

Resource-Significant Activity Costing in Offshore Structure Construction Projects Using Artificial Neural Network [†]

Mofiyinfoluwa Tobi Olowe *  and Michael Ayomoh 

Department of Industrial and Systems Engineering, University of Pretoria, Pretoria 0028, South Africa

* Correspondence: u23980321@tuks.co.za

[†] Presented at the 1st International Online Conference on Designs (Designs 2026), 9–10 February 2026; Available online: <https://sciforum.net/event/Designs2026>.

Abstract

Fixed-bottom or floating offshore structures are the foundations, platforms, and associated infrastructure that allow for oil and gas production systems, offshore wind turbines, and cabling. The remote nature of these structures and the harsh environment with high variability in wind, waves, currents, and weather make construction activity very difficult and unpredictable; the cost of variation in the schedule can lead to high construction vessel and personnel costs. The adoption of artificial intelligence using trends observed in historical data can help achieve more accurate construction costs and schedule predictions, reducing the capital expenditure cost of installation. A resource-significant activity, sometimes called a resource-critical activity or high-resource-demand activity, is an activity on a construction or project schedule that consumes a disproportionately large share of one or more resources compared with others. Plant Design Modelling (PDM) is a digital process that creates and manages a detailed 3D model of a building's physical and functional characteristics and semantic information, such as cost and schedule. PDM serves as a single source of truth for multidisciplinary activities and, therefore, serves as a rich data source for various construction applications, including project scheduling and cost estimation. Neural networks (NNs), a subset of machine learning algorithms inspired by the human brain, excel at identifying patterns in complex datasets and making predictions, such as forecasting costs based on non-linear relationships and historical trends. Data from an offshore structure modification project were extracted from Aveva's Everything PDM, focusing on installation activities to create a dataset for machine learning model training. The structured data extracted exhibit non-linear patterns; therefore, linear, regularised linear, robust linear, and the ensemble (tree-based) models and supervised neural network models with varied architecture and hyperparameter values were evaluated and compared. The best performance was obtained using the deep-optimised ANN model. The result obtained is consistent with previous studies. The neural network models show a superior ability to predict the non-linear nature of offshore construction activities' time.

Keywords: offshore construction; offshore structures; Plant Design Modelling; Construction Cost Estimation; Artificial Neural Networks



Academic Editor: Elena Lucchi

Published: 7 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The construction industry is labour-driven (Golnaraghi et al., 2019) [1] and plays a significant role in national infrastructure and the economy [2]. The ref. [3] survey revealed

that the construction industry accounts for approximately 13% of the world's Gross Domestic Product (GDP), with a projection of reaching 15% in 2020. The construction industry creates about 5–10% of the global workforce [4].

Unlike in manufacturing where the production environment is controlled, construction projects are characterised by variability owing to many moving parts at the construction site. This leads to a highly dynamic environment, which causes disruptions, obstructions, and safety issues, as well as variations in costs and schedules.

In the construction industry, cost overruns due to schedule delays and rework result from inaccurate schedule and cost estimates during the planning phase. Labour costs account for approximately 30–50% of the overall construction cost. Therefore, the inability to properly capture, access, and respond correctly to the variabilities that arise from the dynamic and remote nature of the construction and installation site will have a massive impact on the overall project costs.

Industrial Construction Projects (ICPs) are large-scale, high-cost, and complex projects that often require a high level of technological input. While most of these projects are onshore, increasing opportunities to harness natural resources such as wind, gas, and crude oil are being enabled by advanced exploration technologies [5].

According to [6], energy employment as a share of global employment was approximately 2% (approximately 65 million people) in 2019, and approximately 11.8 million people worked in the oil and gas supply industry. In 2024, the global trade of Crude Petroleum reached \$1.31 T. Crude Petroleum ranked first in global trade value, accounting for 5.71% of world trade [7,8].

Offshore structures, also known as marine structures, are important components of modern energy systems. These structures aid in the exploration, production, and distribution of oil and gas, the ocean's renewable energy through wind farms, mariculture, and subsea installations.

The planning and implementation of offshore construction/installation projects are complex, because offshore structures are located in remote and harsh environments. Offshore construction environments are characterised by high variability in wind, waves, currents, and weather [9], which makes construction activities difficult and unpredictable.

The unpredictability of weather and other environmental factors in offshore construction results in the inability to obtain an accurate cost estimate at the bidding stage. This also makes it possible to estimate the impact of environmental and resource variations on the construction schedule and cost vis-à-vis the project baseline during construction and installation works.

Plant Design Modelling (PDM) is a digital process that creates and manages a detailed 3D model of a building's physical and functional characteristics and semantic information, such as the cost and schedule. PDM serves as a single source of truth for multidisciplinary activities and is therefore a rich data source for various construction applications, including project scheduling and cost estimation.

The adoption of artificial intelligence using trends observed in historical data can help achieve more accurate construction costs and schedule predictions, thereby reducing the capital expenditure cost of installation [1]. Neural networks (NNs), a subset of machine learning algorithms inspired by the human brain, excel at identifying patterns in complex datasets and making predictions, such as forecasting costs based on non-linear relationships and historical data [10,11]. Artificial Neural Networks (ANNs) can correctly establish the relationships between non-linear dependent and independent variables in construction data.

This study aims to develop a predictive Artificial Neural Network (ANN) model for the resource-significant Hook-up and Commissioning (HuC) activity using data from

the Ariosh construction planning repository. To achieve the objective, the following were carried out:

1. A desk study and exploratory data analysis of the Hook-up and Commissioning (HuC) activity construction planning repository were performed to determine the resource-significant activity and generate a dataset for training and testing the models.
2. Statistical and ANN models were trained and tested on the generated dataset.
3. Evaluation and selection of the model with satisfactory metric scores—R2, MAE, and RMSE.

The model developed will help construction business development professionals and project managers during project bidding and construction management.

Offshore construction projects have several stages. Conceptual Design, Front-End Engineering Design (FEED), Detailed Engineering Design, Procurement, Construction and Fabrication, Transportation and Installation, Hook-up and Commissioning (HuC), and start-up. The scope of the research is limited to the Hook-up and Commissioning scope of the project data in the project Job Card Monitoring System (JCMS). The research is then further delineated to include only resource-significant activity in the JCMS.

This study analysed offshore construction management and planning information, data, and databases in the JCMS and other construction planning models, databases, and documents. A dataset was created and used to train and test the models.

1.1. Hook-Up and Commissioning (HuC)

Offshore construction work includes the pre-commissioning of systems in the project, which is typically carried out in a controlled environment onshore, followed by the loadout of the components, equipment, and structures which make up the systems to the jetty, from where they are transported to the offshore site. These components, equipment, and structures are then lifted and hooked up in place to their position in the case of a greenfield installation or to an existing system in the case of a revamp or extension of a brownfield project. The coupled systems are then commissioned before the offshore start-up. The installation components, equipment, and structures may include skids, piping, flowlines, and compressors, among others, in the existing facility in the case of brownfield projects.

HuC works in offshore construction, involving installation tasks that occur between the fabrication of offshore structures and equipment modules and the start-up of installations. HuC accounts for approximately fifty percent of the total project cost [12].

1.2. Resource-Significant Activity

A resource-significant activity is an activity in a construction or project schedule that consumes a disproportionately large share of resources compared to others. The developed ANN model in the research is limited to piping features such as piping material, nominal pipe size, pipe schedule, and spool length available in the JCMS.

1.3. Labour Man-Hour Estimation

The impact of cost overruns can be evaluated using the triple constraint model. The three constraints are scope (project requirements and deliverables), time (project schedule and duration), and cost (project budget for resources). The triple constraints model states that cost is a function of time and scope; therefore, one of the constraints cannot change without corresponding trade-offs from the others [13]. In a typical offshore structure project, material costs are the largest share of construction costs, followed by labour-related costs. Between 30% and 60% of the cost of a construction project goes to the labour component of the cost. Labour costs are mostly estimated using man-hours [14,15]. In this study, we focused on the labour man-hours of the construction cost, as this is the productivity

component that can be managed internally. Other components, such as labour wages, are not always controlled by internal factors but by unions, regulations, etc.

Labour costs are divided into direct and indirect costs. Direct costs are incurred by construction personnel, whereas indirect costs are used to manage construction personnel. Indirect costs are associated with government and regulatory levies and taxes, insurance, retirement, leave, training costs, and professional membership payments. Labour costs involved establishing the base man-hours and cost/hour for the disciplines required to carry out the tasks. To account for indirect costs, a labour burden multiplier was included (Labour burden = Indirect costs/Direct costs). A productivity multiplier is also included to account for the team's productivity level in performing the tasks:

$$\text{Labour cost, } y = (1 + c) \times a \times d \quad (1)$$

where c = labour burden = (a/b) ; a = direct cost = $(e \times f)$; d = labour productivity rate; b = indirect cost; e = man-hours; and f = labour cost/hour.

