



# Machine Learning based Material Demand Prediction of Construction Equipment for Maintenance

Poonam Katyare<sup>1</sup>, Shubhalaxmi S. Joshi<sup>2</sup>

<sup>1</sup> Department of Computer Science and Application, Dr. Vishwanath Karad MIT World Peace University, Pune, India

<sup>2</sup> Department of Computer Science and Application, Dr. Vishwanath Karad MIT World Peace University, Pune, India  
 poonamkatyare@gmail.com, shubhalaxmi.joshi@mitwpu.edu.in

...

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

**Abstract:** Construction managers faced the Construction Equipment (CE) challenges related to running repair and replacement of spare part materials as well as shortage of materials, sudden damage of spare parts and unavailability of necessary materials at job sites frequently. Regular follow up and track of materials availability and their usage at each stage of requirement phase becomes essential. This study presents Machine Learning (ML) based material demand prediction. Training of ML models utilizes historical maintenance, and procurement periodic data related to materials of the CE. This study highlights the use of Multiple Linear Regression (MLR), Support Vector Regression (SVR), Decision Tree (DT) Regressor and ensemble boosting models as Random Forest (RF) Regressor and Gradient Boosting Regressor (GBR). According to the performance measurement of each model, RF performs better and is used for prediction. Material demand prediction helps in maintenance and operational planning of CE. Subsequently, approach assists in addressing issues early by involving operators and site owners, enabling preventive actions to be taken before the scheduled procurement process. This study addresses the corrective measurement of the model using periodic data. The model performance results indicate that early prediction of maintenance costs based on the quantity of essential materials withdrawn from demand is helpful for budgeting expenditures.

**Keywords:** Construction Equipment, Machine Learning, Material Demand, Maintenance

## 1. INTRODUCTION

Construction Equipment is a key driver for executing successful construction projects. The management of CE concerns efficiently overseeing equipment resources to meet the equipment requirements and to gain maximum returns on equipment for the construction project, which is targeted to be executed in a scheduled and economically viable fashion. Major contractors often have the capability to acquire, operate, and manage a substantial collection of heavy CE units. Therefore, making decisions regarding routine equipment management responsibilities is essential for overall project management. The daily routine involves the procurement process, maintenance process, equipment allocations, equipment operational activity, and replacement and repair activities of equipment spare parts. The day-to-day execution of these activities has financial implications for fleet owners because cost is involved in every activity. Proper and effective budgeting

of any construction project focuses on CE cost bifurcations on various aspects, which involve the initial acquisition cost of the equipment, operating cost, maintenance and repairs cost, operator and labor wages, depreciation cost, financing costs, interest payments, transportation cost, regulatory compliance costs, technology integration costs, and disposal cost of CE. A critical piece of maintenance costs is fundamentally credited to essential materials in form of spare parts for CE [6]. Site owners should maintain an inventory of spare parts associated with all equipment present and currently working on the job site. It is a big challenge to handle equipment failures and face downtime while performing tasks on site using equipment. They need to keep records of all materials in the system with their availability quantity, order details, required quantity, withdrawn quantity, and special operating run hours of the equipment. The cost sheet for each quantity is recorded with date and time. Large numbers of lists of



materials are available that are distributed in similar groups of equipment for simplicity of cost computations.

This investigation aims to avoid manual work for computing demand quantity of materials. The proposed study emphasizes ML-based essential materials demand prediction of CE in advance from a maintenance perspective. This study focuses on the rational study of various ML algorithms.

This paper is arranged as follows. Existing study with limitation is elaborated in part 2. The proposed methodology for predicting the demand quantity of materials with data preprocessing and model fitting is given in part 3. Results and discussion with comparison of ML algorithm performance is indicated in part 4. The inference with concluded work is illustrated in part 5.

## 2. EXISTING STUDY

A significant study was identified related to CE cost prediction, residual value prediction, and sensor-based data analysis. Maintenance of CE study observed using different methods, such as by reviewing the existing techniques used for reliability and fault analysis of CE. ML techniques, graphical methods, fault tree analysis, and probability distribution models have been used; however, ML models have the best accuracy for failure prediction and reliability estimations of CE [3]. The researcher presented a related study on the implications of the Internet of Things with sensor-based technologies attached to the equipment for capturing real-time information of CE with location tracking, movement tracking, working condition of engines, fuel data updates, distance travelled, and battery updates from equipment working on construction job sites. This would help managers analyse the data collected from remote sensors and make proper decisions regarding the equipment's performance. Remote sensing devices identify information related to construction material tracking to handle the supply chain management process along with cloud computing, radio frequency identification, and augmented reality and big data technologies [1][5][7].

Existing study presented the prediction of residual values of CE by an Autoregressive Tree algorithm of data mining using equipment age, make, model, region, horsepower, auction year, condition rating, annual construction investment, Gross Domestic Product features to predict equipment price. This study compared the performance of data mining algorithms with those of neural networks, linear regression algorithms, and deep learning algorithms [2][11]. The authors demonstrated fuel consumption prediction using ML and live data parameters from smart sensing devices, which indirectly impacted to the maintenance cost of CE. Real-time data ensemble methods provide better accuracy than other regression models [4].

