Volume 8, Issue 11, pp. 112-115, 2024.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

# Predictive Analysis Models for Optimizing Construction Equipment Operation Based on Historical Data

Ksenia Zagorskaya
Business analyst, Orange, MA
Waltham, Massachusetts (Bentley university)

***Abstract—*** *The purpose of this study is to analyze and compare regression models, decision trees, and neural networks in terms of their use in the construction industry. The methodology includes studying the main predictive analysis models, conducting a theoretical comparative analysis of their strengths and weaknesses, assessing the practical applicability of each model in the conditions of construction companies, identifying possible shortcomings and prospects for using predictive analysis in construction. The results show that regression models are easy to implement and interpret, but are limited in nonlinear dependencies; decision trees are flexible and interpretable, but are subject to overtraining; neural networks are capable of modeling complex nonlinear dependencies and working with large amounts of data, but require computing resources and are difficult to interpret. The findings suggest that the choice of predictive analysis model should be based on the specifics of the construction company, data availability and resources, while predictive analysis has the potential to improve the operating efficiency of construction equipment, reduce maintenance costs and improve productivity*.

***Keywords—*** *Predictive analysis, construction equipment, regression models, decision trees, neural networks*.

## I. INTRODUCTION

The use of machinery is a key factor in the successful implementation of projects in modern construction. However, the high cost of equipment, strict project deadlines, and increasing market competition demand that construction companies optimize their machinery operation processes. In this context, predictive analysis, which relies on historical data, offers opportunities to forecast the technical condition of equipment, reduce downtime, prevent accidents, and optimize maintenance costs.

The aim of this study is to analyze and compare predictive analysis models for optimizing construction equipment performance based on historical data. To achieve this objective, the following tasks need to be addressed:

- Examine the primary predictive analysis models used in the construction industry (regression models, decision trees, and neural networks).
- Conduct a theoretical comparative analysis of the strengths and weaknesses of these models.
- Evaluate the practical applicability of each model in construction company settings.
- Identify potential limitations and future prospects for applying predictive analysis in construction.

The novelty of this research lies in analyzing predictive analysis models in terms of their application to the operation of construction machinery.

## II. LITERATURE REVIEW

Predictive analysis in the construction industry entails applying forecasting methods to assess future events and trends based on historical data. For instance, authors such as Yu.V. Kamaeva and L.A. Adamsevich note in their work that predictive analytics uses historical and current data combined with statistical modeling and machine learning to forecast future outcomes. They emphasize that in construction, this approach enables the prediction of project completion times, risk assessment, and optimization of funds and resources [4].

The primary models used in this context are regression, decision trees, and neural networks. Regression is a statistical approach intended to determine a connection between a dependent variable and one or several independent variables. In terms of construction machinery, regression analysis allows for the prediction of quantitative indicators, such as equipment lifespan or the expected time until the next maintenance. Several types of regression exist [9] (see Fig. 1):

- Linear Regression assumes a linear relationship between variables and is applied when a direct proportionality exists between parameters.
- Multiple Regression extends linear regression to multiple independent variables to account for complex relationships.
- Polynomial Regression is used when the relationship between variables is nonlinear.

The figure shows two graphs: linear regression (illustrating a simple relationship between an independent and a dependent variable) where data points are marked in blue, and the red line represents the linear model fitted to the data, and polynomial regression, which demonstrates a more complex nonlinear relationship. Here, data points are marked in green, and the red curve represents a second-degree polynomial model that better describes the nonlinear relationship between the variables.

Essential conditions for applying regression analysis include accounting for potential issues such as multicollinearity, heteroscedasticity, and incorrect model specification.