In modelling and predicting construction project man-hours, many evaluation techniques ranging from statistical to machine learning approaches have been studied. Predicting man-hours using traditional approaches is very slow and often inaccurate [1]. Several researchers have studied the application of machine learning in construction [11]. Based on modelling ease and prediction accuracy, the use of supervised ML algorithms, such as Artificial Neural Networks, has proven to be more effective than statistical regression techniques.

Ref. [16] posited that ML enhances site surveillance and automated detection, such as road surveys, crack detection, detection of people, real-time prediction, bridge inspection, sewer inspection, structural health monitoring, and site operations monitoring, reducing the demand on construction personnel and increasing safety. Ref. [17] conducted a comprehensive review of machine learning and artificial intelligence techniques that have been applied to solve problems in the oil and gas industry and concluded that they enhance decision-making and reduce the risk and cost of maintenance.

Ref. [18] was the first to explore the use of ANN. They developed an ANN model to estimate the markup under different bidding conditions. Ref. [19] adopted the back-propagation algorithm in a multilayer feed-forward network to measure the productivity of reinforcement installation in building projects in India. Ref. [20] trained three neural networks for formwork assembly, steel fixing, and concrete pouring activities.

Ref. [21] constructed two models for managing the repairs and maintenance of construction projects. They developed a cost-significance model using the weights of cost-significant items from the work packages. They compared the prediction of the cost-significance model with that of a feed-forward ANN using a backpropagation learning technique. They confirmed that "the artificial neural network may be an attractive model compared to cost significance work packages model." They posited that cost models can help managers with project estimation and budgeting.

Ref. [22] combined multivariate regression (MVR), a linear statistical model, and a multilayer feed-forward ANN method to develop a model that estimates shovel capital costs. They recommended the use of a hybrid model when modelling complex systems that generate datasets with at least one non-linear pattern. Ref. [23] developed an ANN model using a dataset constructed from 132 projects to evaluate the cost of engineering consultancy work. Ref. [24] predicted the man-hours of scaffolding activity using historical scaffolding data from a heavy industrial project. They compared the prediction outputs of algorithms such as multiple linear regression, decision tree regression, random forest regression, and Artificial Neural Networks (ANNs).

The aforementioned demonstrated that machine learning, especially Artificial Neural Networks, when applied to construction data, can help model the complex nature of

construction activities. This research, therefore, applies ANN to the offshore construction man-hour modelling using historical data, using the traditional statistical model as a benchmark.

The structure of the rest of this paper is as follows: Section 2 describes the model development and how archived historical data was collected, transformed, and used to train models for man-hour prediction. Section 3 presents the results of the linear, regularised linear, robust linear, and the ensemble (tree-based) models, which were first examined to establish a benchmark. The neural network models were trained and evaluated. Section 4 concludes on the appropriateness and the merit of ANNs for model development and evaluation. Section 4 also recommends further work to improve the research.

2. Model Development

This study was conducted in two broad stages: a repository data review and model training. Figure 1 provides an overview of this study.

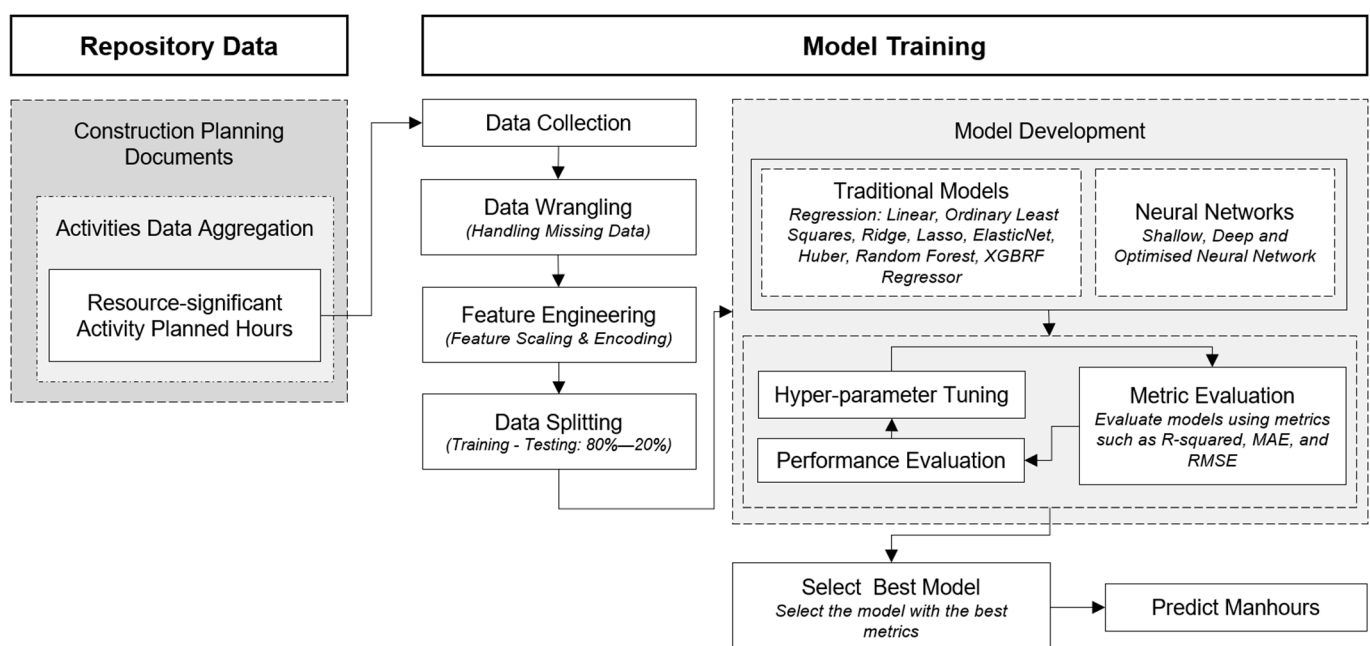


Figure 1. Research conceptual model framework.

2.1. Repository Data Exploration

Archived historical data are unstructured and messy. Therefore, they must be cleaned to remove noise and augmented to bring them to a useful state. In this study, information from Ariosh Limited's construction planning repository was used. The repository includes construction planning documents, the Aveva E3D model and database, Detailed Engineering Design documents and drawings, the Job Card Management System (JCMS), installation procedures and drawings, and Facility Information Management (FIM).

Aveva Everything 3D (version 3.1) (Aveva E3D), a Plant Design Modelling (PDM) software, serves as a single source of truth for multidisciplinary activities in the evaluated project. It serves as a rich data source for Virtual Design and Construction, installation sequencing, project scheduling, and cost estimation. The information extracted from Aveva E3D includes the geometry, material types, quantities, and spatial relationships between elements. Table 1 details the data and data format of the archived construction planning documents.

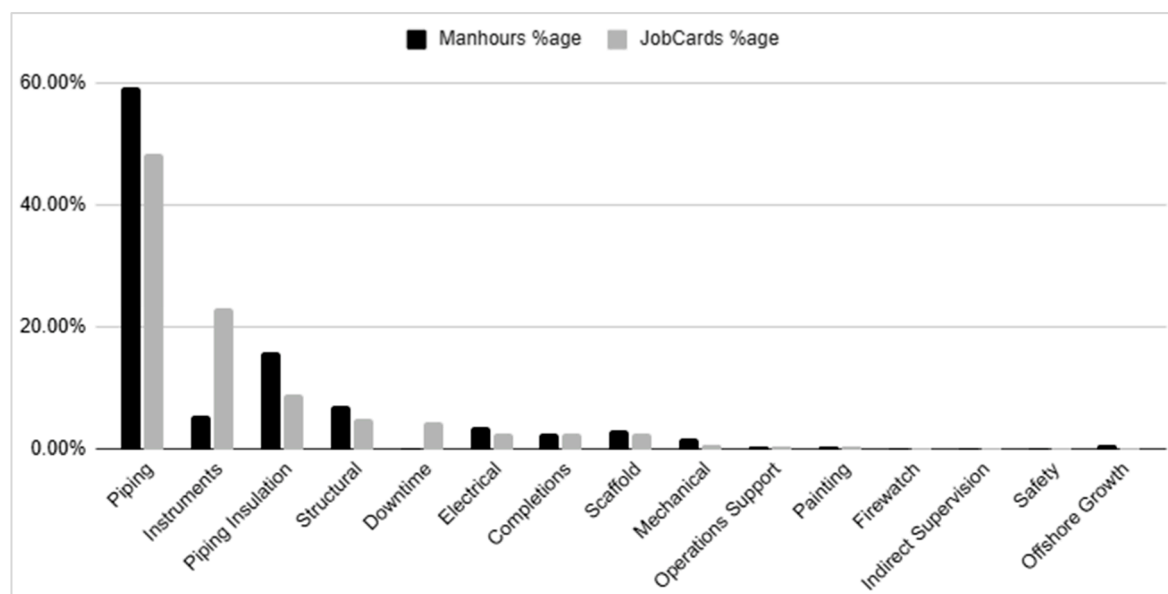
Table 1. Data and data format of the archived construction planning documents.

