Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 3, pp. 427–437
DOI : 10.47974/JIOS-1363

Real time data modeling for forecasting fuel consumption of construction equipment using integral approach of IoT and ML techniques

Poonam Katyare *
Shubhalaxmi S. Joshi †
Department of Computer Science
Dr. Vishwanath Karad MIT World Peace University
Pune
Maharashtra
India

Sheetal Rajapurkar §
School of Computer Science
Dr. Vishwanath Karad MIT World Peace University
Pune
Maharashtra
India

Abstract

The Internet of Things (IoT) plays a vital role in the automation of Construction Industry. The real time data of the construction equipment is monitored using IoT devices. An integral approach of IoT based sensing data and Machine Learning (ML) models helps to predict the fuel consumed by the equipment. This paper presents the real time data modeling to estimate the fuel consumption for a trip travelled by the construction equipment using IoT enabled remote data along with machine learning algorithms. The Random Forest, Extreme Gradient Boosting (XGBoost) ensemble methods and Lasso Cross Validation (LassoCV), Support Vector Machines Regression models are used in this study. These models are fitted on dataset and splits the data into training and testing data. Based on the comparative analysis of coefficient of determination, LassoCV technique produces more accurate results along with the other models using Models' accuracy measures. This study would help the decision

* *E-mail:* `poonamkatyare@gmail.com (Corresponding Author)`
† *E-mail:* `shubhalaxmi.joshi@mitwpu.edu.in`
§ *E-mail:* `sheetal.rajapurkar@mitwpu.edu.in`

makers for cost estimation of the construction project which includes fuel consumption as major component of cost.

Subject Classification: *60G25 Prediction theory, 60G42 Martingales with discrete parameter.*

Keywords: *Construction industry, Internet of Things (IoT), Machine learning (ML).*

1. Introduction

In the current era, the Construction Industry is significantly contributed to the economic growth of the country. It has been observed that the growth in Construction Industry is outstanding and remarkable with the fast and successful completion of the Construction projects. Construction industry is highly dependent on the Social and Infrastructure development [20]. The construction equipment, workers and resources are the key factors used in the construction projects. The construction industry regularly requires an optimized tool for cost estimation to maximize the profit and making accurate decisions. Fuel is highly consumed for utilization of the construction equipment leading to the largest cost in the construction projects. Essentially, all the decisions for heavy construction equipment related to their scheduling, allocation and performance measurement depends on fuel consumption. Traditionally, the equipment performance handbook or the guidance from experienced professionals used for estimating the fuel consumption required for the construction equipment on the construction site. [8] Indian Construction industry faces a challenge in the estimation of fuel consumption due to the lack of digitization [15]. There is a need of estimation of fuel consumption required for construction equipment on job site.

Internet of Things (IoT) is one of the emerging technologies in the world which enable the use of wireless sensing mechanism to capture data remotely. The Real time construction equipment data is easily available and stored data is used for analysis and decision making by the top management. Future prediction becomes easy by using the historical data of construction equipment. Machine learning (ML) is a technology which emphasis machines relearn from experience. ML plays a key role in decision making for business. ML can be defined as a computer program which perform operations from the training data or historical data. Major two types of ML algorithms are present such as Supervise Learning and Unsupervised Learning. Supervised learning is used to classify the data or to predict the output variable which depends on inputs for the algorithm. Unsupervised learning algorithms used to analyze the given input data and to find the patterns or behavior of the input variables. This paper highlights the use of supervised learning to estimate the Fuel consumption. Supervised learning algorithms leading towards creation of a mathematical model of a set of data that contains both the inputs and the expected output. The data is divided into training data and testing data. Predicted output is evaluated on the inputs which are not the part of training data. An accurate output depends on an optimized algorithm with an improved accuracy.

This study provides an integral approach of IoT enabled real time data of Transit Mixer and ML algorithms applications on the data. The main objective of this study is to forecast trip wise fuel consumption of transit mixer on the real time data provided by remote server. This study is performed with the multiple regression algorithms as Support Vector Machine, LassoCV along with some ensemble bagging and boosting method as XGBoost and Random Forest. ML Models are fitted on the dataset and best algorithm for this transit mixer data with the highest accuracy (LassoCV) is selected and used for the forecasting. This study provides the methodology to train the ML models and predict the fuel consumption on the new instance of the data. This study will be used to estimate the cost associated with the transit mixer device working on the site and to analyze the performance of the transit mixer and identify the pattern of working of this device for the contractors.