The second model is decision trees, which are algorithms that recursively split data into subsets based on feature values to maximize differences among target variables. They are used for both classification and regression tasks. The main steps in

International Journal of Scientific Engineering and Science
Volume 8, Issue 11, pp. 112-115, 2024. ISSN (Online): 2456-7361

constructing a decision tree include, first, determining the feature that best divides the data. For this, criteria such as Gini impurity or information gain are employed. Second, the data is split into subsets, and the process is repeated for each subset until a stopping criterion is met, such as maximum tree depth or minimum number of observations in a node [2].

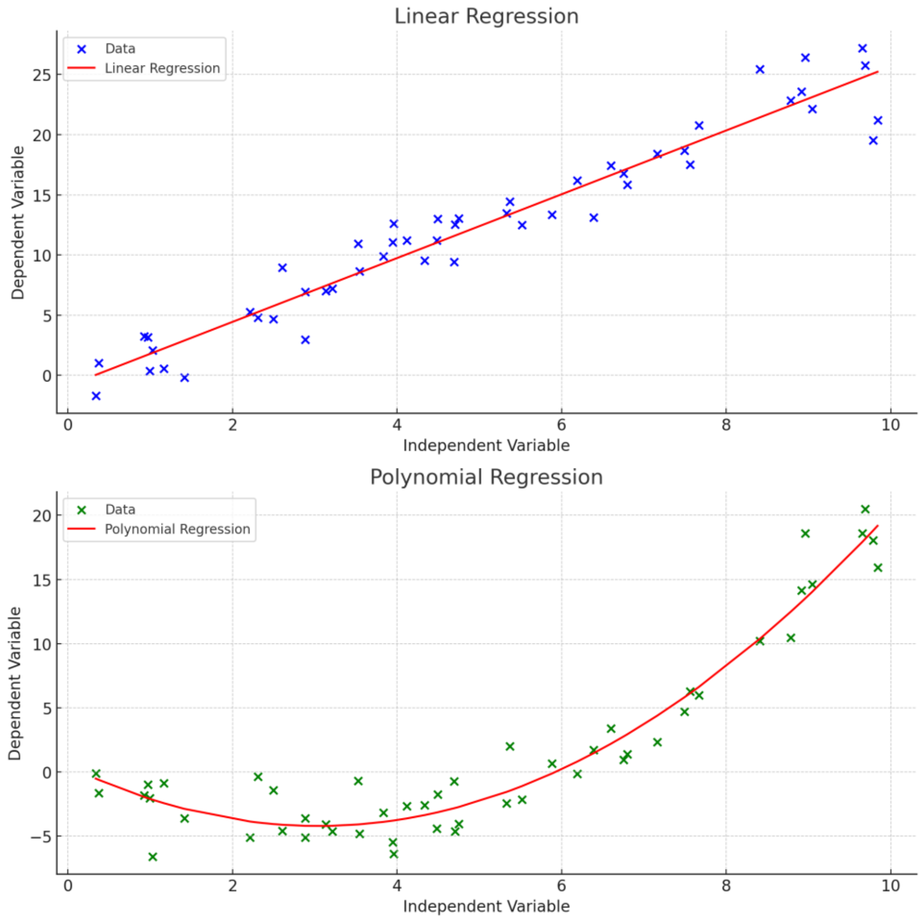


Figure 1. Examples of Linear and Polynomial Regression (source: compiled by the author based on own research)

For example, Figure 2 presents a decision tree for predictive analysis of construction equipment operation, illustrating the decision-making process based on input parameters such as equipment operating hours, maintenance status, ambient temperature, and fuel type. Each node in the tree represents a criterion for splitting the data to determine whether equipment operation will be optimal or not [3]. The leaf nodes display the final classes: optimal or non-optimal operation.