S/No.	Data	Data Format
1.	Facility Information Management (FIM)	Aveva E3D models extracted to Naviswork (nwd) format Piping isometrics (generated PDF drawings from the Aveva E3D model)
2.	Detailed Engineering Design Drawings	AutoCAD (version 24.1) drawings (generated from Aveva E3D model)
3.	Installation Procedures	Microsoft Word documents AutoCAD drawings (generated from the Aveva E3D model)
4.	Installation Storyboard Drawings	AutoCAD drawings (generated from Aveva E3D model)
5.	Job Card Management System	Microsoft Access database
6.	Basis of Estimate	Microsoft Excel documents

2.2. Data Extraction

The extracted information includes the geometry, material types, quantities, and spatial relationships between elements. The dataset used for training the statistical, machine learning, and neural network models was extracted from the information in the documents, database, and 3D model. Data preparation accounts for a significant percentage of the overall work.

The scope of the research is limited to the Hook-up and Commissioning scope of the project data in the project Job Card Monitoring System (JCMS). The research is then further delineated to include only resource-significant activity (spool installation) in the JCMS. JCMS is a Microsoft Access database which contains work packages (work packs) for the Engineering, Procurement, Construction, and Installation (EPCI) projects. The work packs in the JCMS are subdivided into job cards based on the individual project's work breakdown schedule (WBS). The distribution of disciplines for the project being evaluated, as found in the JCMS database, by job cards and man-hours, is shown in Figure 2.

**Figure 2.** Database's discipline distribution by job cards and man-hours.

Each job card contains the planned man-hours for all activities involved in executing the activities involved. Job cards were prepared for all discipline activities. Each job card

contains information about the man-hours, materials, and reference documents required for each work activity, as well as the trade responsible for the activities.

The JCMS has 506 activities across different disciplines. The piping discipline accounted for 48.42% of the piping activities and 59.42% of the project's man-hours. The analysis indicates that piping work is a resource-significant discipline. Therefore, the piping discipline is the focus of this study. This confirms the observation of [25] that piping installation is a resource-intensive activity.

2.3. Resource-Significant Activity—Piping Installations

Piping installations convey fluid or gas between pipe-to-pipe, pipe-to-equipment, or open drain. The materials used in the project under study included carbon steel, light carbon steel, and stainless steel. The piping installation man-hours are affected by the length of the spool, the impact of height installation, and the difficulty of installation due to density in the final installation area. During the constructability reviews, pipes with long route lengths were reduced to manageable lengths (i.e., spools). The height of installation and density of installation location were not found to be factors that could impact installation man-hours, as scaffolds were erected to level out the height differences. In scenarios where this cannot be achieved, field welds are introduced to the spool. The JCMS project has 245 piping activities. Spool installation accounted for 40.41% of the piping activities and 66.40% of the piping man-hours. Therefore, it can be concluded that the spool installation activity is the most resource-significant.

Therefore, the focus of this research was further narrowed to deal with the 99 spools installation job cards. Piping spool installation man-hours are influenced by the material types, piping grades, pipe bore size (diameter of the pipe), piping schedule (wall thickness), the pressure rating of the pipe and flanges, and the location of installation (offshore or onshore). Figure 3 shows the material counts and the pipe bore sizes by material type, while Figure 4 further shows the breakdown by piping material and pipe bore sizes.

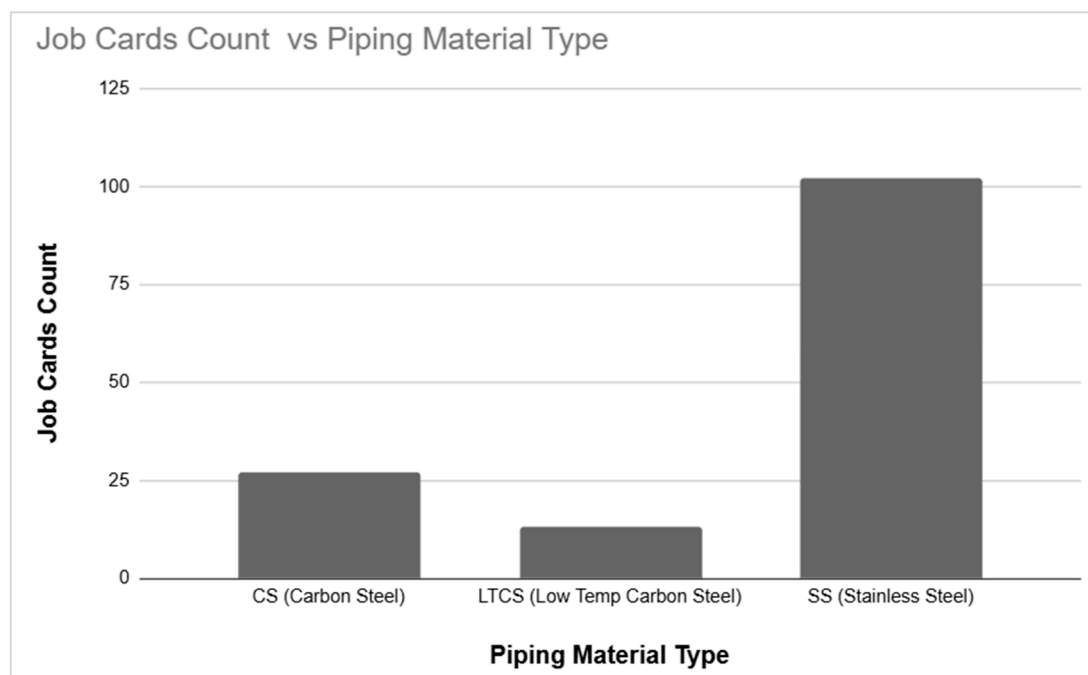


Figure 3. Spool installation observations by material type.

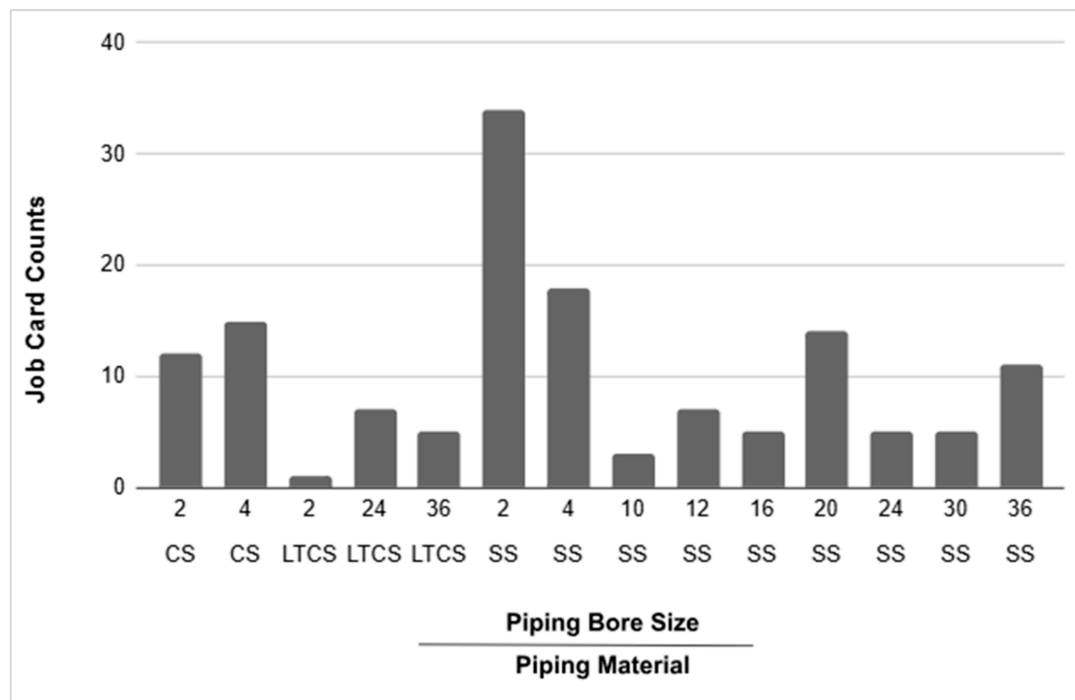


Figure 4. Spool installation observations by pipe bore sizes by material type.