2. Literature Review and Limitations of Existing Techniques

Researcher presented the study related to establishing and examining the idle time of construction equipment and notify the user if the threshold value set for the idle time was exceeded [1] This study used sensor data and threshold values for idling as input and produces alerts if threshold time was exceeded and estimates fuel consumption without the use of ML techniques. Similarly, one presented the study of predicting the maintenance cost of the construction equipment by using time series approaches [2] where monthly data is collected from contractors. Fuel consumption of long-distance public bus with inward and outward journey off traveling using ML algorithms found in the study [3][9] with small dataset of inward and outward journey. Literature contributed towards prediction of fuel consumption of truck, Container, ships along with hourly fuel consumption at different time zone places with ML techniques [4-12][17] [19] with different features like road, acceleration, weather, travel, speed, load and the duration of the route. Some study demonstrated prediction of Residual value of construction equipment [14]. It has been observed that, very few study related to automation of construction sites with the integration of IoT enabled sensing data and ML approaches to estimate fuel consumption of the construction equipment is present. Researchers used ML approaches on manual data while others used sensing data of ships, public bus, trucks, containers to predict maintenance cost, Residual value, idle time, and fuel. Existing ML algorithms as Decision Tree, Random Forest and Neural Network is inferior in case of large datasets. ML methods is used to predict the level of emissions from heavy construction equipment using advanced techniques as cameras and computer vision which is very costly to use [16][18].

3. Proposed Objective and Methodology

The main objective of the study is to forecast the fuel consumption during a trip of the construction equipment (Transit mixer) along with real time IoT based data and ML approaches. The proposed methodology captures the real time data from the sensors attached to the construction equipment. The collected data is preprocessed and partitioned into training and testing data set. The ML models are applied on data set and the best model is selected for fuel consumption prediction. The proposed model for fuel consumption estimation is given below.

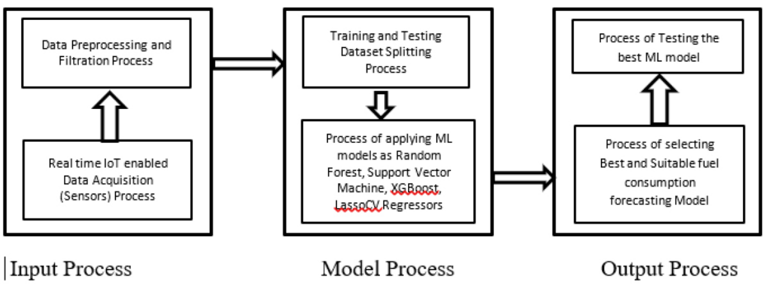

Figure 1

Model for fuel consumption estimation

3.1 *Real time IoT-enabled Data Acquisition and Preprocessing as Input Process*

The trip wise fuel consumption data of Transit Mixer were collected from 12 April 2022 to 18 November 2022. Many features of transit mixer related to trips are recorded such as date time stamp, equipment location by latitude and longitude, trip start time trip end time, total trip duration in seconds, fuel consumption from start point fuel level value the endpoint fuel level value, total engine run hours from start point and end point engine run hour value, distance covered in a trip, average speed during a trip of transit mixer and battery voltage. All the features are extracted from onboard sensors and some external sensors like Fuel Level sensor, Hour Meter, Bluetooth sensor and Global Positioning System (GPS), wireless network, Geographic Information System (GIS) attached to the device. The data transmission from device to the data server trough the IoT gateways happened in every minute with 308120 datapoints for 214 days. Trip wise 950 data points were captured from the start location to end location information which consist of the run hour data, fuel consumption, time interval, distance, and speed along with position of each trip of transit mixer where fuel consumption is calculated by the difference between the end point fuel level with start point fuel level with milliliter unit.

Data Preprocessing of the raw data is performed by removing outliers or noise values and duplicates from the dataset. Data points with less than two-kilometer distance trip proposed as outlier and removed from the dataset before the model construction. Standardization of feature scaling process is applied to match the input or independent features with the target feature before ML process. Table1 denotes the descriptive Statistics of the dataset features as

Table 1

Descriptive Statistics

<table>
<tr><th>Supportive Features</th><th>Main Features</th><th>Mean</th><th>Std. Dev</th><th>Min</th><th>Max</th></tr>
<tr><td rowspan="4">Trip Start & End Latitude
Trip Start & End Longitude
Trip Start & End Engine Run Hour
Trip Start & End Fuel Level
Trip Start & End Time
Trip Max Speed</td><td>Trip Distance</td><td>11.53</td><td>8.94</td><td>2.11</td><td>57.97</td></tr>
<tr><td>Engine Run Hours</td><td>0.72</td><td>0.45</td><td>0.08</td><td>4.44</td></tr>
<tr><td>Average Speed</td><td>30.32</td><td>9.78</td><td>2.09</td><td>67.31</td></tr>
<tr><td>Fuel consumption</td><td>5.12</td><td>3.91</td><td>0.2</td><td>27.3</td></tr>
</table>

3.2 *Developing Machine Learning Models as Model Process*

Machine Learning is a technique for transforming data into actionable knowledge. Various supervised ML techniques are available for prediction which associated with the historical data for predicting the new instances of data points with the relation of dependent variables along with independent input values [13]. ML model follows the Data collection, Data preprocessing, spiling data into training and testing dataset, apply the model on training dataset and evaluate the performance of the model. Finally best model with highest accuracy is selected for predicting new instance of the data. We used the different ML Regression algorithms such as Random Forest, Support Vector Machine, XGBoost and LassoCV models for forecasting the trip wise fuel consumption of transit mixer.