Another study demonstrated a prototype model that effectively reduces labour costs and mitigates challenges associated with equipment maintenance decision-making by presenting a data-driven methodology that integrates three key skills as reliability maintenance focused on reliability, modelling of building data and live tracking system. This includes critical components for ensuring the optimal functionality of buildings [8][9]. Data-driven approaches have been observed in many studies that attempt to manage equipment information data using data analysis from huge amounts of data and manage the data for decision making [10].

Manufacturers in the construction machinery parts industry must manage inventories in a timely manner, optimize production processes for efficient and swift product manufacturing, and promptly deliver finished products to customers. To solve this problem, an existing study elaborated demand estimating for spare parts in the construction machinery industry using regression and artificial intelligence models [12]. Similarly, the heavy equipment of specific group demand forecasting is also performed by the researcher using the Support Vector Machine Regressor, which is very useful for the equipment owner [13]. Multivariate time series analysis performs better for the construction raw materials of steel products prediction [14]. A related study investigated the prediction of heavy equipment prices with precision by employing ML algorithms on sales data obtained from a website. [15].

Another study uses an artificial neural network-based methodology to measure uncertainties and generate forecasting intervals for predicting prices of construction material, with a specific focus on asphalt and steel. This study provides supplementary information to enhance the effective management of project cost-related risks through estimate intervals to project managers. The proposed optimal Lower Upper Bound Estimation (LUBE) cost function yields highly precise estimate intervals [16]. The Analytical Hierarchy Process within the thematic domain involves the development of a modified decision model in CE procurement. This model designed to order parameters that influence equipment procurement. The approach is particularly tailored to address the unstructured aspects of the selection method [17].

Research showcased prediction of maintenance costs related to breakdown and planned maintenance activity events for essential plant resources, and the developed model exhibited strong predictive accuracy. The methodology integrates a stochastic mathematical modelling approach that considers both unplanned breakdowns and scheduled maintenance. This technique generates a pseudo-random number to simulate the magnitude of an impending maintenance cost event [18]. Time series maintenance and fuel consumption data were

used to anticipate the CE cost of maintenance using a neural network time series model [19].

replacement of material, and damage of material at the site. They should add all these records manually to log

