Utilizing Machine Learning Approach to Forecast Fuel Consumption of Backhoe Loader Equipment

Poonam Katyare, Shubhalaxmi Joshi, Mrudula Kulkarni
Department of Computer Science and Application, Dr. Vishwanath Karad MIT World Peace University, Pune, India

***Abstract*—This study addresses the challenge of forecasting fuel consumption for various categories of construction equipment, with a specific focus on Backhoe Loaders (BL). Accurate predictions of fuel usage are crucial for optimizing operational efficiency in the increasingly technology-driven construction industry. The proposed methodology involves the application of multiple machine learning (ML) models, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), and Decision Tree Regression (DT), to analyze historical data and key equipment characteristics. The results demonstrate that Decision Tree models outperform other techniques in terms of precision, as evidenced by comparative analysis of the coefficient of determination. These findings enable construction firms to make informed decisions about equipment utilization, resource allocation, and operational productivity, thereby enhancing cost efficiency and minimizing environmental impact. This study provides valuable insights for decision-makers in construction project cost estimation, emphasizing the significant influence of fuel consumption on overall project expenses.**

Keywords*—*Machine learning; construction equipment; fuel consumption

I. Introduction

The construction sector holds significant importance in driving the global economy, driving infrastructural development, and shaping urban landscapes. Within this dynamic sector, efficient management of resources, particularly fuel consumption, is paramount for ensuring project viability, sustainability, and profitability. As construction companies face increasing pressure to optimize operational efficiency while minimizing environmental impact, the need for accurate forecasting of fuel consumption has become more pronounced [19]. This research investigates the utilization of ML techniques for predicting the fuel usage of BL used at construction sites.

Backhoe Loader demonstrates versatility as it combines the functionalities of a tractor loader and a backhoe in a single machine, making them versatile for various tasks like digging, loading, lifting, and transportation. Backhoe loaders provide excavation and Loading as they excel at excavation tasks such as digging trenches, foundations, and holes, as well as loading materials onto trucks or other equipment. Their relatively compact size compared to dedicated excavators or loaders makes them suitable for job sites with limited space or access.

Incorporating ML methodologies offers a promising avenue for addressing the complexities of fuel consumption forecasting in construction. Historical data and key equipment characteristics help to develop robust predictive models capable of generating accurate forecasts [22]. Such predictions enable construction firms to make well-informed choices regarding equipment utilization, resource distribution, and project scheduling.

The utilization of ML techniques has increasingly become a focal point in the area of manufacturing and construction, by sharing constructive perceptions along with predictive analytics that boost decision-making methods. In construction, where efficiency and cost-effectiveness are paramount concerns, the ability to forecast fuel consumption for different categories of equipment holds immense significance [4][6]. As fuel represents a substantial portion of operational expenses in construction projects, accurate forecasting outcomes can result in enhanced resource allocation, refined project scheduling, and, ultimately, financial savings.

This research endeavors to explore the utilization of ML approaches specifically tailored to forecast fuel consumption across diverse categories of construction equipment [14]. ML models can offer predictive capabilities that traditional methods may struggle to achieve using past data and major features of the equipment, such as engine specifications, load capacity, and operational conditions.

The focus on ML techniques stems from their ability to handle complex datasets, identify patterns, and adapt to changing conditions, thus providing more accurate and reliable forecasts. By utilizing MLR, SVR, and DT, this research aims to identify the most efficient method for fuel consumption forecasting for the construction industry [8]. The outcomes of the study hold promise for construction companies and project managers, offering them a data-driven approach to estimating fuel consumption for various types of equipment. Such insights can inform strategic decisions related to equipment deployment, maintenance schedules, and project budgeting, ultimately contributing to improved operational efficiency and cost management. Through this research, the study connects conventional methodologies with contemporary technological innovations, fostering a pathway toward a construction sector that is both sustainable and resource-efficient. Valuable insights from this research offer constructive guidance for decision-makers involved in cost estimation and project planning within the construction industry. By shedding light on the significant role of fuel consumption in project expenses, our study contributes to the broader goal of promoting efficiency, sustainability, and cost-effectiveness in construction operations.

The manuscript is meticulously organized, beginning with Section II, which delves into comprehensive reviews of fuel consumption estimation in construction equipment using ML methods. Section III elaborates on the proposed methodology with real-time dataset preprocessing techniques and ML model implementation for forecasting. Section IV provides detailed experimental results and a discussion of the outcomes. Within Section V, the manuscript ends by presenting final thoughts with conclusions derived from the research endeavor.

II. Literature Review

This literature review covers the major use of ML techniques in various estimation processes. The systematic literature review presented in the document focuses on the application of ML methods in predictive maintenance (PdM) to enhance equipment maintenance practices in industries. The review emphasizes the importance of selecting appropriate machine-learning techniques to optimize PdM applications. Key findings include the benefits of ML in reducing maintenance costs, minimizing equipment faults, increasing production efficiency, improving operator safety, and facilitating planned management. Practical cases of effective PdM applications using ML are highlighted, showcasing the potential of these methods in preventing equipment failures and enhancing overall maintenance operations. The review also discusses the use of Random Forest (RF), Neural Networks, and SVR models with challenges and opportunities, within the realm of predictive maintenance [1].