3.2.1 Random Forest and XGBoost

Random Forest and XGBoost are an ensemble learning method is the way of collectively associate all the predictions together. This activity will help to create more accurate ML model. Ensemble methods can decrease variance using bagging approach, bias using a boosting approach, or improve predictions using stacking approach. Bagging and Boosting are ensemble learning methods. Bagging collects more than one learner to reduce the variance of predictions. Random Forest is the example of ensemble learning technique which trains the number of decision tree on different random subset data. Voting is performed on final prediction. Boosting is a method for reducing the bias and variance. Boosting algorithm help in tracing the failed model with incorrect prediction. The overfitting problem can be resolved in Boosting algorithms. XGBoost (Extreme Gradient Boosting) model is used in Boosting technique. Model performance and speed is increased using gradient-boosting decision trees implemented in XGBoost method.

3.2.2 Support Vector Regressor (SVR) and LassoCV

"Support Vector Machine" (SVM) is a supervised ML algorithm which can be introduced for regression as Support Vector Regressor (SVR). Hyperplane with maximum number of points within **decision boundary line or** support vectors should be present within that boundary line is main objective of SVM. To predict the continuous values hyperplanes decision boundaries are used. Input data is converted into the required form using this set of mathematical functions defined as kernels [4]. SVR tries to fit within the distance between the hyperplane and boundary lines. With the help of linear kernel, the prediction of well consumption of transit mixer is possible. Liner Regression is also performed by Lasso model which is known to reduce the number of features and to prevent overfitting in the model. Lasso denotes as Least Absolute Shrinkage and Selection Operator (Lasso). When cross validation applied on Lasso it denotes as LassoCV which has one of the parameter inputs as "cv" to signifies several folds to be measured while applying cross-validation. LassoCV is linear regression mode with regularization technique that regularizes the coefficient estimates towards zero which is an extra information to prevent overfitting. Study focused on LassoCV performance with best accurate model for prediction.

4. Performance measurement of Machine Learning Models as Output Process

The linear relation between the dependent and independent features is analyzed using the Correlation matrix. In our study, the distance covered during trip and engine run hours captured during a trip are observed as the strong correlated parameters with our target fuel consumption feature (Figure 2).

The Random Forest, Support Vector Machine, XGBoost and LassoCV regressor models were trained and tested on the proposed real dataset. Performance measurement of the regression models is evaluated using Mean Squared Error (MSE), Mean absolute error (MAE), Root Mean Squared Error (RMSE), and Coefficient of determination(R-Squared) where MAE denotes the average of difference between the original and estimated values, MSE denotes the average of squared difference between the original and estimated values. RMSE (Root Mean Squared Error) is the error rate by the square root of MSE. R-squared (Coefficient of

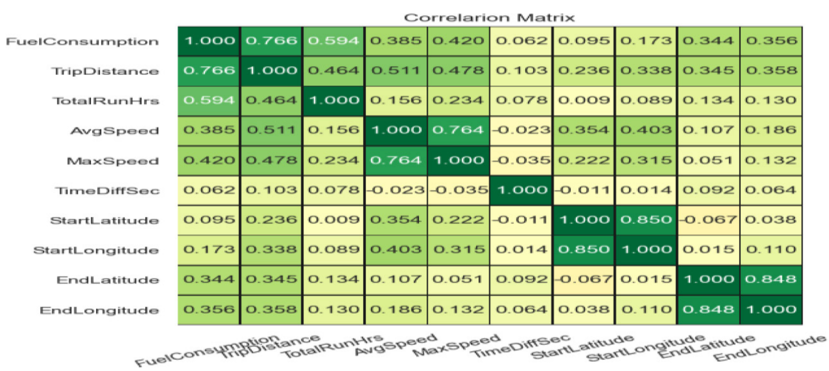

Figure 2

Correlation Matrix

Table 2

Performance Measurement of Models

Model	MAE	MSE	RMSE	R^2
XGBoost	1.26	2.89	1.7	0.79
LassoCV	0.09	0.01	0.12	0.99
Random Forest	1.31	3.38	1.84	0.75
Support Vector Machine Regressor	1.47	4.94	2.22	0.64

determination) denotes the coefficient of estimated values compared with the original values. The percentages signify the value from 0 to 1. Table2 denotes the comparative analysis of the performance of the models. The higher the value denotes better performance of the model.