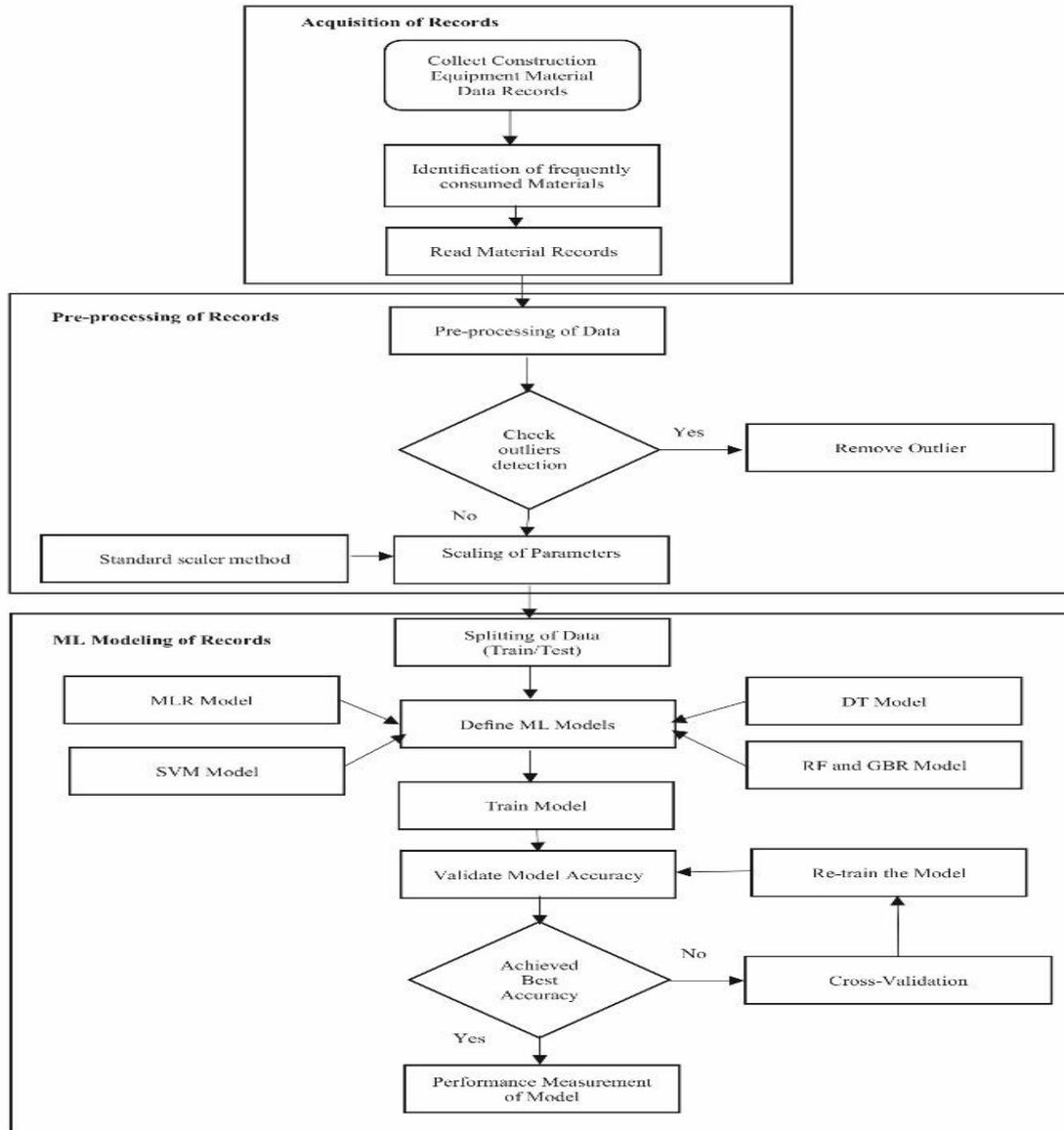


Figure. 1. Prediction Model Flow Diagram

**A. Limitations of Existing Study**

Existing studies have demonstrated the prediction of residual values of equipment, prediction related to the cost of equipment, prediction of failure or breakdown of equipment, fuel consumption, and maintenance cost estimation of equipment. This study presents qualitative and quantitative data analysis using ML models, time series analysis, and factor analysis. Contractors or site owners need to maintain records of equipment spare parts or materials available, required or demand quantity, and how much quantity is utilized manually. They face the issues of failure of materials,

sheets, order the materials as per the requirement, and wait for procurement of those materials. Meanwhile, there could be downtime at the site for that equipment because of the unavailability of the materials required for that equipment aligned with the working conditions. The frequency of such situations or challenges frequently occurred at the job site. There is a need to estimate such a material demand quantity. The proposed study predicts the quantity of equipment required for material demand from the operating hours and history data, which would



help to maintain the required materials stock in advance at the job site.

### 3. PROPOSED METHODOLOGY

This emphasizes analysis of data related to CE materials, detailing the quantity available for each machine, used run hours, the specifics of each machinery order, and the anticipated future demand for each. The goal is to predict the demand for construction machinery. Figure 1. illustrates the Prediction model flow diagram with the detailed steps for estimating the quantity of materials. This flow diagram interprets the Data Acquisition, Data Preprocessing and Data Modelling steps.

#### A. Acquisition of Records

Daily logs of the repair and replacement of materials are maintained at the construction site. New orders are placed to purchase materials for replacement. Industries keep these records in their Enterprise Resource Planning system. The proposed study acquires order data and material quantity data from the organizational system from 2017 to 2023 from various sources. Interviews were conducted with experts and contractors working on job sites, and literature review data were used to finalize the features required for the proposed study. The acquired dataset is in a daily basis format with consideration of the days when orders are placed. The collected data has features related to order, material, and material quantity details. Order details include order number, order creation date, order completion date, and requirement date.

Material details features represent Site number, Material number, Equipment number, Material group, Operating run hours, and equipment manufacturer. The

major material groups are classified into various categories. Material quantity-related features denote the available quantity of material in the inventory, the required quantity at the time of replacement and repair, and the withdrawn quantity of materials represent the total quantity of materials used. Frequently consumed materials were observed during the study. Major material categorial groups of materials are highlighted in the dataset. Material group codes are present in the dataset, and mapping of all materials under groups will be used in future studies. Table 1. presents the material groups used in the dataset.

Table 1: Material Groups

No.	Material Groups
1	Spares
2	Structural Steel
3	Welding Materials
4	Pipe & Pipe Fittings
5	Hardware, Painting & Chemicals
6	Electrical Items
7	Rubber Goods
8	Lubricant and Oil
9	Tools
10	Miscellaneous
11	Consumables (Anchor/Pilling/Drilling)

#### B. Preprocessing of Records

The demand quantity of material estimation is related to the quantity of material withdrawn from historical records. The available and required quantities are major contributors to predicting the withdrawn quantity demanded. Table 2. Statistical Description of parameters represents statistical values for major parameters.

The attributes relation is identified from the correlation matrix denoted in Table 3. Correlation matrix

TABLE 2. Statistical description of parameters

	Withdrawn Qty	Material Group	Run Hrs	Quantity Available	Requirement Qty
Mean	1.88	209	6552.90	2.33	2.13
Std	2.14	45.55	4246.44	3.91	3.36
Min	0	200	2	0.004	0
25%	1	200	3350	1	1
50%	1	200	5639	1	1
75%	2	200	8983	2	2
Max	39	500	21022	286	91

TABLE 3. Correlation matrix of parameters

	P1	P2	P3	P4	P5	P6	P7	P8
P1	1	0.03	0.32	-0.06	0.32	0	0.31	0.96
P2	0.03	1	0.11	0.19	0.11	0.3	0.05	0.03
P3	0.32	0.11	1	-0.06	1	-0.04	0.39	0.32
P4	-0.06	0.19	-0.06	1	-0.06	-0.21	-0.04	-0.05
P5	0.32	0.11	1	-0.06	1	-0.04	0.39	0.32
P6	0	0.3	-0.04	-0.21	-0.04	1	-0.01	0
P7	0.31	0.05	0.39	-0.04	0.39	-0.01	1	0.32
P8	0.96	0.03	0.32	-0.05	0.32	0	0.32	1



of parameters where P1 is withdrawn quantity, P2 is site, P3 is material number, P4 is equipment number, P5 is material group, P6 is run hours, P7 is quantity available and P8 is required quantity of material.

Correlation Matrix tests can be used to check whether the information focuses are independent and indistinguishably distributed. It observes the relationship of the independent parameters with target parameter [4][13]. In this study, the material was highly correlated with the quantity available, quantity required, and operating run hours. The withdrawn quantity is also related to the available and required material quantity. Finally, the major features selected for the predictive modelling of records are material number, Equipment number, Material group, Operating run hours, quantity available, and required quantity to predict the withdrawn quantity demand of material. Outliers are identified and removed from the data using the quantile method of outlier removal [18].