2.4. Research Scope—Limitation

During desktop review and data exploration, it was observed that only a few pipe bore, pipe material, and spool length categories were present in the JCMS; these are as shown in Figure 4. Therefore, the models developed and evaluated are limited to the piping features, such as the piping material, nominal pipe size, pipe schedule and Spool length available in the JCMS. The model may, therefore, not be effective in predicting parameters outside the ranges not included in the JCMS. Table 2 highlights the piping features in the general piping standard compared to the piping features in the repository used to generate the dataset for model training.

Table 2. The standard general piping features compared to the ones in the dataset.

S/No	Feature Description	Generally Available in Standards	Included in Dataset
1.	Nominal Pipe Sizes	1/8", 1/4", 3/8", 1/2", 3/4", 1", 1 1/4", 1 1/2", 2", 2 1/2", 3", 3 1/2", 4", 5", 6", 8", 10", 12", 14", 16", 18", 20", 24", 30", 36", 42", 48"	2", 4", 10", 12", 16", 20", 24", 30", 36", 40"
2.	Pipe Schedules	5, 5S, 10, 10S, 20, 30, 40, 40S, 60, 80, 80S, 100, 120, 140, 160, STD, XS, and XXS	20, 40, 40S, 60, 80, 160, STD, and large outer diameter (OD) pipes (30", 36", 40") with wall thicknesses (WT) of '0.750 WT', '0.875 WT', '1.25' WT
3.	Material	Carbon Steel, Light Carbon Steel, Stainless Steel, Cast Iron, Plastic, Copper, etc.	Carbon Steel, Light Carbon Steel, and Stainless Steel
4.	Spool Length	0–40 feet (0–480 inches)	0.960–58.84 inches

2.5. Data Exploration—Dataset Attributes

Using the information and references on the job cards, the 3D BIM model, isometrics, installation procedure document, and estimating sheets were examined.

There are 99 spool installation job cards. The unbundled job cards produced 149 spools. A total of 149 spool features were aggregated using their piping isometrics, the 3D model, and estimate sheets in the JCMS.

The piping installation activities in the JCMS have 142 spools with flanges at both ends (Figure 5), while a total of seven spools have a service “DO—Open Drain”, which have just one flange and the other end is open to the air (Figure 6).

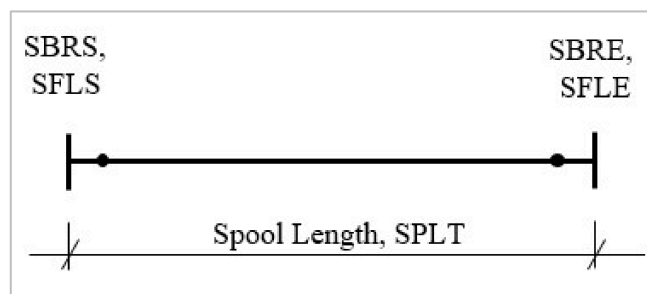


Figure 5. Spools with a flange on both ends.

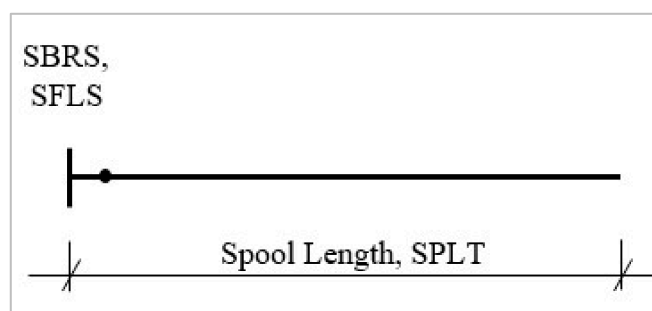


Figure 6. Spools with flanges on one end.

The seven spools with service “DO—Open Drain” were removed from the dataset, as the lack of piping parameters for the open end of the spools resulted in “NaN” values in the dataset. The 142 spool observations became the dataset used for developing the statistical and ANN models.

A total of 142 data observations (of spool installation) were shortlisted for developing the linear and ANN models. Table 3 highlights the dataset variables, indicators, and data type. The dataset was split (80–20): 80% (113) for training and 20% (29) for testing.

Table 3. Dataset features and data type.

Variables	Spool Feature	Indicator	Data Types
Predictor variables	Pipe Material Type	MTRL	Categorical
	Pipe Schedule (wall thickness)	PSCH	Categorical
	Spool Bore Diameter—Start	SBRS	Numerical
	Spool Flange Pressure Rating—Start	SFLS	Categorical
	Spool Bore Diameter—End	SBRE	Numerical
	Spool Flange Pressure Rating—End	SFLE	Categorical
	Spool Length	SPLT	Numerical
Response variable	Installation Man-hours	IMHR	Numerical

2.6. Data Exploration—Scatter Matrix for Numerical Features

The scatter matrix in Figure 7 provides a visual overview of the relationships between the numerical variables (SBRS, SBRE, SPLT, and IMHR). SPLT shows some positive correlation with IMHR but also some variability, indicating a non-linear relationship. IMHR exhibits a wide range of values, and its relationships with other features appear complex, with potential non-linear trends and varying distributions. There is a need to investigate IMHR further.

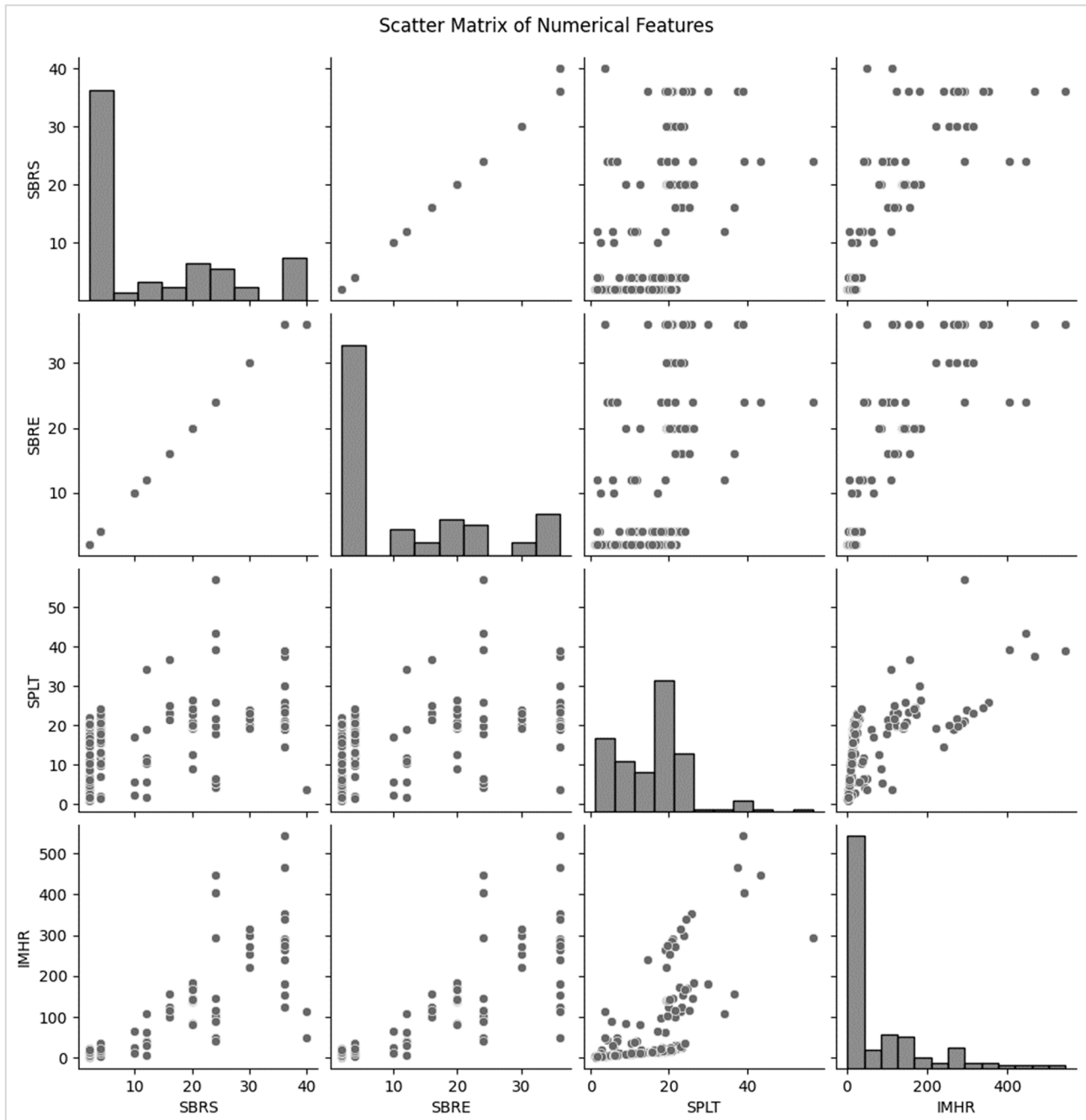


Figure 7. Scattered matrix for numerical features.