Prior research focused on constructing a machine learning-based model to anticipate the consumption of fuel utilizing ship data of service. The research aims to enhance energy efficiency in the maritime industry and contribute to the advancement of eco-friendly ships. The study addresses existing gaps in the literature related to fuel consumption models with operational performance optimization in case of shipping sector. Utilizing statistical methods and domain-knowledge-based approaches, researchers used two models Artificial Neural Network (ANN) and MLR to forecast fuel consumption using the data gathered. The models aim to overcome multicollinearity issues and select statistically significant variables for accurate predictions. Outcomes share visions for improving energy efficiency, operational performance, and sustainability in the maritime industry via the usage of ML methods in fuel consumption prediction models [2].

Another investigation examines the significance of precisely forecasting fuel oil consumption (FOC) within the maritime sector to mitigate environmental impact, lower costs, with enhanced operational efficiency. It focuses on creating models using sensor data and weather information to forecast FOC for Very Large Crude Oil Carriers (VLCCs), emphasizing main engine consumption prediction as a key factor. Multivariate Polynomial Regression and ANN were evaluated, with eXtreme Gradient Boosting (XGBoost) showing excellent working. The study provides practical solutions for improving FOC forecasting in maritime operations, with a review of existing literature on FOC prediction methodologies and data sources. The research points out the impact related to high-quality data in boosting prediction precision [3].

Heavy-duty trucks (HDT) are significant fuel consumers within the US highway transportation system, making it essential to have a precise fuel consumption model for evaluating energy-saving strategies. The proposed model utilizes the longitudinal acceleration of the truck and is trained on field test data sets using a deep-learning neural network. By accurately estimating engine power, the model improves the fuel consumption model with reduced error rates. Including a Long Short-Term Memory (LSTM) allows for the accurate depiction of fuel consumption during engine braking scenarios, a feature not commonly found in conventional HDT fuel consumption models. The model architecture, evaluation metrics, and validation against extensive field data sets are discussed in detail. The study demonstrates that the deep learning engine power model provides accurate fuel consumption estimates and has the potential for various applications in transportation planning and traffic operation studies along with utilizing big data analytics, and a Decision Tree model [4] [5].

An available data summarization approach based on distance for developing individualized ML models for fuel consumption was presented by the existing study with a 1 km window showing high predictive accuracy for fuel consumption. Previous work includes physics-based and ML models. Technologies like V2I with dynamic traffic management can further optimize fuel efficiency at the vehicle, route, and time level. The paper utilizes a data summarization approach based on distance for developing ML models for fuel consumption. The speed and road grade of the vehicle are used in the model. The model aggregates predictors with window sizes of distance covered. Input features are adjusted to account for widely varying means in the model. The model's performance depends on the training procedure with the validation procedure [6].

The review covers carbon emission accounting models. A bottom-up procedure for detailed carbon emission analysis at the microscopic level, involving inventory analysis of building materials and energy use lists. The Economic Input-Output method is presented as a top-down approach for macro-level carbon emission analysis and ANN regression model compared with the SVR model for prediction [7]. Linear Regression, K-nearest neighbor, and ANN algorithms were used to forecast the consumption of fuel related to heavy vehicles with cross-validation to define the best model. The method provides reliable estimates of the true model error. Hyperparameters for the algorithms were defined in the inner loop, and outer loop with the model of best-performing [8]. The researcher discussed a relevant study concerning the utilization of sensor-based technologies integrated with construction equipment to capture real-time data using RF, SVR, XGBoost ensemble method, and Lasso Cross Validation (LassoCV). This data includes location tracking, movement tracking, engine condition, fuel consumption, distance traveled, and battery status. The objective is to enable managers to analyze data collected by remote sensors and make informed decisions regarding equipment performance. Additionally, remote sensing devices are utilized to track construction materials, facilitating supply chain management [9] [10] [19].

Usage of fuel consumption in heavy-loaded truck data minimized using economically optimal control strategies. Along with methodologies included nonlinear time-based formulations penalizing fuel consumption and braking effort, as well as a linear distance-based convex formulation balancing energy expenditure and velocity profile tracking [11]. Quantitative methods, such as simulation techniques, mathematical models, and decision-making methods, are commonly used in energy efficiency research. Qualitative methods like content analysis and in-depth interviews are also utilized in some studies. Different types of journal articles, including articles, conference proceedings, and review articles, contribute to the diverse methodologies used in energy efficiency research [12]. The study involves data collection for fuel consumption estimation via sensors. The methodology involved the framework design for assessing consumption based on load, slope, distance, and pavement type, enhancing optimization tools' accuracy. The IoT framework collected data from sensors in the truck, storing it for SVR, RF, and ANN algorithms' use. Sensor acquisition was implemented using Python with threads for modularity and fault tolerance [13] [14].