1) *Outlier Removal Method*

A statistical approach, Interquartile Range (IQR) is used to remove the outliers [27]. This approach identifies the distribution of the mid fifty percentage of the records. Equation (1) represents the formula for computing IQR as the subtraction of the 75th record percentile as QT3 with the 25th record percentile as QT1.

$$IQR = QT3 - QT1 \tag{1}$$

Where,

QT1 = Upper bound with value less than 25% of records lie.

QT3 = Upper bound with value less than 75% of records lie.

This approach handles the skewed record distribution with outliers and provides a list of outliers.

2) *Feature Scaling*

It is an approach of transforming values of features from records into similar scales, which helps to define the

equal contribution of all features. Scaled features have a greater impact on performing ML models accurately. Figure 2. Yearwise Materials Utilization

Standardization is an approach that denotes that the values of features are central to the mean with a unit of standard deviation [25]. This supports the retention of the relationship between record points from the data mentioned in equation (2). It is computed as

$$(DT - \text{mean (all DTs)}) / SD \tag{2}$$

Where,

DT = Data Point

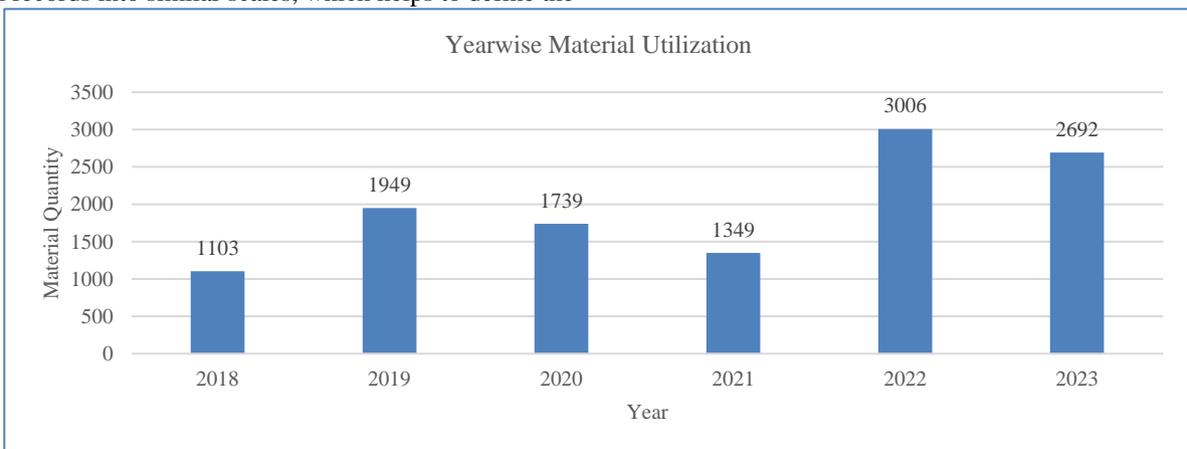
DTs = All Data Points

SD = Standard deviation of all DTs

C. *Preliminary Analysis of Records*

Preliminary analysis of the pre-processed data helps to observe year-wise material usage. This dataset is real time data of maintenance which involves materials details that were repaired and replaced at the time maintenance. The issues are related to running repairs, breakdown order, breakdown repair, calibration changes, maintenance after specific run hours, order the machinery, defective material indication, regular servicing, and handling damage materials. Every record of the issue along with order details of materials were kept as logs in the ERP system of the organization.

This dataset contains Site number, Equipment number, Material number, Material Group, Run hours, Available quantity of materials in stock, Requirement quantity and withdrawal quantity is the quantity used as the major features along with order number, order creation date, and order completion date as the minor features. Figure 2. Year-wise material usage from 2018 to 2023. As per the increment in project scheduling, the increase in the order of material usage is listed. The years 2022 and 2023 highlighted more use of materials than prior years.





#### D. Modelling of Records

ML is a strategy for changing information into noteworthy information. Different directed ML methods are accessible for expectation, which is related to the verifiable information for anticipating new occasions of data of interest with the connection of target factors alongside independent information values. ML model follows information assortment, Information preprocessing, and modelling with different algorithms. Finally, the model with the better measurement is chosen for predicting new occasions of information. We used different ML regression models, for example, MLR, SVR, DT along with ensemble regressors as RF and GBR models for determining the demand quantity of materials.

##### 1) Multiple Linear Regressor (MLR)

MLR is a basic and generally involved method for displaying the association of a dependent variable with at least one independent variable. The model anticipates a linear relationship within the dependent and independent factors, suggesting that they can be represented as a straight line. MLR is a measurable investigation technique used to determine the dependent quantitative connection between at least two factors. Target factors in the regression examination are perceived or assessed [25]. Independent factors are the factors that are remembered to significantly affect the target variable attempted for assessment. Forecasts can be made by estimating the connections within factors using examination.