LassoCV lead to the best performance model for our study as compared to the XGBoost, Random Forest and Support Vector Machine (Table2). LassoCV with k-fold cross validation approach of selecting data performs better as it evaluates the entire training data and executes a grid search to find an ideal choice of the regularization strength.

5. Discussion and Conclusion

The study developed Machine Learning models for forecasting trip wise fuel consumption of Transit Mixer with IoT enabled Construction equipment real time data . The topmost important features extracted from ML model are the total distance covered during trip, total engine run hours and average speed. This study adopted that data and applied Supervised ML models as Support Vector Machine Regressor (SVR), Lasso Cross Validation (LassoCV) model and some ensemble methods with regression as Random Forest, XGBoost for forecasting trip wise fuel consumption. Amongst these models LassoCV provides the accurate results for prediction.

Future study would focus to predict the fuel consumption on the various types of construction equipment and provides the data analysis which results in making business decisions.

6. Acknowledgment

Authors are grateful for KTL industry for proving the dataset for the analysis.

References

[1] R. Akhavian and A. H. Behzadan, "Simulation-based evaluation of fuel consumption in heavy construction projects by monitoring equipment idle times," Proc. 2013 Winter Simul. Conf. - Simul. Mak. Decis. a Complex World, WSC 2013, no. Lim 2002, pp. 3098–3108 (2013), doi: 10.1109/WSC.2013.6721677.

[2] 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.

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

[4] F. Perrotta, T. Parry, and L. C. Neves, "Application of machine learning 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.

[5] A. Rahman and A. D. Smith, "Predicting fuel consumption for commercial buildings with machine learning algorithms," Energy Build., vol. 152, no. July, pp. 341–358 (2017), doi: 10.1016/j.enbuild.2017.07.017.

[6] A. Schoen, A. Byerly, B. Hendrix, R. M. Bagwe, E. C. Dos Santos, and Z. Ben Miled, "A Machine Learning 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. Bousonville, M. Dirichs, and T. Kruger, "Estimating truck fuel consumption with machine learning using telematics, topology and weather data," Proc. 2019 Int. Conf. Ind. Eng. Syst. Manag. IESM 2019 (2019), doi: 10.1109/IESM45758.2019.8948175.

[8] R. Yan, S. Wang, and Y. Du, "Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship," Transp. Res. Part E Logist. Transp. Rev., vol. 138, no. March, p. 101930 (2020), doi: 10.1016/j.tre.2020.101930.

[9] E. Moradi and L. Miranda-Moreno, "Vehicular fuel consumption estimation using real-world measures through cascaded machine

learning modeling," Transp. Res. Part D Transp. Environ., vol. 88, no. October, p. 102576 (2020), doi: 10.1016/j.trd.2020.102576.

[10] Z. Hu, T. Zhou, M. T. Osman, X. Li, Y. Jin, and R. Zhen, "A novel hybrid fuel consumption prediction model for ocean-going container ships based on sensor data," *J. Mar. Sci. Eng.*, vol. 9, no. 4 (2021), doi: 10.3390/jmse9040449.

[11] T. Anh Tran, "Comparative analysis on the fuel consumption prediction model for bulk carriers from ship launching to current states based on sea trial data and machine learning technique," *J. Ocean Eng. Sci.*, vol. 6, no. 4, pp. 317–339 (2021), doi: 10.1016/j.joes.2021.02.005.

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

[13] 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. November 2020, p. 103827 (2021), doi: 10.1016/j.autcon.2021.103827.

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

[15] Katyare P, Joshi S., "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

[16] S. Kristombu, S. Thilakarathna, and J. Shalitha, "Artificial Intelligence and Smart Vision for Building and Construction 4 . 0 : Machine and Deep Learning Methods and Applications Automation in Construction Artificial intelligence and smart vision for building and construction 4 . 0 : Machine and deep learning methods and applications," Autom. Constr., vol. 141, no. June, p. 104440 (2022), doi: 10.1016/j.autcon.2022.104440.

[17] I. M. I. Ramadan, O. H. Ahmed, and M. S. Abdel-monem, "Modelling Vehicle Emissions and Fuel Consumption Based on Instantaneous Speed and Acceleration Levels," vol. 51, no. 4, pp. 106–116 (2022).

[18] F. Shahnavaz, "Automated Estimation of Construction Equipment Emission Using Inertial Sensors and Machine Learning Models" (2022).

[19] Panda H., Das M.,Sahu B., "Received signal strength prediction model for wireless underground sensor networks using machine learning algorithms," *Journal of Information and Optimization Sciences,* vol. 43, Issue 5, pp. 949-962 (2022).

[20] Liu S., Pan N.,"Integrating the diamond theory and the 3C model: The construction companies in Taiwan," *Journal of Statistics and Management Systems*, vol. 15, issue 6, pp. 601-625 (2012).