2.7. Exploring the Dependent Variable—Installation Man-Hours (IMHR)

The histogram and box plot of installation man-hours (IMHR) shown in Figures 8 and 9, respectively, confirm that it is a highly skewed distribution, with the presence of numerous outliers. The outliers were identified and examined: It was found that the outlier values

are real values that represent installation works on the project being examined; they are, however, associated with high man-hour values and few occurrences; we therefore elected to keep the values in the dataset. This heterogeneity, therefore, implies the need for robust models that are robust to outliers and capable of capturing non-linear patterns. This will be beneficial for accurate activity modelling prediction.

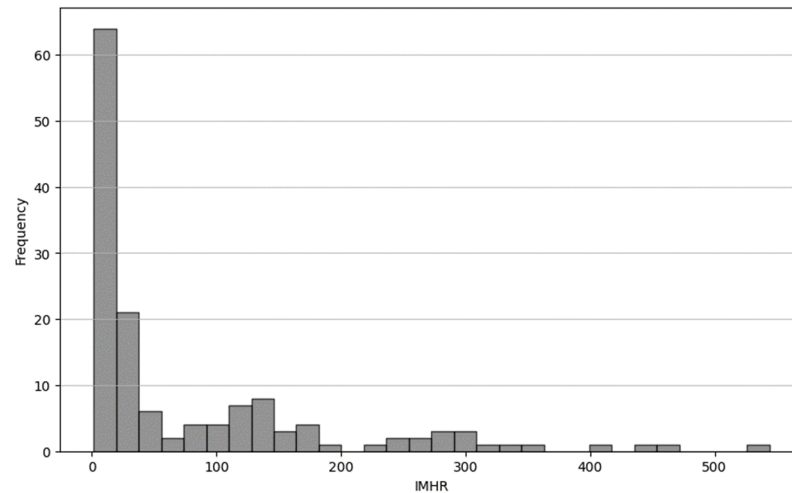


Figure 8. Histogram for Installation man-hours, IMHR.

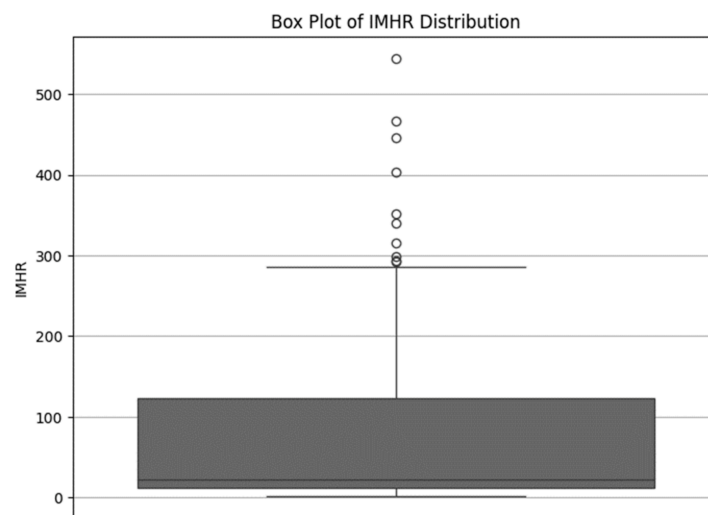


Figure 9. Box plot of installation man-hours, IMHR.

2.8. Model Development

Two methods have been used by researchers in the modelling of man-hours forecasting/prediction. To accurately predict the man-hour value for resource-significant activity, piping, various statistical approaches, and machine/deep learning (ML/DL) approaches were examined. The regression analysis works best where the relationship between the dependent and independent variables is linear, while ANN performs well where the relationship between the dependent and independent variables is non-linear or complex. For the ML/DL models, various ANN architectures and hyperparameters were interactively examined using R2, MAE, and RMSE metric scores.

The performance of the ANN model was compared with the linear, regularised, and robust regression models using the performance metrics. The model's performance was evaluated using R2, MAE, and RMSE. There was a training and testing phase (80–20%

training and testing data split). The models were developed in Python (version 3.14.6) using the Google Colab notebook environment.

2.9. Model Performance Evaluation Metrics

The metrics used to evaluate the models are:

$$\text{Mean absolute error, } MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2)$$

$$\text{Root mean squared error, } RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$\text{Coefficient of determination, } R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

where SS_{RES} = the residual sum of squared errors, SS_{TOT} = the total sum of squared errors of the model, x_i = the actual value, n = the number of data points, $\hat{y}$ = the predicted value of the data point, y = the actual value of the data point, and $\bar{y}$ = the mean of all the values of y .

2.10. Models Evaluated

Due to the presence of outliers in the dataset, the evaluation models must be robust to outliers and capable of capturing non-linear patterns. The sub-sections below highlight the merits, demerits, and relevance of the models to the work and dataset being evaluated and models selected.

2.10.1. Models Considered Characteristics

Linear Regression finds the “best-fitting” line through a set of independent and dependent values of a dataset. It serves as the fundamental baseline fit for the dataset. It assumes linearity and produces a set of coefficients for the independent variables that is easy to interpret [26,27].

Lasso Regression (L1) adds a penalty equal to the absolute value of the coefficients, making the coefficient of the less important independent variables exactly zero [28–30]. *Ridge Regression (L2)* adds a penalty equal to the square of the coefficients, shrinking them evenly to prevent multicollinearity. It prevents any single feature from dominating the prediction [31].

ElasticNet Regression (L1–L2) is a hybrid of Ridge and Lasso. Its value requires the careful tuning of two hyperparameters, *alpha* and *l1_ratio*, to get the optimum result. The dual L1–L2 penalty can bog down the model [29].

Huber Regression, as robust regression, addresses extreme sensitivity to outliers, a flaw of the standard Ordinary Least Squares (OLS) Regression. It introduces a piecewise loss function, controlled by a threshold parameter called *epsilon*. The *epsilon* value is the threshold for what the model considers an “outlier” [32,33].

Random Forest Regressor is an ensemble method that builds hundreds of independent decision trees on random subsets of data and averages their predictions. It is highly relevant in this research because of the presence of categorical and numerical data. Random Forest does not require strict mathematical scaling, and it captures non-linear interaction well [34,35]. *XGBRF Regressor* is the XGBoost framework’s native implementation of a Random Forest.

An *Artificial Neural Network (ANN)* is a network of interconnected nodes (neurons) spread across hidden layers, capable of learning highly complex, non-linear mappings [36,37].

From the literature review, a multilayer feed-forward neural network with a backpropagation learning technique tends to map the non-linear relationship between input and

output parameters better than statistical methods [1,38]. Table 4 provides the key properties of the models considered.

Table 4. Model category, model, and their key properties.

Model Category	Model	Key Properties
Linear Model	Linear Regression (sklearn)	Interpretable, assumes linearity
	OLS Regression (Statsmodels)	Interpretable, statistical rigour
Regularised Linear Model	Ridge Regression	Handles multicollinearity, reduces coefficients
	Lasso Regression	Feature selection, sparse coefficients
	ElasticNet Regression	Combines Lasso/Ridge, handles multicollinearity and sparsity
Robust Linear Model	Huber Regressor	Robust to outliers, less sensitive to extreme values
Ensemble (Tree-based)	RandomForestRegressor (Optimised)	Non-linear, handles interactions
	XGBRF Regressor	Gradient boosting + Random Forest
Deep Learning (Neural Network, NN)	Initial Neural Network	Basic feed-forward NN, non-linear capabilities
	Optimised Neural Network (Shallow)	Optimised shallow NN, improved generalisation
	Neural Network (Deeper Optimised)	Deeper architecture, regularisation for robustness, and the highest accuracy

2.10.2. Models Selected

The models selected in this research include the OLS Regression implemented using the OLS Regression in the Statsmodels library. It was selected to serve as the baseline model. In the Regularised Linear Model category, the ElasticNet Regression model was selected, as it is a hybrid of Ridge and Lasso. Huber Regressor was also selected for its robustness to outliers observed in the dataset. In the ensemble (tree-based) category, an optimised RandomForestRegressor was selected for its ability to handle the mix of categorical and numerical data present in the dataset. The neural network was evaluated.