Considering input parameters as  $X = (x_1, x_2, \dots, x_k)$ , the basic statistical model of MLR is stated by Equation (3).

$$Y(x,c) = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_k = c_0 + \sum_{i=1}^K c_i x_i \quad (3)$$

where  $c$  can be projected using the least squares method as

$$\hat{c} = \operatorname{argmin} \left\{ \sum_{j=1}^N (y_j - c_0 - \sum_{i=1}^K c_i x_{ji})^2 \right\} \quad (4)$$

where  $x_1, x_2, \dots, x_k$  are the observed values of independent parameters,  $c_1, c_2, \dots, c_k$  are the regression coefficients,  $c_0$  is intercept term,  $N$  is the sample count size with  $K$  represents input parameters, and  $y$  are stated value of dependent parameter.

##### 2) Support Vector Regressor (SVR)

A function provided by the SVR signifies the relationship existing within the dependent and independent parameters with a reducing error factor. The fundamental aim of SVR is to discover a hyperplane with the largest number of points within the decision boundary line or support vectors that should be present within that boundary line. Decision boundaries are used with hyperplanes to anticipate continuous values. This assortment of numerical activities known as kernels is

used to change input information into important configurations. SVR attempts to fit between the boundary lines and the hyperplane [4] [25].

The formula for a SVR can be expressed as follows:

$$\hat{y} = \sum_{i=1}^{n_{sv}} a_i K(x_i, x) + b \quad (5)$$

Where,

$\hat{y}$  - the predicted dependent value.

$n_{sv}$  - the count of support vectors.

$x_i$  - the  $i^{\text{th}}$  support vector.

$b$  - bias term

$K(x_i, x)$  - function kernel, which calculates similarity between the  $i^{\text{th}}$  support vector and the input sample  $x$  allowing for nonlinear relationships between features.  $a_i$  are coefficients associated with the support vectors. A hyperplane is calculated to fit the training data while minimizing margins. This aims to find coefficients as  $a_i$  and bias term as  $b$  that minimize the empirical risk as the variation in the anticipated and real values subject to a margin of tolerance  $\epsilon$ . This optimization problem is typically described as a quadratic programming problem and is solved using optimization techniques. Common kernel functions include sigmoid, linear, polynomial kernels. The choice of the kernel function varies depending on the complexity of the relationships between features and the nature of the data. SVR is intensely efficient for datasets with dense relationships and high-dimensional feature spaces. SVR ensures robust predictions and reduced sensitivity to outliers by expanding the margin among the hyperplane and the data points.

##### 3) Decision Tree Regressor (DT)

It is an extensively applied supervised learning algorithm. It supports regression and classification analysis. A DT is a progressive model used in portraying decisions and their expected results, consolidating chance occasions, asset costs, and utility. This algorithmic model uses contingent control proclamations in the form of statements. It is a nonparametric supervised learning method helpful for both regression and classification analysis. The tree structure contains a root node and subtrees with branches followed by interior nodes, and leaf nodes frame a hierarchical, tree-like construction [25].

The DT regression model can be represented by the following formula:

$$\hat{y} = \sum_{i=1}^N w_i \cdot I(x \in R_i) \quad (6)$$

Where:

$\hat{y}$  = forecasted target value.

$N$  = total count of leaf nodes in the DT.

$R_i$  = region as leaf node of the feature space stated as the  $i^{\text{th}}$  leaf node.



$w_i$  is the anticipated value correlated with leaf node. Input value when lies in region, Indicator function proceeds to success. The anticipated value of leaf node is treated as final prediction. Each region with associated leaf node with anticipated value represents the average of the dependent estimates of the training falling within that region [25]. The DT regressor formula essentially represents a piecewise constant function, where feature space partitioned into non-overlapping regions and each section is linked with a constant predicted value. The last estimate for a given input sample is the sum of the predicted values of the leaf nodes into which the sample falls [26].

#### 4) *Random Forest Regressor (RF)*

Ensemble learning models impact the finding of solutions to very complex regression problems. Ensemble learning can be characterized as the method involved in creating different models, such as classifiers, and then accumulating their outcomes to acquire better prescient execution. Two notable outfit learning techniques are boosting and bagging. In supporting, progressive models add additional load to preparing cases that were erroneously predicted by past models. While making the forecast, a weighted vote is considered. Although progressive models are not reliant upon prior models in bagging, each model is freely developed by a bootstrap test of information. Forecasting is created by considering a basic larger part vote. The ensemble predictive model, RF, is built on a set of decision/regression trees. Rather than basing the forecast on a single tree, a group of trees is used to make the determination. RF adds an extra element of randomization to bagging, which sets it apart from other approaches. Like other bagging models, RF uses a bootstrap of sample data to build each decision/regression tree. The process for creating trees is different [8]. Because of this technique, the RF can withstand overfitting and excel in various problems. In addition, working with the assessment of variable significance and exception recognition are different advantages of this calculation [25]. Moreover, RF is sensibly quick to obtain and can be effortlessly parallelized. By backward eliminating predictors according to the specified variable relevance, RF can be improved.

The formula for a RF can be stated as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (7)$$

Where,

$\hat{y}$  is the anticipated target value.

N is the entire trees in the RF.

$T_i(x)$  is the prediction of the  $i^{\text{th}}$  DT sample x.

This model aggregates estimates all multiple DTs to make a final prediction. Every DT is trained using a

bootstrap sample extracted from the training data and allows splitting of features randomly. At last estimate is computed by averages of all individuals' predictions. Overfitting reduces using averaging and enhances performance. Final prediction dependents on the contribution of every tree, regardless of its individual performance. This ensemble approach makes RFs robust and capable of handling noisy data while providing reliable predictions.