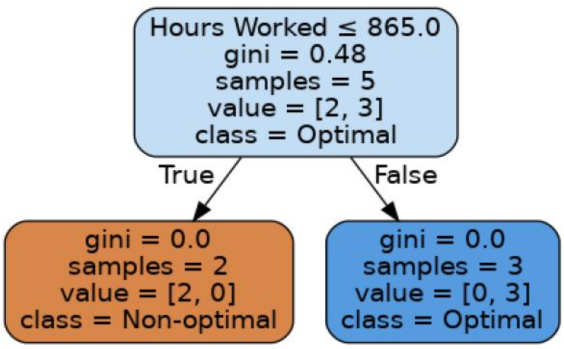


Figure 2. Example of a decision tree for predictive analysis of construction equipment operation (source: compiled by the author based on his own research)

Decision trees offer advantages such as interpretability and the ability to handle both numerical and categorical data; however, they are prone to overfitting, especially with a large tree depth. Pruning methods and ensemble techniques, such as random forests and gradient boosting, are applied to address this issue [1].

Neural networks are algorithmic models motivated by the configuration and operation of biological neurons. They consist of layers of interconnected nodes (neurons) that transform input data into outputs through trainable weights. The main components of a neural network include the input layer, which accepts raw data; hidden layers, which perform nonlinear transformations on the input data; and the output layer, which generates a prediction or classification (see Fig. 3) [6].

Neural networks are capable of modeling complex nonlinear dependencies and handling large volumes of data. In construction equipment operations, they are used for predicting equipment failures, analyzing vibration signals, and other tasks. However, neural networks require significant computational resources and large datasets for training. They are often viewed as "black boxes," which complicates the interpretation of their decisions. To address this, a method using Boolean integral-

International Journal of Scientific Engineering and Science
Volume 8, Issue 11, pp. 112-115, 2024. ISSN (Online): 2456-7361

differential calculus has been proposed to interpret neural network decisions, exploring the logic behind their decision-making and identifying the most influential features in their outputs [7].

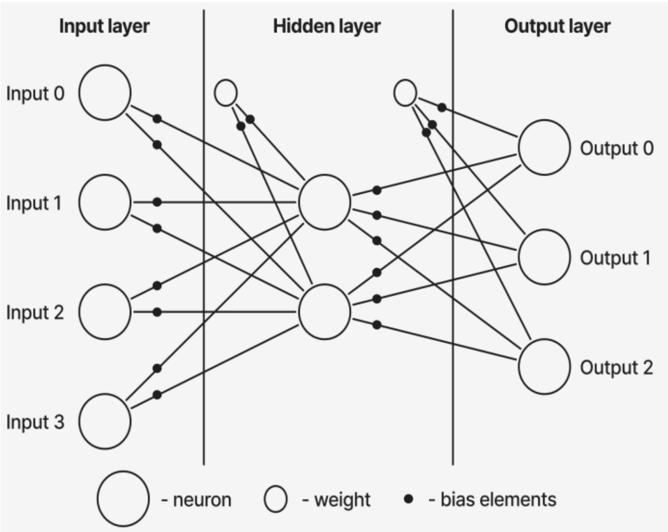


Figure 3. Basic architecture of a neural network (source: compiled by the author based on his own research)

Modern approaches to predictive analytics in the construction industry are based on the integration of big data, machine learning, and artificial intelligence, which process vast amounts of data from various sources, such as sensors on construction sites, drones, and BIM models.

For example, Autodesk has developed the BIM 360 Construction IQ solution, which uses predictive analytics to manage risks and optimize workflows in construction. This system analyzes site data and predicts potential issues, such as delays or safety violations [8].

In Russia, interest in predictive analytics in construction is also on the rise. For example, the company "Pragma" has incorporated a predictive analytics feature into its Pragmacore platform to ensure material and equipment provision for projects, aiming to prevent risks of supply delays.

Predictive analytics applied to forecasting the condition of construction equipment encounters challenges that reduce forecast accuracy. As already noted, reliable predictive models require extensive, high-quality historical data on equipment operation. However, the construction industry often lacks such data due to the absence of systematic data collection and insufficient digitalization of processes. In Russia, specifically, research in the field of predictive analytics for construction remains in its infancy, as this area is still underdeveloped.

