

# Proposal for a Machine Learning Model to Improve Hauling Productivity in an Open-Pit Mine

Ernan Capcha Milla, Jimmy Rosales Huamani

Universidad Nacional de Ingenieria, Lima, Peru  
Email: [ecapcham@uni.edu.pe](mailto:ecapcham@uni.edu.pe), [jrosales@uni.edu.pe](mailto:jrosales@uni.edu.pe)

**How to cite this paper:** Milla, E.C. and Huamani, J.R. (2026) Proposal for a Machine Learning Model to Improve Hauling Productivity in an Open-Pit Mine. *Engineering*, 18, 99-106.  
<https://doi.org/10.4236/eng.2026.183007>

**Received:** December 9, 2025

**Accepted:** March 17, 2026

**Published:** March 20, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This proposal outlines a machine learning-based approach aimed at improving productivity in haulage operations within open-pit mining. Since hauling accounts for up to 60% of total operational costs, predictive models that enable early intervention and optimization are of strategic importance. The proposed methodology involves the use of Gaussian Mixture Models (GMM) for data preprocessing and Random Forest algorithms for predictive modeling, complemented by ensemble techniques such as Gradient Boosting and XGBoost. The model is expected to be trained and evaluated using historical and real-time operational data, including variables such as loading time, truck availability, material type, and travel distance. Evaluation metrics such as MAE, RMSE, and  $R^2$  will be used to assess predictive performance. The aim is to build a framework that enables early warnings of productivity deviations and supports real-time decision-making. This research seeks to contribute to Mining 4.0 through the development of an interpretable and scalable tool for haulage optimization in real operational settings.

## Keywords

Machine Learning, Haulage Productivity, Open-Pit Mining, Random Forest

## 1. Introduction

The open-pit mining industry faces significant challenges in optimizing loading and hauling operations, critical components that represent between 50% and 60% of the total mineral extraction cost. In this context, accurately predicting the productivity of these systems has become a pressing need to improve operational efficiency and reduce costs in global mining operations.

At the international level, the problem of optimizing loading and hauling systems has been widely documented. [1] provided a comprehensive perspective on the solution strategies used in truck dispatch systems for surface mines, identifying the inherent complexities in the efficient management of mining fleets. This problem is compounded by the stochastic nature of mining processes, where variables such as ground conditions, weather, and the characteristics of the extracted material introduce significant variability into cycle times. [2] conducted a comparative study of truck cycle time prediction methods in surface mining, highlighting the need for more accurate approaches to productivity estimation in these operations.

The implementation of machine learning technologies has emerged as a promising solution to address these challenges. [3] evaluated predictive machine learning models for mineral prospectivity, comparing neural networks, random forests, regression trees, and support vector machines, demonstrating the superiority of these approaches over traditional methods. In the specific context of mining optimization, [4] applied random forest models using multi-year blast data from an open-pit mine, demonstrating the ability of these algorithms to predict complex outcomes in mining operations. Similarly, [5] employed machine learning methods to predict real-time truck travel times in open-pit mines, demonstrating the viability of these technologies for operational optimization.

In the North American context, particularly in the oil sands operations of Alberta, Canada, significant research has been conducted on mining systems optimization. [6] discussed the development of the Alberta oil sands, highlighting the importance of a strong energy sector to ensure Canada's future prosperity. This development has driven the need to optimize large-scale mining operations, where load-haul systems play a critical role. [7] improved production planning in oil sands mining by analyzing and simulating truck cycle times, identifying specific optimization opportunities in this type of operation.

The application of advanced machine learning techniques in the mining context has shown promising results in various applications. [8] used GBDT, XGBoost, and LightGBM algorithms to predict pillar stability in hard rock, achieving accuracy levels superior to traditional methods. Similarly, [9] used gradient boosting regression trees to estimate mineral grades, demonstrating the effectiveness of these approaches in complex regression problems in mining. [10] developed ensemble learning models to assess the strength of coal-slurry materials, demonstrating the ability of these techniques to handle the inherent complexity of mining data.

A critical aspect of implementing machine learning in mining operations is the handling of large volumes of data with complex characteristics. [11] addressed the preprocessing of large datasets using Gaussian mixture modeling to improve the prediction accuracy of truck productivity at mining sites, demonstrating the importance of advanced data preparation techniques. This approach is particularly relevant considering that mining operations continuously generate large volumes of operational data that require efficient processing for use in predictive models.