2.11. Hyperparameter Tuning—Cross-Validation

Because we have a limited amount of data entry (142 rows), cross-validation (CV), a resampling technique, was used to ensure the model trains on well-distributed data and generalises when tested on an unseen dataset. To reduce trial and error in the determination of the best models and ANN architecture that will achieve the optimal performance, the *GridSearchCV()* function was used to search for indications of the potential optimal number of layers and hyperparameter values. The hyperparameters applied to the models are as detailed in Table 5.

Table 5. Model design features.

Model	Key Hyperparameters	Architecture/Layers	Key Properties	Optimiser/Solver	Loss Function	Activation Functions	Epochs	Batch Size	Cross-Validation	Momentum
OLS Regression (Statsmodels)	None	Single layer (linear combination)	Interpretable, statistical rigour	N/A (Analytical/OLS)	Mean Squared Error	N/A	N/A	N/A	N/A (Direct fit)	N/A
ElasticNet Regression	alpha = 0.1, l1_ratio = 0.9, max_iter = 10,000	Single layer with L1+L2 regularisation	Combines Lasso/Ridge, handles multicollinearity and sparsity	Coordinate Descent	Mean Squared Error + L1/L2 Penalty	N/A	N/A	N/A	GridSearchCV (cv = 3)	N/A
Huber Regressor	alpha = 0.001, epsilon = 2.0, max_iter = 10,000	Single layer (robust loss function)	Robust to outliers, less sensitive to extreme values	lbfgs/sgd (Default)	Huber Loss	N/A	N/A	N/A	GridSearchCV (cv = 3)	N/A
Random Forest Regressor (Optimised)	n_estimators = 100, max_features = 0.6, max_depth = None, min_samples_split = 2	Ensemble of decision trees	Non-linear, handles interactions	N/A	Mean Squared Error (Implicit)	N/A	N/A	N/A	GridSearchCV (cv = 3)	N/A
Initial Neural Network	activation = relu, epochs = 100, batch_size = 32	Input layer + 2 hidden layers (64, 32 neurons) + output layer	Basic feed-forward NN, non-linear capabilities	Adam	Mean Squared Error	ReLU	100	32	N/A (direct fit, val_data used for monitoring)	N/A (Adam is adaptive)
Optimised Neural Network (Shallow)	activation = relu, epochs = 100, batch_size = 16, learning_rate = 0.001, neurons_layer1 = 128, neurons_layer2 = 64	Input layer + 2 hidden layers (128, 64 neurons) + output layer	Optimised shallow NN, improved generalisation	Adam	Mean Squared Error	ReLU	100	16	GridSearchCV (cv = 3)	N/A (Adam is adaptive)
Neural Network (Deeper Optimised)	activation = relu, epochs = 100, batch_size = 16, learning_rate = 0.001, neurons_layer1 = 128, neurons_layer2 = 64, neurons_layer3 = 32, dropout_rate = 0.0, l1_reg = 0.0001, l2_reg = 0.0001	Input layer + 3 hidden layers (128, 64, 32 neurons) with dropout and L1/L2 regularisation + output layer	Deeper architecture, regularisation for robustness, and the highest accuracy	Adam	Mean Squared Error	ReLU	100	16	GridSearchCV (cv = 3)	N/A (Adam is adaptive)

2.12. Regularisation and Dropout

Regularisation helps to avoid overfitting. L2 and L1 regularisation and Dropout were used in this research. In the ANN models developed, L1–L2 regularisation was applied to prevent model overfitting. Dropout randomly selects and removes neurons in the neural network architecture's input and hidden layer(s). The regularisation parameters were set using the *RandomSearch()* function in the Keras library.

3. Results

The models are examined in two steps: the linear, regularised linear, robust linear, and the ensemble (tree-based) models were first examined to establish a benchmark, and the neural network models were then examined.

3.1. Linear, Regularised Linear, Robust Linear, and the Ensemble (Tree-Based) Models: Result Analysis

Table 6 shows the summarised ranked performance of the linear, regularised linear, robust linear, and the ensemble (tree-based) models. The Huber Regressor (result boldened) stands out as the best-performing model, with an R-squared score of 0.983 and the lowest MAE value of 11.23 and RMSE of 14.24. This shows that the Huber regressor is robust to outliers.

Table 6. Model performance ranking of linear, regularised linear, robust linear, and the ensemble (tree-based) models.

Model	MAE	RMSE	R-Squared
Huber Regressor	11.235922	14.250461	0.983133
OLS Regression (Statsmodels)	13.458345	17.774829	0.973759
ElasticNet Regression	16.820361	20.749254	0.964242
RandomForestRegressor (Optimised)	16.195724	41.542930	0.856660

3.2. Neural Network Result Analysis

Figure 10 shows the Training vs. Validation Loss plots for the three neural network models, indicating their learning process and performance. It informs us as to whether a model is overfitting, underfitting, or generalising well.

The initial neural network's Training Loss steadily decreases over epochs, indicating that the model is learning to fit the training data. The model learned the training data very well but struggled to generalise to unseen data (the validation set). This model performed worse than most of the previous models. Therefore, there is a need for hyperparameter tuning. Table 7 shows the summarised ranked performance of the neural network models.

The Optimised Neural Network (shallow) shows a better alignment between the Training Loss and Validation Loss. Further hyperparameter tuning—optimal learning rate, batch size, and number of neurons—led to a more effective learning process. The model found a good balance between fitting the training data and performing well on unseen data, indicating reduced overfitting compared to the initial model. The presence of regularisation (L1/L2, Dropout) in the Neural Network (Deeper Optimised) (result boldened) helps prevent the model from becoming overly complex and overfitting. The deeper architecture and regularisation exhibited strong generalisation capabilities. The consistent decrease and convergence of both losses indicate that the model has learnt complex patterns from the data efficiently and has prevented overfitting.

