and site (L.M)	Low	
18.	Crews Nationality	Arab	
19.	Crew Experience	High	
20.	crew supervision	High	
21.	Drawings availability	High	
22.	Work Front availability	High	
23.	Working at heights	YES	

DISCUSSION

Industrial modular construction produces better-quality goods and increases efficiency by assembling parts at a shop before shipping them to construction sites. This procedure uses pipe spools, which are prefabricated pipe segments that enable quicker on-site assembly. However, a number of variables, including labour productivity, capacity of the shop, loading circumstances, and material availability, affect the estimation of fabrication time [29]. Consequently, the purpose of the current research was to help production managers plan how to provide materials and staff. The deep learning ANN model was found to be the most effective in predicting piping erection per day, outperforming other ensemble methods like CatBoost, LGBM, and XGB. The ANN model significantly outperformed advanced regression models in predictive accuracy, suggesting that the complexity and non-linearity of the dataset may have been better captured by its architecture. Notably, according to Chakraborty et al.(2020), using information from previous projects, researchers have created ANN models that generate accurate and timely cost estimates. Besides, with historical data on significant building expenses, they use the CCI for concrete structures. It demonstrated how well ANN and regression models performed in comparison when it came to forecasting simulated cost contingencies brought on by price fluctuations in steel reinforcing. Building owners and decision-makers may use a regression model for cost forecasting for mid-rise green office buildings to estimate development costs, weigh them against traditional solutions, and choose the most appealing one[30]. Similarly, as mentioned by Kulkarni et al.(2017), cost, time, quality, and safety are all areas where construction management is uncertain, which makes the construction process extremely unpredictable. ANNs are used to interpret unclear data and draw insightful conclusions. It showed the efficient use of ANNs in construction-related tasks, including forecasting costs, safety, risks, tender bids, labour, and equipment productivity, which is helpful in accurately understanding insufficient input data[31]. In the similar manner, the results of the current research suggested that deep learning models hold significant promise in improving predictive accuracy for estimating piping erection per day, underscoring the potential of sophisticated ML techniques. The cases presented in this research based on predictive ML modelling provide a detailed analysis of the predicted daily rates of piping erection in Egypt and Qatar under different conditions. Egypt's daily piping erection rate is high, with a predicted output of approximately 9,752.6 D.I/day. However, the implementation of HSE restrictions in Qatar leads to a significant decrease in the daily output, dropping to approximately 8,262 D.I/day. High Heat Index in Qatar also significantly impacts the daily output, dropping to about 5,577.5 D.I/day. The introduction of "Working at heights" in Qatar further reduces the estimated daily rate, dropping to around 5,415.5 D.I/day. These cases highlight the importance of regulatory compliance, environmental conditions, and working conditions in optimizing production output in construction projects. The findings underscore the need for careful consideration and management of various parameters to optimise production output in construction projects.

Limitations and Future Implications

The current study limits in terms of test data and the dynamic nature of construction projects. Additionally, it only addresses the prediction estimation of commodities in the Middle East, mainly Egypt and Qatar. In the future, the research can be extended in terms of data and countries. Additionally, the integration of ML and AI in the construction industry holds significant potential for improving cost estimation, particularly for commodities and labour productivity. In the future, these technologies can analyse vast datasets, including historical project data, market fluctuations, and labour trends, enabling the creation of predictive models that provide a more nuanced understanding of project costs by many construction companies in the Middle East. This allows construction professionals to make informed decisions, anticipate budgetary challenges, and enhance project financial management.

Conclusion

The application of ML and AI in cost estimation not only addresses historical uncertainties but also contributes to the industry's adaptability, fostering a proactive and resilient approach to project planning. This shift is expected to lead to a more efficient and sustainable construction landscape. Most models prioritise skilled labour, project location, HSE restrictions, temperature/heat index, political issues, and material type. AI and ML are crucial for future success, and pioneers of the future consider these in decision-making and management techniques. Historical data is essential for building effective models, and companies should collect daily reports, areas of concern, manpower reports, equipment reports, accidents, incidents, and daily temperatures. This research can be applied to various trades like shuttering, steel fixing, piping fabrication, equipment installation, concrete pouring, and steel erection, allowing for a standard rate for all trades.