The study developed black-box and white-box models using RF and XGBoost to estimate the fuel consumption of ships. A simulation of several winds along with wave strategies was handled to validate the estimated outcomes, showing the effectiveness of the data-cleaning method in modeling fuel consumption accurately. The models achieved acceptable accuracy in forecasting fuel consumption, highlighting the significance of data quality and the impact of acceleration and deceleration processes on prediction reliability [15]. Related work directs on forecasting the fuel consumption of a public bus applying ML techniques. Predictor variables like distance, speed, longitude, latitude, elevation, and day of the week were used for forecasting fuel consumption. Exploratory data analysis was conducted on the dataset collected from the bus in Sri Lanka, considering factors like route, time, and terrain. ML models such as random forest, gradient boosting, and neural networks were compared for predictive accuracy, with random forest showing the best performance [16]. The study applies deep learning, and linear and non-linear simulations to fuel utilization modeling of trucks using telematic data and road characteristics. Random Forest (RF) algorithm is used to classify the influence of parameters. The research includes 14 variables significantly correlated with fuel consumption, such as gross vehicle weight, road gradient, and engine revs, which are used in developing the models. The RF algorithm allows for the selection of significant variables and is robust to outliers, making it widely used for fuel consumption predictions in various fields [17].

The methodology for data preparation and feature engineering was detailed in the study. Granville's method was mentioned to calculate the hull fouling directly [18]. Another study related to heavy vehicle consumption estimation included the ensemble method as well as consumption in commercial buildings. The study utilizes a dataset comprising relevant variables such as vehicle specifications, driving conditions, and environmental factors to train and evaluate the models. The existing study presents a comprehensive study on the application of RF, SVR, and ANN algorithms to forecast fuel consumption in commercial buildings. The research aims to address the challenges associated with accurately predicting fuel usage in diverse building types and operational contexts. The study begins by compiling a dataset comprising relevant variables such as building characteristics, occupancy patterns, and fuel consumption information. The findings indicate that ML models outperform traditional statistical methods of anticipating fuel utilization in construction devices. Research influences the understanding of forecasting systems, offering insights that can inform decision-making processes in transportation, logistics, and fleet management industries. Site managers can optimize fuel usage, reduce operational costs, and enhance sustainability in heavy vehicle operations [20] [21].

The related study delves into the utilization of ML methodologies to forecast fuel consumption in mining excavators. The study addresses the critical need for accurate fuel consumption prediction in the case of the mining sector to optimize running costs along boost efficacy. To initiate the research, a comprehensive dataset is assembled, encompassing pertinent variables such as excavator specifications, operating conditions, environmental factors, and historical fuel consumption records. This dataset involves training along with testing RF, Gradient Boosting, KNN, and MLR techniques. Various ML techniques are employed, including linear, nonlinear, and ensemble methods. Each model undergoes training on the dataset to recognize correlations of independent features with fuel consumption metrics. Machine Learning phases from exploratory data analysis to performance measurement of models evaluated. In conclusion, the research influences effective perceptions for the estimation of fuel consumption in mining excavators, offering a data-driven approach to optimize fuel usage, reduce operational costs, and improve sustainability in mining operations. By implementing ML techniques, contributors in the mining industry can make informed decisions to enhance productivity, profitability, and environmental stewardship [22].

The application of predictive maintenance techniques to enhance the reliability and efficiency of construction equipment was demonstrated. The study focuses on harnessing log data generated by the equipment during operation to predict potential failures and schedule maintenance proactively. The methodology involves collecting and preprocessing log data from construction equipment, including variables such as operating conditions, sensor readings, and maintenance logs. Feature engineering techniques are employed and prepared for model training. RF, Logistic Regression, and XGBoost algorithms are applied to the preprocessed data with classification models. These models are trained to classify equipment conditions as either normal or indicative of a potential failure. Cross-validation techniques and performance are employed for the predictive maintenance models. The outcomes exhibit the possibility and efficacy related to the predictive maintenance of log records from construction equipment [23].

The authors investigate the enhancement of thermal conductivity in green buildings through the application of nano insulations. The study employs Gaussian Process Regression (GPR), SVR, and DT methods to optimize the selection and deployment of nano insulations for improved thermal insulation performance. The methodology involves collecting data on various nano insulation materials, including their properties, compositions, and thermal conductivity characteristics. ML algorithms are utilized to analyze this data and identify correlations between nano insulation attributes and thermal conductivity improvement. The most influential factors contributing to thermal conductivity enhancement were observed. Regression algorithms are employed to develop predictive models for estimating thermal conductivity improvement based on the selected features. The algorithms are tested using experimental data of nano insulation performance in real-world green building scenarios. The findings related to the study provide insights into the optimal selection and deployment of nano insulations for improving thermal conductivity in green buildings. It can reduce heating along cooling costs, and promote sustainability in building design and construction [24].