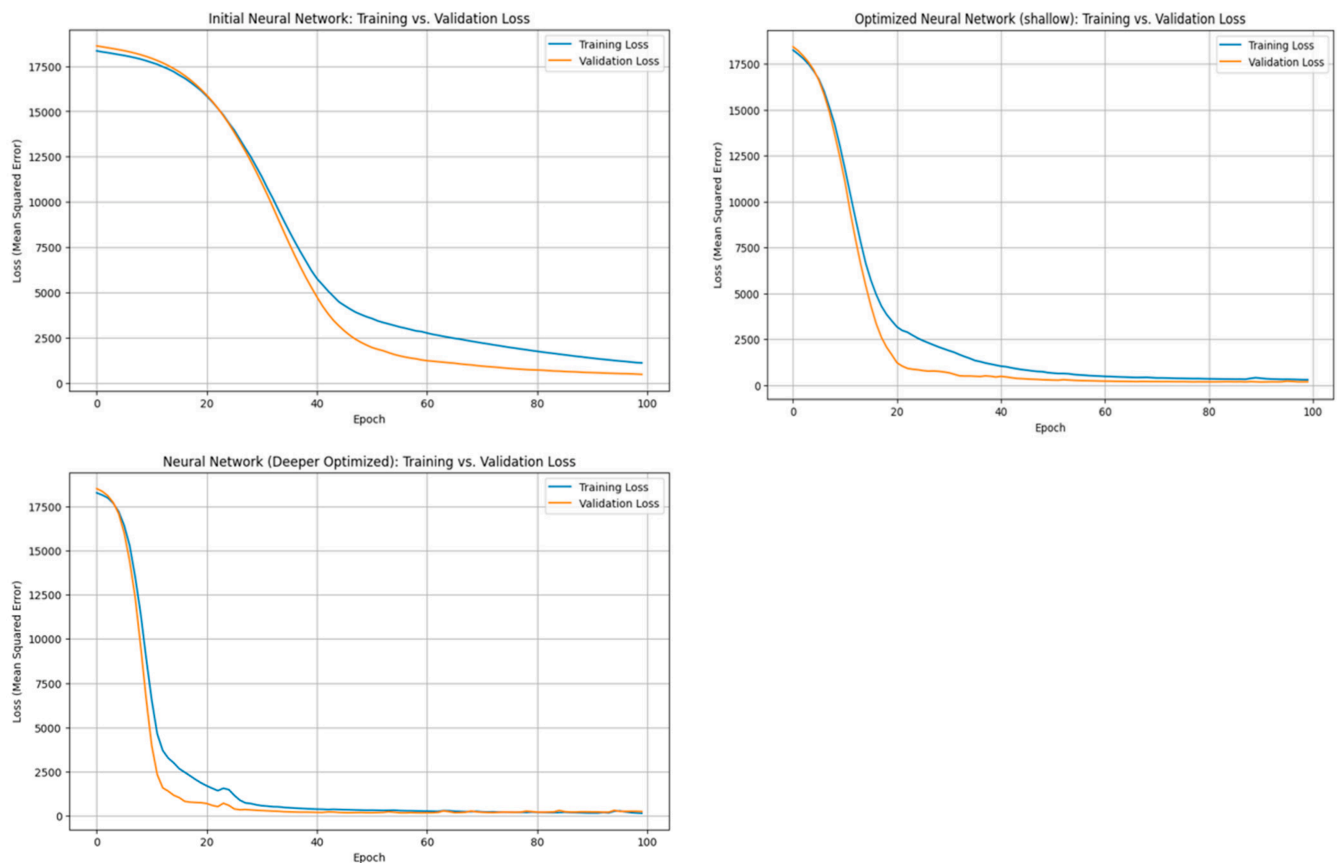


Figure 10. Actual vs. predicted man-hours (IMHR) values for the neural network models.

Table 7. Performance ranking of neural networks.

Model	MAE	RMSE	R-Squared
Neural Network (Deeper Optimised)	5.253364	13.558393	0.984732
Optimised Neural Network (Shallow)	6.696214	13.570539	0.984704
Initial Neural Network	16.726062	23.401505	0.954516

3.3. All Models' Results Analysis

Figure 11 shows the actual vs. predicted man-hours (IMHR) values for all the key models—Huber, XGBRF Regressors, and the neural networks. The Plot, with points closely aligned with the ideal prediction line, indicated higher accuracy for prediction. The Neural Network (Deeper Optimised) is the most effective model for predicting “IMHR”. Its ability to handle potential outliers while maintaining a high predictive accuracy makes it the best model for this study.

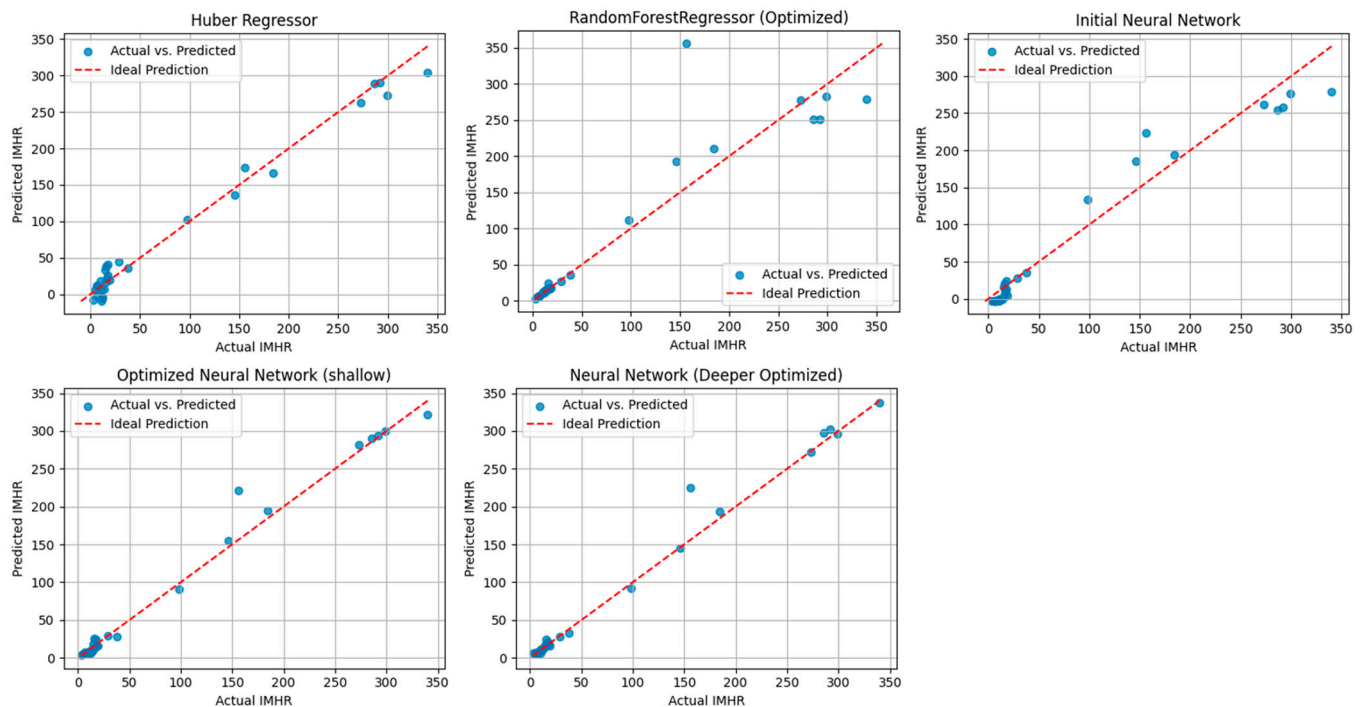


Figure 11. Actual vs. predicted man-hours (IMHR) values for all models.

3.4. Final Model Performance Ranking

Table 8 shows the performance ranking of all the models evaluated. After a comprehensive evaluation of various machine learning models for predicting “IMHR”, including linear models, tree-based ensembles, and neural networks. The Neural Network (Deeper Optimised) model (result boldened) achieved the highest prediction metric compared to the other models. This indicates its superior predictive accuracy and robust performance on the test set. The Huber Regressor also shows robustness to outliers.

Table 8. Performance ranking of all models.

Model	MAE	RMSE	R-Squared
Neural Network (Deeper Optimised)	5.253364	13.558393	0.984732
Optimised Neural Network (Shallow)	6.696214	13.570539	0.984704
Huber Regressor	11.235922	14.250461	0.983133
OLS Regression (Statsmodels)	13.458345	17.774829	0.973759
ElasticNet Regression	16.820361	20.749254	0.964242
Initial Neural Network	16.726062	23.401505	0.954516
RandomForestRegressor (Optimised)	16.195724	41.542930	0.856660

The consistently strong performance of the Huber Regressor in the statistical models and the neural networks confirms its effectiveness in handling potential outliers or complex relationships within the data. However, the neural network models are found to be more suitable for the modelling task.

3.5. Model Scalability

Unlike traditional statistical models, ANNs can handle and interpret datasets with complex non-linear relationships. To ensure that the model generalises well to unseen data, a three-fold cross-validation was applied during training. Its performance on the

test data confirms that the model is robust and highly adaptable. For practitioners, the proposed Artificial Neural Network (ANN) deployed in this research offers a framework for high-volume datasets.

4. Conclusions

The ability to predict and estimate man-hours during an offshore structure project bid exercise or project planning is one of the important components of a successful offshore campaign, as this is tied to the ability to estimate costs, book the amount of time the offshore construction vessels need, and manage very tight-scheduled and high-cost offshore construction work requirements. This is especially important considering the level of competition for projects in the offshore construction sector. An accurate man-hours estimate will increase the profitability and competitiveness of construction contractors.

This research attempts to predict man-hours of a resource-intensive activity in an offshore construction project using data extracted from archived construction management documents and a database for an offshore construction project. The study set out to develop an ANN model that will help estimators in the HuC sector to predict man-hours for a resource-significant task, which in this case is piping installation.

Linear, regularised linear, robust linear, and the ensemble (tree-based) models and supervised neural network models with varied architecture and hyperparameter values were evaluated and compared. The best performance was obtained using the deep-optimised ANN model. The results obtained are consistent with those of previous studies. The ANN model developed has the potential to eliminate the manual man-hour estimation. At worst, it can serve as a benchmark when performing critical estimates.

This research highlights the importance of turning the historical information in the repositories of construction companies into insight. This can be leveraged when estimating and planning future projects. For construction companies to maximise the benefits of historical data, as highlighted in this research, they must come up with a data vision and strategy.

One limitation of this study is that the findings are limited to the collected data range and parameters considered in the study. It should be noted that the developed model does not involve any parameters that directly account for management strategies and skills or any project-specific conditions.