On the other hand, predictive analytics models must consider factors such as equipment technical specifications, operating conditions, and external influences. Thus, developing these models requires in-depth knowledge of machine learning and an understanding of the specific features of construction equipment. Implementation of these systems is complicated by the fact that applied predictive analytics relies on expert knowledge of the observed technological object [5].

Furthermore, acquiring a predictive analytics system involves financial costs for purchasing equipment, software, personnel training, and ongoing investments in maintenance and updates, making it economically impractical for some companies.

The complexity of integrating predictive models with existing equipment management and monitoring systems is another issue, as the lack of standardization and compatibility between various systems hampers data exchange and diminishes analytical effectiveness.

Construction equipment operates in diverse conditions (such as different climatic zones, soil types, and usage intensities), and thus, accounting for all these factors in predictive models is a challenge that cannot be ignored.

Many predictive models do not adapt well to changes in operational processes or equipment upgrades, necessitating continuous updating and adjustment of models, which increases labor costs and reduces the speed of decision-making.

Lastly, there is a shortage of specialists in the labor market who have expertise in data analysis, machine learning, and the specifics of construction equipment.

## III. MATERIALS AND METHODS

To develop predictive analysis models for construction equipment, data from various sources are utilized, including technical documentation, operational reports, manufacturer data, and specialized monitoring systems. Data sources are classified into two main categories: primary and secondary. Primary data contain equipment operating parameters (such as hours of operation, average load, and temperature conditions) obtained directly from the equipment via sensors and monitoring systems. Secondary data consist of reports and technical specifications provided by manufacturers or gathered from official operational reports, which include information on equipment types, maintenance recommendations, and historical failure records. The data undergo normalization, standardization, and filtering to remove outliers and anomalies.

Model quality is evaluated using MSE, MAE, and the coefficient of determination ($R^2$) to accurately measure deviations and the explained variance.

## IV. DISCUSSION

In this study, three primary predictive analysis models used in the construction industry were analyzed: regression models, decision trees, and neural networks. Based on the comparative analysis, the following conclusions were reached:

Regression models are characterized by their relative simplicity in implementation and result interpretation. Linear and multiple regression are effective when there are linear relationships between variables and allow for predicting quantitative indicators, such as equipment lifespan or time until the next maintenance. However, their limitations arise with nonlinear data, where polynomial regression complicates the model and can lead to overfitting. Additionally, regression models are sensitive to multicollinearity and heteroscedasticity, requiring careful data preprocessing.

Decision trees offer a more flexible approach to analysis, capable of handling both numerical and categorical data. Their

International Journal of Scientific Engineering and Science
Volume 8, Issue 11, pp. 112-115, 2024. ISSN (Online): 2456-7361

visual interpretability facilitates understanding and decision-making based on the model. However, a significant disadvantage is their tendency to overfit, especially with deeper trees. This necessitates the use of pruning methods or ensemble methods, which complicate the model and increase computational costs.

Neural networks have a high capacity for modeling complex nonlinear relationships and are effective when processing large volumes of data, taking into account multiple factors simultaneously. This makes them beneficial for predicting equipment failures and analyzing complex processes. However, their primary drawback is the “black box” nature—the difficulty of interpreting internal decision-making mechanisms. Furthermore, neural networks require substantial computational resources and large datasets for training.

The practical applicability of these models for construction companies depends on their resources and needs. Regression models can be implemented relatively easily and quickly, providing useful forecasts with minimal costs. Decision trees are suitable for companies needing in-depth insights and willing to invest in skilled personnel and technology. Neural networks are more appropriate for large organizations with access to big data and high computational capacity, where the accuracy of forecasts outweighs the costs of implementation.

The potential for improving construction equipment efficiency through predictive analysis is significant, as the use of these models allows for the optimization of maintenance schedules, prediction of potential failures, and reduction of equipment downtime. This leads to reduced repair costs, extended equipment lifespan, and overall productivity improvement for the company.