The existing study explores vehicle trip data for model estimation with artificial intelligence methods to analyze trip-specific variables and accurately forecast fuel consumption. Using ANN, MLR, and RF methods, investors in the transportation industry can optimize fuel usage, improve route planning, and reduce operational costs for heavy-duty vehicle fleets [25].

Another study investigates a method for predicting vehicle fuel consumption using driving behavior data obtained through smartphones. The study utilizes sensor-based data embedded in smartphones for analyzing driving patterns and developing RF, SVR, and Back Propagation neural network predictive models for fuel consumption. The methodology involves collecting driving behavior data from smartphones, including variables such as acceleration, braking, speed, and route information. Feature engineering techniques are applied to preprocess the data and extract relevant features indicative of fuel consumption patterns. This reveals the feasibility of using smartphone-based driving behavior data to predict vehicle fuel consumption accurately and leads towards fuel-efficient driving strategies, optimize vehicle performance, and reduce fuel costs for drivers and fleet operators [26].

A. *Research Gap*

As is commonly understood, predicting fuel usage may depend on factors such as route features, vehicle specifications, and driving habits. This study tackles the scientific hurdle of identifying which factors have the highest influence related to fuel consumption in vehicles. A significant challenge lies in the difficulty of obtaining accurate consumption from equipment. Reliable consumption data of fuel are essential for accurately training ML algorithms, making it imperative to secure these data with certainty. However, it's common for this information to be unreliable, often underestimating the actual values.

The complexity of advanced recent tools makes it impractical for integration into such uses. Nevertheless, with the rapid advancement of remote monitoring systems, ML applications in this field have achieved success across various sectors. These encompass earthworks productivity, slope safety, jet grouting compressive strength, as well as pavement management and monitoring. The Indian construction sector encounters difficulty in accurately estimating fuel consumption owing to its limited digitalization. Estimating the fuel consumption necessary for construction equipment at job sites is imperative.

III. PROPOSED METHODOLOGY

The objective of the proposed study is to forecast the fuel consumption of construction equipment from the IoT-enabled sensor data received from devices. This study used the highly utilized equipment on the job site. These include the Backhoe Loader equipment data. Real-time data is collected from the smart sensing devices in daily behavior. The systematic flow of the proposed study is represented in Fig. 1. Fuel consumption forecasting flow diagram. The proposed system is majorly distributed in phases of data preprocessing, feature computation, and selection, forecasting of data, and performance evaluation phase.

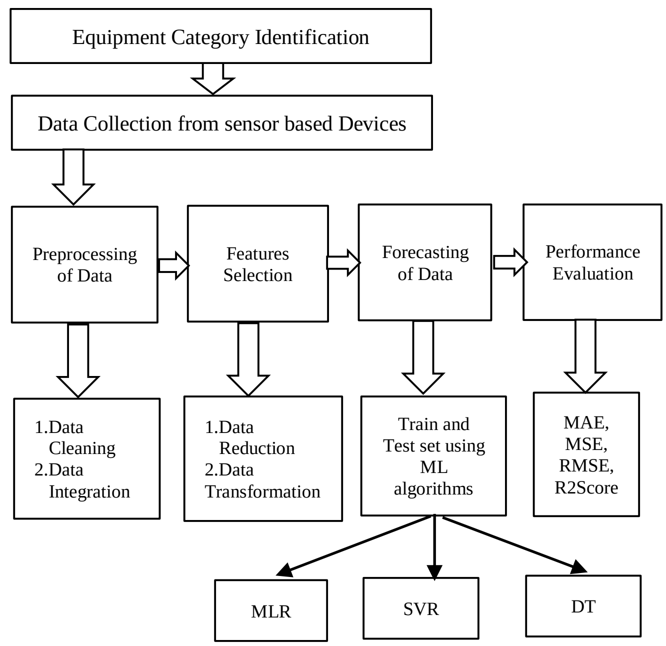

Fig. 1. Fuel consumption forecasting flow diagram.

A. *Data Collection from Devices*

IoT-enabled smart sensor data from devices is captured and daily logs are received from March 2022 to January 2023 for the construction equipment as Backhoe Loader. The data contains features related to fuel, features for operating run hours, features for distance, and speed. Distance covered by the equipment is captured from the latitude and longitude, Fuel-related features are captured from fuel level sensors while actual operating run hours, start run hours, and end run hours are captured from the hour meter sensor.

IoT-enabled smart sensing devices are attached to the construction equipment which help to capture data through IoT gateways. Onboard sensors and external sensors are capable of sharing these features by transmitting the data to the server using IoT gateways. Start fuel level and end fuel level values are captured. Start run hours and end run hours values are captured. Distance is computed using latitude and longitude values along with speed calculations from distance computed value.