Future Research Recommendation

Machine Learning is a growing field of study, and applications are being sought. ML has a wide scope and many approaches; it, therefore, cannot fit into the scope of a single research work. For future research, we recommend the use of more extensive data from different projects and companies for the development of a model that can predict man-hours for activities of different disciplines. The dataset used was manually extracted; we recommend using an automated ML approach to extract this information in future research.

Author Contributions: Conceptualisation, M.T.O.; methodology, M.T.O. and M.A.; validation, M.T.O. and M.A.; writing—original draft preparation, M.T.O.; writing—review and editing, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable.

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

References

1. Golnaraghi, S.; Zangenehmadar, Z.; Moselhi, O.; Alkass, S. Application of Artificial Neural Network(s) in Predicting Formwork Labour Productivity. *Adv. Civ. Eng.* **2019**, *2019*, 5972620. [\[CrossRef\]](#)
2. Choi, K.; Haque, M.; Lee, H.W.; Cho, Y.K.; Kwak, Y.H. Macroeconomic Labour Productivity and Its Impact on Firm's Profitability. *J. Oper. Res. Soc.* **2013**, *64*, 1258–1268. [\[CrossRef\]](#)
3. McKinsey Global Institute. *Reinventing Construction: A Route of Higher Productivity*; McKinsey Global Institute: New York, NY, USA, 2017.
4. ILO. *Good Practices and Challenges in Promoting Decent Work in Construction and Infrastructure Projects*; International Labour Organization: Geneva, Switzerland, 2015.
5. Randolph, M.F.; Gaudin, C.; Gourvenec, S.M.; White, D.J.; Boylan, N.; Cassidy, M.J. Recent Advances in Offshore Geotechnics for Deep Water Oil and Gas Developments. *Ocean Eng.* **2011**, *38*, 818–834. [\[CrossRef\]](#)
6. IEA. *World Energy Employment Report*; International Energy Agency: Paris, France, 2022.
7. The Observatory of Economic Complexity (OEC). Crude Petroleum. Available online: <https://oec.world/en/profile/hs/crude-petroleum-oils> (accessed on 7 April 2026).
8. International Labour Office (ILO). *Opportunities and Challenges for a Just Transition to a Future of Work That Contributes to Sustainable Development*; International Labour Office: Geneva, Switzerland, 2022.
9. Gerwick, B.C., Jr. *Construction of Marine and Offshore Structures*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2007; ISBN 978-0-429-12502-7.
10. Flood, I. Next Generation Artificial Neural Networks for Civil Engineering. *J. Comput. Civ. Eng.* **2006**, *20*, 305–307. [\[CrossRef\]](#)
11. Zakeri, Z.; Mansfield, N.; Sunderland, C.; Omurtag, A. Cross-Validating Models of Continuous Data from Simulation and Experiment by Using Linear Regression and Artificial Neural Networks. *Inform. Med. Unlocked* **2020**, *21*, 100457. [\[CrossRef\]](#)
12. Claridge, J. *The Planning and Control of Offshore Hook-Up and Commissioning*; Society of Petroleum Engineers: Richardson, TX, USA, 1982; p. SPE-12802-MS.
13. Van Wyngaard, C.J.; Pretorius, J.H.C.; Pretorius, L. Theory of the Triple Constraint—A Conceptual Review. In *Proceedings of the 2012 IEEE International Conference on Industrial Engineering and Engineering Management*; IEEE: Hong Kong, China, 2012; pp. 1991–1997.
14. Hanna, A.; Boodai, F.; El Asmar, M. State of Practice of Building Information Modeling in Mechanical and Electrical Construction Industries. *J. Constr. Eng. Manag.* **2013**, *139*, 04013009. [\[CrossRef\]](#)
15. Jarkas, A.M.; Bitar, C.G. Factors Affecting Construction Labor Productivity in Kuwait. *J. Constr. Eng. Manag.* **2012**, *138*, 811–820. [\[CrossRef\]](#)
16. Xu, Y.; Zhou, Y.; Sekula, P.; Ding, L. Machine Learning in Construction: From Shallow to Deep Learning. *Dev. Built Environ.* **2021**, *6*, 100045. [\[CrossRef\]](#)
17. Sircar, A.; Yadav, K.; Rayavarapu, K.; Bist, N.; Oza, H. Application of Machine Learning and Artificial Intelligence in Oil and Gas Industry. *Pet. Res.* **2021**, *6*, 379–391. [\[CrossRef\]](#)
18. Moselhi, O.; Hegazy, T.; Fazio, P. Neural Networks as Tools in Construction. *J. Constr. Eng. Manag.* **1991**, *117*, 606–625. [\[CrossRef\]](#)
19. Mistry, J.M.; Jayaraj, G.K. Analyzing Labor Productivity for Reinforcement Installation Using Artificial Neural Network in India. In *Recent Trends in Civil Engineering*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 1093–1106. [\[CrossRef\]](#)
20. Ezeldin, A.S.; Sharara, L.M. Neural Networks for Estimating the Productivity of Concreting Activities. *J. Constr. Eng. Manag.* **2006**, *132*, 650–656. [\[CrossRef\]](#)
21. Mohamed, B.; Mouloud, B.; Mourad, M. Project Management Using Cost Significant Items. In *Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management, Istanbul, Turkey, 3–6 July 2012*; Industrial Engineering and Operations Management Society (IEOM): Southfield, MI, USA, 2012; p. 8.
22. Yazdani-Chamzini, A.; Zavadskas, E.; Antucheviciene, J.; Bausys, R. A Model for Shovel Capital Cost Estimation, Using a Hybrid Model of Multivariate Regression and Neural Networks. *Symmetry* **2017**, *9*, 298. [\[CrossRef\]](#)
23. Matel, E.; Vahdatikhaki, F.; Hosseinyalamdary, S.; Evers, T.; Voordijk, H. An Artificial Neural Network Approach for Cost Estimation of Engineering Services. *Int. J. Constr. Manag.* **2019**, *22*, 1274–1287. [\[CrossRef\]](#)
24. Siddappa, K. A Predictive Model for Scaffolding Manhours in Heavy Industrial Construction Projects: An Application of Machine Learning. Master's Thesis, Concordia University, Montreal, QC, Canada, 2019.
25. Dmaidi, N.; Al-Khatib, I. Modeling the Oil Industry Norms. *Bethlehem Univ. J.* **2000**, *19*, 49–73.
26. Barzegar, R.; Asghari Moghaddam, A. Combining the Advantages of Neural Networks Using the Concept of Committee Machine in the Groundwater Salinity Prediction. *Model. Earth Syst. Environ.* **2016**, *2*, 26. [\[CrossRef\]](#)
27. Mirahadi, F.; Zayed, T. Simulation-Based Construction Productivity Forecast Using Neural-Network-Driven Fuzzy Reasoning. *Autom. Constr.* **2016**, *65*, 102–115. [\[CrossRef\]](#)
28. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [\[CrossRef\]](#)

29. Zou, H.; Hastie, T. Regression Shrinkage and Selection via the Elastic Net, with Applications to Microarrays. *J. R. Stat. Soc. Ser. B* **2004**, *67*, 301–320.
30. Melkumova, L.E.; Shatskikh, S.Y. Comparing Ridge and LASSO Estimators for Data Analysis. *Procedia Eng.* **2017**, *201*, 746–755. [[CrossRef](#)]
31. Tikhonov, A.N. On the Stability of Inverse Problems. *Dokl. Akad. Nauk SSSR (NS)* **1943**, *39*, 195–198.
32. Huber, P.J. Robust Estimation of a Location Parameter. In *Breakthroughs in Statistics: Methodology and Distribution*; Kotz, S., Johnson, N.L., Eds.; Springer: New York, NY, USA, 1992; pp. 492–518, ISBN 978-1-4612-4380-9.
33. Feng, Y.; Wu, Q. A Statistical Learning Assessment of Huber Regression. *J. Approx. Theory* **2022**, *273*, 105660. [[CrossRef](#)]
34. Breiman, L. Statistical Modeling: The Two Cultures. *Stat. Sci.* **2001**, *16*, 199–215. [[CrossRef](#)]
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
36. Csáji, B.C. Approximation with Artificial Neural Networks. Master's Thesis, Faculty of Sciences, Eötvös Loránd University, Budapest, Hungary, 2001; p. 46.
37. Nielson, M. *Neural Network and Deep Learning*; Determination Press: San Francisco, CA, USA, 2015.
38. Ok, S.C.; Sinha, S.K. Construction Equipment Productivity Estimation Using Artificial Neural Network Model. *Constr. Manag. Econ.* **2006**, *24*, 1029–1044. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.