#### 5) *Gradient Boosting Regressor (GBR)*

It has a place in the class of ensemble learning techniques that explicitly boost calculations. It is known for its high prescient accuracy. It functions admirably with both linear and nonlinear connections between the dependent and target factors. It can deal with complex information with high dimensionality and countless factors. It can detect complex communications among factors and precisely model non-direct connections. It can deal with missing qualities in the dataset without requiring attribution. It divides information by considering accessible elements and continues to prepare the model. It is hearty handles to anomalies in the information. This method uses a collection of weak learners and limits the effect of exceptions to the iterative process. The regressor includes significance scores, allowing comprehension of the elements that are most compelling in forecasting. This can be useful for highlighting determination and identifying hidden examples in the information. The regressor is less inclined to overfitting in contrast with other complex models like profound brain organizations. This is because it assembles trees successively, improving the blunders made by the past trees. The regressor considers tweaking hyperparameters like the quantity of trees, tree profundity, learning rate, and misfortune capability, giving adaptability in model streamlining [21][22]. It tends to be used for an extensive variety of regression undertakings, including expectation of nonstop factors. GBR is a flexible and strong model reasonable regression undertaking, particularly when high prescient precision and interpretability are required [23][24]. The formula for a GBR model stated as:

$$\hat{y} = \sum_{i=1}^M \gamma_i h_i(x) \quad (8)$$

Where,

$\hat{y}$  - predicted target value.

M - total count of trees.

$h_i(x)$  - estimate of the  $i^{\text{th}}$  base learner for the input sample x.

$\gamma_i$  - learning rate associated with the  $i^{\text{th}}$  base learner.

In GBR, the model successively develops an ensemble of weak learners, typically DTs, and joins them for strong learner. Each subsequent base learner focuses residuals as



the variation involving the actual with predicted estimates of the preceding predictions. By iteratively fitting new base learners, GBR gradually improves the model's ability to trap complicated data associations. The key idea behind gradient-boosting regression is to minimize a loss function as squared error loss. Each base learner training helps to reduce the loss with respect to residuals of previous predictions. GBR is a compelling method for building predictive models for handling complex datasets. However, it is important to adjust trees as boosting iterations and the learning rate to prevent overfitting and achieve optimal performance.

#### 6) Cross Validation (CV) Technique

K-fold CV is a strategy utilized to assess performance by dividing the first dataset into k equivalent estimated subsamples, called folds. The cycle includes iteratively preparing the model k times. This permits us to obtain k arrangements of assessment measurements, ordinarily finding the middle value to obtain a more reliable prediction. This cross-validation technique of data splitting at the training-validation split can mitigate the overfitting issues and retain consistent estimate of the model execution. A more robust estimate of the model performance is provided by this technique. It uses multiple training validation splits and averages the performance. This is specifically used to select the most suitable model performance and perform a comparative evaluation of the model measurement [20].

## 4. RESULTS AND DISCUSSION

The study depicts expectation task completed to examine a bunch of elective models for predicting the material demand quantity of the chose dataset. We assessed the suitability of MLR, SVR, GBR, DT, and RF models for predicting the material demand quantity of CE. Using the real dataset, regression models were trained and evaluated.

Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R2 score) are used to assess performance measurement of the regression models presented in equation (3), (4), (5), and (6), respectively. Mean squared variance of actual with projected values assign as MSE, mean variance within the original and estimated values denotes MAE and square root of the MSE associated with error rate along with coefficient of estimated values in relation to the original values is indicated by the R2 score. The percentages represent values between 0 and 1.

$$MSE = \frac{1}{N} \sum_{i=1}^N (PRED_i - ACT_i)^2 \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |PRED_i - ACT_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (PRED_i - ACT_i)^2} \quad (5)$$

$$R2 = 1 - \frac{\sum(PRED_i - ACT_i)}{\sum(PRED_i - ACT_i)^2} \quad (6)$$

where  $PRED_i$  and  $ACT_i$  denote the  $i^{\text{th}}$  predicted and actual material demand quantity values.

The comparative examination of the models' performances is shown in Table 4. Performance measurement of models. The MLR, SVR, GBR, DT regression and RF regression models with k-fold cross validation are measured with MAE, MSE, RMSE and R2 score values and compared to predict the withdrawn material quantity demand of the CE.

TABLE 4. Performance measurement of models

Model	MAE	MSE	RMSE	R2
MLR	0.82	3.12	1.74	0.32
SVR	0.28	1.35	1.13	0.72
GBR	0.21	0.37	0.59	0.92
DT	0.06	0.24	0.46	0.95
RF	0.08	0.22	0.42	0.96

MLR can be utilized for material quantity assessment when there is a reasonable direct connection within the input factors as equipment number, material available quantity along with run hours and the target variable as the withdrawn quantity of material. In MLR, the weighted amount of the variables' coefficients is used to predict material quantity. MLR might give a decent beginning stage, yet its capacity to discover complex connections between different highlights might be restricted when predicting the withdrawn quantity of material demand. More complex models may be expected to represent nonlinear impacts. Figure 3. Performance Measurement of models represent visualization of models with R2 score. SVR with RBF kernel meets a useful ability for anticipating material quantities by really discovering complex relationships between input parameters and the target parameter. Appropriate information preprocessing, model preparation, hyperparameter tuning, and assessment are critical stages in utilizing this methodology for precise expectations in material quantity prediction assignments.