## V. Conclusion

This study reviewed and analyzed three main predictive analysis models—regression models, decision trees, and neural networks—applied to the operation of construction equipment. The comparative analysis demonstrated that each model has its own advantages and disadvantages.

The effectiveness and applicability of the models discussed depend on the specific characteristics of the construction company, data availability, and resources. However, it is clear that predictive analysis holds the potential to improve equipment operation efficiency, reducing maintenance costs and increasing productivity.

Promising directions for further research include:

- Practical testing of predictive analysis models on empirical data from real construction projects.
- Development of hybrid models that combine the strengths of different approaches to enhance the accuracy and robustness of predictions.
- Improvement of data collection and processing methods, including the use of monitoring systems and equipment sensors, to create high-quality historical data.
- Investigation of methods for integrating predictive models into existing management and decision-making systems within construction companies.

## References


1. Antonenkov D. V., Matrenin P. V. Study of ensemble and neural network methods of machine learning in the problem of short-term forecasting of electricity consumption of mining enterprises // ES i K. 2021. No. 3 (52). URL: https://cyberleninka.ru/article/n/issledovanie-ansamblevyh-i-neyrosetevyh-metodov-mashinnogo-obucheniya-v-zadache-kratkosrochnogo-prognozirovaniya (date of access: 06.11.2024).
2. Ashikhmin O. V., Shestakova A. P. Digitalization of processes of making technological and design decisions in modern construction // Architecture, construction, transport. 2022. No. 2. URL: https://cyberleninka.ru/article/n/tsifrovizatsiya-protsessov-prinyatiya-tehnologicheskih-i-proektnyh-resheniy-v-sovremennom-stroitelstve (date of access: 03.11.2024).
3. Dormidontova T. V., Gareeva L. Kh., Solkaryan N. G. Application of the "Decision Tree" method and a planned experiment to select the best options for given criteria in transport construction // Bulletin of Eurasian Science. 2015. No. 2 (27). URL: https://cyberleninka.ru/article/n/primenenie-metoda-dereva-resheniy-i-planirovannogo-eksperimenta-dlya-vybora-luchshih-variantov-pri-zadannyh-kriteriyah-v-transportnom (date of access: 06.11.2024).
4. Kamaeva Yu. V., Adamtsevich L. A. Prospects for using predictive analytics in construction // Construction and architecture. 2023. No. 2. pp. 12–12. DOI: https://doi.org/10.29039/2308-0191-2023-11-2-12-12 (date of access: 03.11.2024).
5. Kolmykov I. A. Predictive analytics. Practical reflections // IndaSoft. 2021. URL: https://indusoft.ru/media/articles/1401/ (date of access: 03.11.2024).
6. Krevsky M. I., Bozhday A. S. Predictive neural network model for business process management // News of universities. Volga region. Technical sciences. 2023. No. 3 (67). URL: https://cyberleninka.ru/article/n/prediktivnaya-neyrosetevaya-model-dlya-upravleniya-biznes-protsessami (date of access: 05.11.2024).
7. Lyutikova L. A. Using logical methods to analyze neural network decisions // Modeling, optimization and information technology. 2023. Vol. 11. No. 4. Pp. 1–12.
8. Miroshkin N. V. Technical analysis of existing solutions within the framework of creating an IT system for designing landscaping and construction works // Bulletin of Science. 2024. No. 1 (70). URL: https://cyberleninka.ru/article/n/tehnicheskiy-analiz-suschestvuyuschih-resheniy-v-ramkah-sozdaniya-it-sistemy-proektirovaniya-rabot-po-blagoustroystvu-i (date of access: 05.11.2024).
9. Breiman L., Friedman J., Olshen R.A., Stone C.J. Classification and Regression Trees. 1st ed. Chapman and Hall/CRC, 1984. DOI: https://doi.org/10.1201/9781315139470.