The statistics of the data are explored in Table I.

TABLE I. DATA CHARACTERISTICS

Main Features	Mean	Std. Dev	Min	Max
Trip Distance	6.55	5.41	2.11	31.75
Run Hours	4.88	2.66	0.08	12.53
Average Speed	5.96	3.75	2.09	21.04
Fuel consumption	18.21	10.78	0.2	55

B. *Preprocessing of Data*

Preprocessing of data is crucial for ensuring with reliability of the dataset. Adequate and proper data is responsible for the effective and accurate ML model building. Data Preprocessing involves data cleaning along with data integration steps.

The data cleaning (see Fig. 2) step handles the duplicate data and noisy data. Equipment duplicate raw data points were removed from the iterations.

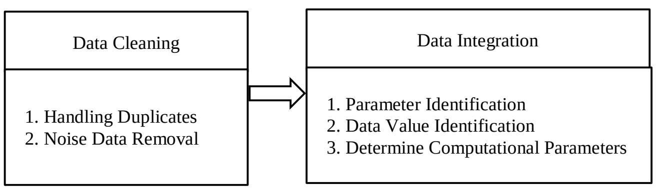

Fig. 2. Data preprocessing steps.

Noisy data is identified and removed using the outlier detection method. The Z-score outlier removal method is a statistical technique used to discover as well as eliminate outliers built on respective variance from average in terms of standard deviations. The importance of outlier detection and removal confirms integrity with consistency related to research findings. the concept of Z-scores and how they are used to measure the deviation of individual data points. The formula for calculating the Z-score of a data point is explained in Eq. (1).

$$Z = \frac{(D-M)}{\mathrm{SD}} \tag{1}$$

Where,

D - Data point,

M - Mean,

SD - Standard deviation.

This method highlights the number of standard deviations of data points left with the average data point. A value of 0 reveals the intimation of lying the point exactly at the mean, positive scores signify above the average data point, and negative scores denote below the average point. The threshold criterion commonly used for outlier detection is based on Z-scores, such as considering points beyond a certain threshold as outliers. The process of identifying outliers using Z-scores is more effectively used.

The data Integration step involves the parameter identification for where the dataset contains MachineId as well as Equipment number, Entrydate as Date, these types of parameters were present in the dataset Correct parameters are identified and integrated into the dataset. Data value identification is performed on the datatypes of the parameters to collect all parameters in the same data type format such as for Entrydate in DateTime format, other Run hours were present in DateTime. To apply ML models numeric data need to calculate. All parameters in numeric format ensure scaling on a similar scale for feature engineering, leading to performing ML models effectively.

C. *Features Selection*

Feature selection is an essential phase for the machine learning pipeline that involves selection with a group of significant features. Subgroups of features were identified from the original parameter set. This step ensures the improvement in the model's performance, reduces overfitting problems, and boosts interpretability. Complex feature spaces with high dimensions may result in heightened computational demands, diminished model efficacy, and susceptibility to overfitting. A correlation matrix is used to select highly correlated features. Distance covered by equipment, total run hours of that equipment with average speed of the equipment are the highest correlated parameters for fuel consumption prediction.

D. *Forecasting of Data using ML Model*

ML models play a crucial role in predicting values across various domains due to several important reasons. Machine learning models demonstrate proficiency in recognizing intricate patterns and correlations within datasets that conventional statistical approaches may overlook. This capability allows them to capture intricate associations of input parameters with the target parameter, enabling accurate predictions. ML models are highly adaptable. This flexibility allows them to effectively model a wide range of real-world phenomena and achieve estimates with diverse data. ML models can scale efficiently and manage substantial amounts of data, making them suitable for applications where massive datasets are involved. Whether it's analyzing millions of transactions in finance or processing vast amounts of sensor data in IoT applications, ML models can effectively manage the workload. Many real-world phenomena exhibit non-linear associations of input parameters with the target parameter. Non-linear ML algorithms are capable of capturing and modeling these non-linear relationships, allowing for more accurate predictions compared to linear models.

Some ML models can constantly be trained for new instances of data that reflect estimates that persist appropriately with accurate behavior in vibrant circumstances. ML models can automate the process of forecasting, removing the necessity for manual examination and human involvement in repetitive tasks. This automation not only saves time and resources but also reduces the likelihood of errors associated with manual prediction methods. ML methods share transparency that allows stakeholders to understand how predictions are made and gain perceptions of the factors influencing the anticipated consequences.

1) Multiple Linear Regression (MLR): It is a foundational and extensively employed approach for modeling the association between a dependent variable and numerous independent variables, assuming a linear correlation between them, implying that they can be represented as a straight line. MLR is a statistical method used to ascertain the quantitative association among two or more variables [2] [22] [25]. In regression analysis, target variables are observed or measured, while the independent variables are factors considered to significantly impact the target variable under evaluation. Predictions can be made by estimating the relationships between variables through analysis.