Decision trees can deal with different categories of data. The CE materials dataset represents numerical and categorical parameters implications. DT is a very simple method of decision making at each stage of

splitting nodes. DT are inclined to overfitting, particularly when the tree develops intensely in the training data. This can prompt unfortunate speculation execution on inconspicuous information. Ensemble methods as RF and GBR helps for resolving data overfitting and provide better results for forecasting material withdrawn quantities and demand. RF enhances decision trees by joining various trees' forecasts. It acquires complex associations between parameters and material quantities. It can handle multi feature data.

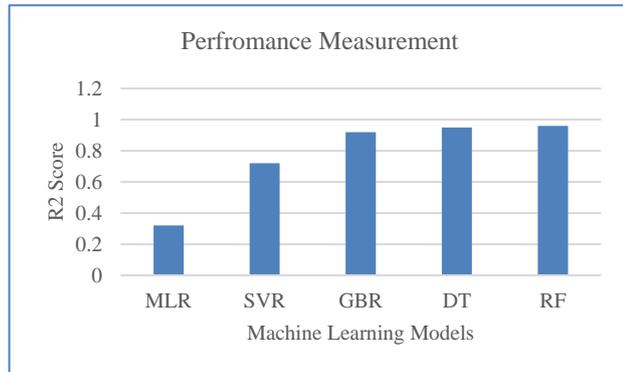


Figure 3. Performance Measurement

## 5. CONCLUSION

This study focuses on the machine-learning-based material-demand prediction of CE. Maintenance data records were analyzed in this study. The limitations of the existing study were acknowledged and summarized using ML technologies, and a model to predict the material demand quantity was proposed. This study helps to estimate the material demand in advance for maintenance and to maintain the maintenance cost associated with the estimated materials in planning. This study provides various ML based regression models, such as MLR, SVR, GBR, DT, and RF regression model performance. The results reveal the viability of utilizing ML techniques to overcome the difficulties in predicting material quantities. The RF model predicts material quantities accurately and performs better than other regression models. It is critical to handle real-time data for preprocessing, which involves outlier removal, handling missing values, and scaling the features to acquire accurate data for modelling. ML models are very sensitive to the quality of the dataset. This study demonstrates ML applications for material quantity prediction of CE in construction industry for maintenance. The estimation of maintenance and operating costs of materials for CE leads to the financial budgeting of the overall construction project at the job site.

## REFERENCES

- [1] A. Kumar and O. Shoghli, "A review of IoT applications in Supply Chain Optimization of Construction Materials," ISARC 2018 - 35th Int. Symp. Autom. Robot. Constr. Int. AEC/FM Hackathon Futur. Build. Things, no. July 2018, 2018, doi: 10.22260/isarc2018/0067.
- [2] H. Fan, S. AbouRizk, H. Kim, and O. Zaïane, "Assessing Residual Value of Heavy Construction Equipment Using Predictive Data Mining Model," *J. Comput. Civ. Eng.*, vol. 22, no. 3, pp. 181–191, 2008, doi: 10.1061/(asce)0887-3801(2008)22:3(181).
- [3] P. Odeyar, D. B. Apel, R. Hall, B. Zon, and K. Skrzypkowski, "A Review of Reliability and Fault Analysis Methods for Heavy Equipment and Their Components Used in Mining," *Energies*, vol. 15, no. 17, pp. 1–27, 2022, doi: 10.3390/en15176263.
- [4] P. Katyare, S. S. Joshi, and S. Rajapurkar, "Real time data modeling for forecasting fuel consumption of Construction Equipment using integral approach of IoT and ML techniques," *J. Inf. Optim. Sci.*, vol. 44, no. 3, pp. 427–437, 2023, doi: 10.47974/JIOS-1363.
- [5] P. Katyare and S. S. Joshi, "Construction Industry Digitization using Internet of Things Technology," vol. 3307, pp. 243–249, 2022.
- [6] H. Fan, H. Kim, and O. R. Zaïane, "Data warehousing for Construction Equipment management," *Can. J. Civ. Eng.*, vol. 33, no. 12, pp. 1480–1489, 2006, doi: 10.1139/L06-108.
- [7] P. Katyare, S. Joshi, "Construction Productivity Analysis in Construction Industry: An Indian Perspective," *Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems*. Springer, Singapore; 2022. doi: 10.1007/978-981-19-0863-7\_11
- [8] Z. Ma, Y. Ren, X. Xiang, and Z. Turk, "Data-driven decision-making for equipment maintenance," *Autom. Constr.*, vol. 112, no. January, p. 103103, 2020, doi: 10.1016/j.autcon.2020.103103.
- [9] J. C. P. Cheng, W. Chen, K. Chen, and Q. Wang, "Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms," *Autom. Constr.*, vol. 112, no. December 2019, p. 103087, 2020, doi: 10.1016/j.autcon.2020.103087.
- [10] H. Fan, H. Kim, S. AbouRizk, and S. H. Han, "Decision support in Construction Equipment management using a nonparametric outlier mining algorithm," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1974–1982, 2008, doi: 10.1016/j.eswa.2007.02.010.
- [11] O. Alshboul, A. Shehadeh, M. Al-Kasasbeh, R. E. Al Mamlook, N. Halalsheh, and M. Alkasasbeh, "Deep and machine learning approaches for forecasting the residual value of heavy Construction Equipment: a management decision support model," *Eng. Constr. Archit. Manag.*, vol. 29, no. 10, pp. 4153–4176, 2022, doi: 10.1108/ECAM-08-2020-0614.
- [12] A. Aktepe, E. Yanık, and S. Ersöz, "Demand forecasting application with regression and artificial intelligence methods in a construction machinery company," *J. Intell. Manuf.*, vol. 32, no. 6, pp. 1587–1604, 2021, doi: 10.1007/s10845-021-01737-8.
- [13] A. Kargul, A. Glaese, S. Kessler, and W. A. Günthner, "Heavy Equipment Demand Prediction with Support Vector Machine Regression Towards a Strategic Equipment Management," *Int. J. Struct. Civ. Eng. Res.*, pp. 137–143, 2017, doi: 10.18178/ijscer.6.2.137-143.
- [14] C. Lee, J. Won, and E.-B. Lee, "Method for Predicting Raw Material Prices for Product Production over Long Periods," *J. Constr. Eng. Manag.*, vol. 145, no. 1, pp. 1–8, 2019, doi: 10.1061/(asce)co.1943-7862.0001586.
- [15] N. Boyko and O. Lukash, "Methodology for Estimating the Cost of Construction Equipment Based on the Analysis of Important