REFERENCES

1. Gledson, B.J. and D. Greenwood, The adoption of 4D BIM in the UK construction industry: an innovation diffusion approach. Engineering, *Construction and Architectural Management,* 2017. 24(6): p. 950-967.
2. Abioye, S.O., et al., Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering,* 2021. 44: p. 103299.
3. Baduge, S.K., et al., Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction,* 2022. 141: p. 104440.
4. Oh, M., et al., Integrated system for BIM-based collaborative design. *Automation in construction,* 2015. 58: p. 196-206.
5. Park, D. and S. Yun, Construction Cost Prediction Using Deep Learning with BIM Properties in the Schematic Design Phase. *Applied Sciences,* 2023. 13(12): p. 7207.
6. Ikudayisi, A.E., et al., Integrated design process of green building projects: A review towards assessment metrics and conceptual framework. *Journal of Building Engineering,* 2022. 50: p. 104180.
7. Ghadbhan Abed, Y., T.M. Hasan, and R.N. Zehawi, Machine learning algorithms for constructions cost prediction: A systematic review. *International Journal of Nonlinear Analysis and Applications,* 2022. 13(2): p. 2205-2218.
8. Van, T.N. and T.N. Quoc, Research trends on machine learning in construction management: a scientometric analysis. *Journal of Applied Science and Technology Trends,* 2021. 2(03): p. 96-104.
9. Ning, F., et al., Manufacturing cost estimation based on a deep-learning method. *Journal of Manufacturing Systems*, 2020. 54: p. 186-195.
10. Uddin, S., S. Ong, and H. Lu, Machine learning in project analytics: A data-driven framework and case study. *Scientific Reports,* 2022. 12(1): p. 15252.
11. Dixit, M.K., Embodied energy analysis of building materials: An improved IO-based hybrid method using sectoral disaggregation. *Energy*, 2017. 124: p. 46-58.
12. Ma, M., et al., A systematic literature review on price forecasting models in construction industry. *International Journal of Construction Management,* 2023: p. 1-10.
13. Velumani, P. and N. Nampoothiri, Volatility forecast of CIDC Construction Cost Index using smoothing techniques and machine learning. *International Review of Applied Sciences and Engineering*, 2021. 12(1): p. 50-56.
14. Pham, T., T. Le-Hong, and X. Tran, Efficient estimation and optimization of building costs using machine learning. *International Journal of Construction Management,* 2023. 23(5): p. 909-921.
15. Sammour, F., et al., Forecasting demand in the residential construction industry using machine learning algorithms in Jordan. Construction Innovation, 2023.
16. Rafiei, M.H. and H. Adeli, Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of construction engineering and management,* 2018. 144(12): p. 04018106.
17. Jung, S.-H., O.-B. Gwon, and J.-H. Son, A Study on the Analysis and Estimation of the Construction Cost by Using Deep learning in the SMART Educational Facilities-Focused on Planning and Design Stage. *Journal of the Korean Institute of Educational Facilities,* 2018. 25(6): p. 35-44.
18. Jo, Y.-H. and S.-H. Yun, Analysis of impact factors for the improvement of conceptual cost estimation accuracy for public office building. *Journal of the Korea Institute of Building Construction,* 2021. 21(5): p. 495-506.
19. Al Janabi, M.A., Optimization algorithms and investment portfolio analytics with machine learning techniques under time-varying liquidity constraints. *Journal of Modelling in Management,* 2022. 17(3): p. 864-895.
20. Hale, J., Which Evaluation Metric Should You Use in Machine Learning Regression Problems? Towords Datascience, 2020.
21. Gu, Z., Complex heatmap visualization. Imeta, 2022. 1(3): p. e43.
22. Ray, S. A quick review of machine learning algorithms. in 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). 2019. IEEE.
23. Maulud, D. and A.M. Abdulazeez, A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends,* 2020. 1(4): p. 140-147.
24. Schmidhuber, J., Deep learning in neural networks: An overview. Neural networks, 2015. 61: p. 85-117.
25. Schonlau, M. and R.Y. Zou, The random forest algorithm for statistical learning. *The Stata Journal,* 2020. 20(1): p. 3-29.
26. Prokhorenkova, L., et al., CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems,* 2018. 31.
27. Alzamzami, F., M. Hoda, and A. El Saddik, Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation. *IEEE access*, 2020. 8: p. 101840-101858.
28. Chen, T. and C. Guestrin. Xgboost: A scalable tree boosting system. in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.
29. Mohsen, O., C. Petre, and Y. Mohamed, Machine-Learning Approach to Predict Total Fabrication Duration of Industrial Pipe Spools. *Journal of Construction Engineering and Management,* 2023. 149(2): p. 04022172.
30. Chakraborty, D., et al., A novel construction cost prediction model using hybrid natural and light gradient boosting. Advanced Engineering Informatics, 2020. 46: p. 101201.
31. Kulkarni, P., S. Londhe, and M. Deo, Artificial neural networks for construction management: a review. *Journal of Soft Computing in Civil Engineering,* 2017. 1(2): p. 70-88.

****