2) Support Vector Regressor (SVR): The Support Vector Regression (SVR) function denotes the relationship between dependent and independent parameters while minimizing error. Its core objective is to identify a hyperplane with the maximum number of support vectors within the decision boundary, allowing for continuous value predictions. This involves employing kernels, a set of numerical operations, to transform input data into meaningful configurations. SVR aims to fit between the boundary lines and the hyperplane, adjusting coefficients within a specified tolerance margin [1][7][13][14]. SVR computes a hyperplane to fit the training data while minimizing margins, aiming to find coefficients and a bias term that reduces the variance of anticipated value with original values within a tolerance margin. This optimization problem is typically formulated as a quadratic programming problem and solved using optimization techniques. Kernel functions such as sigmoid, linear, and polynomial are commonly used, chosen based on the complexity of feature relationships and data nature. SVR is highly effective for datasets with dense relationships and high-dimensional feature spaces, ensuring robust predictions and reduced sensitivity to outliers [19][21[24][26].

3) Decision Tree Regressor (DT): This algorithm is widely employed in supervised learning, supporting both regression and classification analyses [4][5]. It operates by sequentially portraying decisions and their potential outcomes, encompassing chance events, asset prices, and utility considerations. This model utilizes conditional control statements in the form of branching rules, making it a versatile tool for analyzing various types of data. A nonparametric supervised learning technique, the DT algorithm constructs a tree-like structure comprising root nodes, interior nodes, and leaf nodes. Each branch and leaf node represent decision criteria and predicted outcomes, forming a hierarchical representation of the data. The DT regressor essentially represents a piecewise constant function, partitioning the feature space into non-overlapping regions, each linked with a constant predicted value. The final prediction for a given input sample is the sum of the predicted values of the leaf nodes to which the sample belongs [24].

E. Performance Evaluation

Measuring the performance of ML models is necessary for assessing their efficiency and determining their appropriateness related to real-world purposes. Measuring metrics are generally used to estimate regression models, with Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) score.

1) Mean Squared Error (MSE): The Mean Squared Error (MSE) quantifies the average of the squared variances between predicted and actual values, assigning more significance to larger deviations, thus rendering it responsive to outliers. Calculated by averaging these squared variances across the dataset, lower MSE values signify superior model accuracy.

2) Mean Absolute Error (MAE): MAE evaluates the average absolute disparity between predicted and actual values, offering less sensitivity to outliers compared to MSE since it does not square the errors. MAE is computed by averaging the absolute disparities between predicted and actual values across the entirety of the dataset. Similar to MSE, superior model performance is indicated by lower MAE values.

3) Root Mean Squared Error (RMSE): RMSE, being the square root of MSE, offers a comprehensible scale for interpretation. It gauges the average magnitude of errors in units akin to the dependent variable. RMSE is calculated by taking the square root of the MSE. Lower RMSE values indicate better model performance, and it is often preferred when errors are expected to be normally distributed.

4) Coefficient of determination (R2) score: The R2 score signifies the fraction of the variability in the dependent variable elucidated by the independent variables in the model. Its scale spans from 0 to 1, where a score of 1 denotes an impeccable fit, while 0 suggests that the model fails to elucidate any variability. In instances where the model performs poorer than a horizontal line, the R2 score can be negative. Enhanced model performance is denoted by higher R2 values, with 1 representing the pinnacle of performance.

When assessing machine learning models, it's crucial to examine a blend of these metrics to obtain a well-rounded view of their performance. While MSE, MAE, and RMSE offer insights into error magnitudes, the R2 score quantifies the model's overall adequacy of fit. By interpreting and comparing these metrics, researchers, and practitioners can make informed decisions about model selection and refinement to achieve optimal results in various applications.

IV. Results and Discussion

This study undertook an expectation task aimed at examining various alternative models for forecasting fuel consumption of related construction equipment in selected datasets. The study evaluated the appropriateness of MLR, SVR, and DT models for this purpose. Using authentic dataset inputs, regression models were trained and assessed. MSE, MAE, RMSE, and R2 scores are used for measuring the performance of ML models. Table II represents the performance measurement of ML models as MLR, SVR, and DT are used to forecast the fuel consumption of backhoe loader equipment.

TABLE II. PERFORMANCE MEASUREMENT OF MODELS

Model	MAE	MSE	RMSE	R2
MLR	9.18	152.54	12.35	0.43
SVR	9.62	259.14	16.09	0.56
DT	9.70	227.38	15.07	0.61

As per the performance metrics offered in Table I, all three fuel consumption prediction models, namely MLR, SVR, and DT, demonstrate relatively similar predictive abilities. MLR achieved the lowest Mean Absolute Error (MAE) of 9.18, indicating its ability to predict fuel consumption with the smallest average absolute deviation from the actual values. DT achieved the top R2 score with 0.61, suggesting the largest proportion of variance in the fuel consumption data among the three models. SVR yielded intermediate results between MLR and DT relations with MAE, MSE, RMSE, and R2 scores, demonstrating moderate performance in predicting fuel consumption.