- Characteristics Using Machine Learning Methods,” J. Eng. (United Kingdom), vol. 2023, 2023, doi: 10.1155/2023/8833753.
- [16] M. Mir, H. M. D. Kabir, F. Nasirzadeh, and A. Khosravi, “Neural network-based interval forecasting of construction material prices,” J. Build. Eng., vol. 39, no. February, p. 102288, 2021, doi: 10.1016/j.jobbe.2021.102288.
- [17] K. Petroutsatou, I. Ladopoulos, and D. Nalmpantis, “Hierarchizing the Criteria of Construction Equipment Procurement Decision Using the AHP Method,” IEEE Trans. Eng. Manag., pp. 1–12, 2021, doi: 10.1109/TEM.2021.3080117.
- [18] D. J. Edwards and G. D. Holt, “Predicting construction plant maintenance expenditure,” Build. Res. Inf., vol. 29, no. 6, pp. 417–427, 2001, doi: 10.1080/09613210152620813.
- [19] H. L. Yip, H. Fan, and Y. H. Chiang, “Predicting the maintenance cost of Construction Equipment: Comparison between general regression neural network and Box-Jenkins time series models,” Autom. Constr., vol. 38, pp. 30–38, 2014, doi: 10.1016/j.autcon.2013.10.024.
- [20] D. Berrar, “Cross-validation,” Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, no. April, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [21] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” Front. Neurobot., vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.
- [22] H. Yang et al., “Optimization of tight gas reservoir fracturing parameters via gradient boosting regression modeling,” Heliyon, vol. 10, no. 5, p. e27015, 2024, doi: 10.1016/j.heliyon.2024.e27015.
- [23] G. Guo, W. Zhu, Z. Sun, S. Fu, W. Shen, and J. Cao, “An aero-structure-acoustics evaluation framework of wind turbine blade cross-section based on Gradient Boosting regression tree,” Compos. Struct., vol. 337, no. June 2023, p. 118055, 2024, doi: 10.1016/j.compstruct.2024.118055.
- [24] A. Shehadeh, O. Alshboul, R. E. Al Mamlook, and O. Hamedat, “Machine learning models for predicting the residual value of heavy Construction Equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression,” Autom. Constr., vol. 129, no. July, p. 103827, 2021, doi: 10.1016/j.autcon.2021.103827.
- [25] Y. Alzubi, “Comparison of Various Machine Learning Models for Estimating Construction Projects Sales Valuation Using Economic Variables and Indices,” J. Soft Comput. Civ. Eng., vol. 8, no. 1, pp. 1–32, 2024, doi: 10.22115/SCCE.2023.364221.1536.
- [26] O. Ö. Ersöz, A. F. İnal, A. Aktepe, A. K. Türker, and S. Ersöz, “A Systematic Literature Review of the Predictive Maintenance from Transportation Systems Aspect,” Sustain., vol. 14, no. 21, 2022, doi: 10.3390/su142114536.
- [27] <https://medium.com/@pp1222001/outlier-detection-and-removal-using-the-iqr-method-6fab2954315d>



**Prof. Dr. Shubhalaxmi S. Joshi** is Associate Dean, Faculty Affairs at School of Computer Science and Engineering at Dr. Vishwanath Karad Maharashtra Institute of Technology World Peace University, Pune. She is a Computer Engineering graduate with MMS in SYSTEMS and Ph. D. in Computer Management from University of Pune. Doctorate is awarded on 4th Jan 2011 and recently completed M.TECH. in Comp. Science. Working with MIT and MITWPU since July 2001. She is a recipient of CSI National award for Best Paper presenter for international conferences. Published more than 33 research papers in National and International conferences and journals. Reviewed research papers for international conferences for the last two years. Published two books and 4 patents. Instrumental in organizing National level students research conference for Computer Science at Departmental level. Initiated MOU with multinational companies for various Computer Science programs.



**Poonam Katyare** is pursuing a Ph.D. program at Dr. Vishwanath Karad Maharashtra Institute of Technology World Peace University, Pune. She completed post-graduation with a degree of Master of Computer Application (MCA) in 2006. She is currently working as Senior Manager in Kloudq Technologies Ltd, Pune. Her area of interest is Data Science, Machine Learning and Data Analysis.