The relatively low R^2 (0.43) indicates that MLR explains only 43% of the variance in the target variable. MAE and RMSE values indicate the average errors, with RMSE being higher due to its sensitivity to larger errors (squared differences). MLR is a linear model that attempts to establish a linear relationship between the input features and the target variable. It minimizes the sum of squared residuals to find the best-fitting linear hyperplane in the feature space. SVR has a higher R^2 (0.56) compared to MLR, indicating it captures more variance in the target variable (56%). This suggests SVR handles non-linear relationships better than MLR. The higher MAE and RMSE values compared to MLR might be due to SVR being more sensitive to outliers or the choice of kernel and its parameters. Despite capturing more variance (higher R^2), the higher errors (MAE, MSE, RMSE) suggest that the SVR model might not be well-tuned or that it could be overfitting/underfitting the data. SVR aims to find a hyperplane in a high-dimensional space that maximizes the margin between the hyperplane and the data points. It uses kernel functions to handle non-linearity. It focuses on minimizing a margin of error and uses support vectors to define the hyperplane. DT has the highest R^2 (0.61), meaning it explains 61% of the variance in the target variable, suggesting it captures the data's underlying structure better than MLR and SVR. The MAE is slightly higher than MLR, but the significant improvement in R^2 indicates that DT models better handle non-linearity and interactions between features. The MSE and RMSE values are lower than SVR but higher than MLR, which may indicate that DT captures more variance. DT models split the data into subsets based on feature values, creating a tree-like structure where each node represents a feature split that contributes to reducing the target variance. It is a non-linear model that can capture complex relationships by recursive partitioning.

Overall, the Decision Tree model appears to offer the best balance between accuracy and explanatory power among the three models evaluated. However, further analysis and comparison with additional metrics may be necessary to make a conclusive determination about the optimal model for predicting fuel consumption.

V. CONCLUSION

This study demonstrates the effectiveness of machine learning approaches in forecasting fuel consumption for construction equipment, with a particular focus on Backhoe Loader (BL) fuel consumption estimation. As the construction industry increasingly integrates technology, accurate predictions become essential to optimizing fuel usage and operational efficiency. Through this analysis, accurate forecasts of fuel consumption are generated, empowering construction companies to facilitate well-informed decisions concerning equipment usage, resource allocation, and equipment productivity. The study employed Multiple Linear Regression, Support Vector Regression, and Decision Tree Regression models, trained on the dataset. Comparative analysis of the coefficient of determination reveals that the Decision Tree technique yields more precise results compared to other models, as indicated by measures of accuracy. The findings of this study provide valuable insights for decision-makers involved in cost estimation for construction projects, highlighting the significant role of fuel consumption in project expenses. By employing advanced ML techniques, construction operations can be enhanced in terms of efficiency, sustainability, cost-effectiveness, and environmental impact mitigation.

AUTHORS' CONTRIBUTION

P.K.: Conceptualization, methodology, writing the original draft. S.J..: methodology, writing—original draft, writing—review and editing, visualization. M.K.: conceptualization, supervision, investigation, writing—review, and editing.

DATA AVAILABILITY STATEMENT

The authors do not have permission to share data.

REFERENCES

[1] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of ML methods applied to predictive maintenance," Comput. Ind. Eng., vol. 137, no. August, p. 106024, 2019, doi: 10.1016/j.cie.2019.106024.

[2] Y. R. Kim, M. Jung, and J. B. Park, "Development of a fuel consumption prediction model based on ML using ship in-service data," J. Mar. Sci. Eng., vol. 9, no. 2, pp. 1–25, 2021, doi: 10.3390/jmse9020137.

[3] C. Papandreou and A. Ziakopoulos, "Predicting VLCC fuel consumption with ML using operationally available sensor data," Ocean Eng., vol. 243, no. June 2021, p. 110321, 2022, doi: 10.1016/j.oceaneng.2021.110321.

[4] Y. Kan, H. Liu, X. Lu, and Q. Chen, "A deep learning engine power model for estimating the fuel consumption of heavy-duty trucks," 6th IEEE Int. Energy Conf. ENERGYCon 2020, pp. 182–187, 2020, doi: 10.1109/ENERGYCon48941.2020.9236554.

[5] M. Zhang, N. Tsoulakos, P. Kujala, and S. Hirdaris, "A deep learning method for the prediction of ship fuel consumption in real operational conditions," Eng. Appl. Artif. Intell., vol. 130, no. June 2023, p. 107425, 2024, doi: 10.1016/j.engappai.2023.107425.

[6] A. Schoen, A. Byerly, B. Hendrix, R. M. Bagwe, E. C. Dos Santos, and Z. Ben Miled, "A ML Model for Average Fuel Consumption in Heavy Vehicles," IEEE Trans. Veh. Technol., vol. 68, no. 7, pp. 6343–6351, 2019, doi: 10.1109/TVT.2019.2916299.

[7] T. H. Kappen and L. M. Peelen, "A Review of Building Carbon Emission Accounting and Prediction models," Curr. Opin. Anaesthesiol., vol. 29, no. 6, pp. 717–726, 2016, doi: 10.3390/buildings13071617.

[8] V. Bodell, L. Ekstrom, and S. Aghanavesi, "Comparing ML Estimation of Fuel Consumption of Heavy-Duty Vehicles," vol. 15, no. 2, pp. 97–101, 2021.

[9] P. Katyare and S. S. Joshi, "Construction Industry Digitization using Internet of Things Technology," vol. 3307, pp. 243–249, 2022.

[10] P. Katyare and S. Joshi, "Construction Productivity Analysis in Construction Industry: An Indian Perspective," pp. 133–142, 2022, doi: 10.1007/978-981-19-0863-7_11.

[11] J. Borek, B. Groelke, C. Earnhardt, and C. Vermillion, "Economic optimal control for minimizing fuel consumption of heavy-duty trucks in a highway environment," IEEE Trans. Control Syst. Technol., vol. 28, no. 5, pp. 1652–1664, 2020, doi: 10.1109/TCST.2019.2918472.

[12] F. S. Hafez et al., "Energy Efficiency in Sustainable Buildings: A Systematic Review with Taxonomy, Challenges, Motivations, Methodological Aspects, Recommendations, and Pathways for Future Research," Energy Strateg. Rev., vol. 45, no. November 2022, p. 101013, 2023, doi: 10.1016/j.esr.2022.101013.

[13] G. Pereira, M. Parente, J. Moutinho, and M. Sampaio, "Fuel consumption prediction for construction trucks: A noninvasive approach using dedicated sensors and ML," Infrastructures, vol. 6, no. 11, 2021, doi: 10.3390/infrastructures6110157.

[14] M. A. Hamed, M. H. Khafagy, and R. M. Badry, "Fuel Consumption Prediction Model using ML," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 11, pp. 406–414, 2021, doi: 10.14569/IJACSA.2021.0121146.

[15] X. Xie, B. Sun, X. Li, T. Olsson, N. Maleki, and F. Ahlgren, "Fuel Consumption Prediction Models Based on ML and Mathematical Methods," J. Mar. Sci. Eng., vol. 11, no. 4, 2023, doi: 10.3390/jmse11040738.

[16] S. Wickramanayake and D. H. M. N. Bandara, "Fuel consumption prediction of fleet vehicles using ML: A comparative study," 2nd Int. Moratuwa Eng. Res. Conf. MERCon 2016, pp. 90–95, 2016, doi: 10.1109/MERCon.2016.7480121.

[17] F. Perrotta, T. Parry, and L. C. Neves, "Application of ML for fuel consumption modelling of trucks," Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017, vol. 2018-January, pp. 3810–3815, 2017, doi: 10.1109/BigData.2017.8258382.

[18] A. Fertier, A. Montarnal, S. Truptil, and F. Bénaben, "Interpretable, Data-driven Models for Predicting Shaft Power, Fuel Consumption, and Speed Considering the Effects of Hull Fouling and Weather Conditions," Decis. Support Syst., no. January, p. 113260, 2020, doi: 10.1016/j.nhres.2023.06.006.

[19] 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.

[20] H. Almér, "ML and statistical analysis in fuel consumption prediction for heavy vehicles," Master Sci. Thesis, p. 75, 2015.

[21] A. Rahman and A. D. Smith, "Predicting fuel consumption for commercial buildings with ML algorithms," Energy Build., vol. 152, pp. 341–358, 2017, doi: 10.1016/j.enbuild.2017.07.017.

[22] A. Saim, F. Kumah, and M. Oppong, "Predicting Mining Excavator Fuel Consumption using ML Techniques," Adv. Eng. Des. Technol., vol. 2, pp. 12–23, 2020.

[23] B. M. Kotriwala, "Predictive Maintenance of Construction Equipment using Log Data A Data-centric Approach (MSc thesis)," Degree Proj. Technol., 2021.

[24] M. Ghalandari et al., "Thermal conductivity improvement in a green building with Nano insulations using ML methods," Energy Reports, vol. 9, pp. 4781–4788, 2023, doi: 10.1016/j.egyr.2023.03.123.

[25] S. Katreddi and A. Thiruvengadam, "Trip based modeling of fuel consumption in modern heavy-duty vehicles using artificial intelligence," Energies, vol. 14, no. 24, 2021, doi: 10.3390/en14248592.

[26] Y. Yao et al., "Vehicle Fuel Consumption Prediction Method Based on Driving Behavior Data Collected from Smartphones," J. Adv. Transp., vol. 2020, 2020, doi: 10.1155/2020/9263605.