The optimization of shovel-truck systems has been the subject of extensive research, recognizing its direct impact on mining productivity. [12] performed an optimization of shovel-truck systems for surface mining, identifying critical parameters that affect operational efficiency. [13] compared truck-and-shovel systems versus mine haulage systems, evaluating the valuable operating time of each configuration. These studies underscore the importance of systematic approaches to optimizing loading and hauling operations.

The inherent variability in mining operations, particularly in truck cycle times, has been identified as a critical factor affecting overall productivity. [14] developed a discrete event model to simulate the effect of truck clustering due to payload variance on cycle time, transported materials, and fuel consumption. This research highlighted the need to consider multiple interrelated variables in the optimization of mining transportation systems.

In the specific context of productivity prediction, several studies have explored the application of machine learning algorithms to improve the accuracy of estimates. [15] compared backpropagation neural networks, gradient boosting trees, and multiple linear regression to predict the calorific value of lignite deposits, demonstrating the superiority of machine learning approaches. [16] applied gradient boosting methods to improve travel time prediction, achieving significant improvements in predictive accuracy.

The implementation of advanced clustering techniques has proven particularly effective in the analysis of complex operational patterns. [17] used Gaussian mixture models and hierarchical clustering to identify typical profiles of daily electricity usage in buildings, a methodology that can be adapted for the analysis of productivity patterns in mining operations. [18] applied k-means and Gaussian mixture models to classify seismic activity, demonstrating the versatility of these techniques in geological and mining applications.

Despite significant advances in the application of machine learning in mining operations, significant challenges remain in accurately predicting productivity in haulage and loading systems. The multivariate and dynamic nature of these systems, influenced by operational, environmental, and equipment factors, requires robust methodological approaches that can capture the inherent complexity of these operations. The integration of advanced machine learning techniques with real-time operational data represents a significant opportunity to improve efficiency and productivity in open-pit mining operations, justifying the need for dedicated research in this field.

In this context, this research contributes to the development of an integrated methodological framework that combines advanced machine learning techniques with real-time operational data analysis for the accurate prediction of productivity in loading and hauling systems in open-pit mining. The main contribution lies in the development of a hybrid methodology that uses Gaussian mixture models for the preprocessing of large volumes of operational data, integrated with ensemble learning algorithms (random forest, gradient boosting and XGBoost) to generate

robust productivity predictions.

This methodological approach captures the stochastic and multivariate nature of loading and hauling processes, considering operational, environmental and equipment variables that have traditionally limited the accuracy of conventional predictive methods. The practical impact of this research is realized in the development of real-time prediction tools that can be integrated into existing mining management systems, providing operational optimization capabilities that result in significant cost reductions, improved equipment utilization efficiency, and the establishment of early warning systems for the proactive identification of deviations from expected productivity. From a theoretical perspective, this research advances knowledge in the application of artificial intelligence to mining operations by establishing a systematic protocol for the evaluation and implementation of predictive models in the specific context of loading and hauling operations, which constitutes a fundamental contribution to the development of more efficient and sustainable mining operations.

## 2. State of Art

Hauling operations in open-pit mining represent one of the most critical and costly stages of the mining cycle, accounting for between 50% and 60% of total operating costs. Due to their highly dynamic and multivariate nature, numerous studies have sought to improve their efficiency through prediction, simulation, and optimization models. [1] presented a comprehensive overview of truck dispatch strategies, while [2] compared traditional truck cycle time prediction methods, highlighting their limitations in the face of operational variability.

The recent development of machine learning (ML) techniques has opened up new possibilities for addressing this problem. Researchers such as [3] [5] demonstrated the superiority of models such as random forests, gradient boosting, and support vector machines over traditional methods in mining contexts, especially in classification and regression tasks with large volumes of operational data.

However, despite these advances, significant challenges remain related to operational data preprocessing, the selection of relevant variables, and the model's real-time adaptability. In response to these limitations, this research proposes a hybrid methodology based on Random Forest algorithms, complemented by Gaussian mixture techniques for data preprocessing, aimed at predicting productivity in loading and hauling systems. This approach allows for efficient modeling of the stochastic complexity of the process, integrating operational, environmental, and equipment factors.

Studies such as those by [11] have already shown that preprocessing with Gaussian mixtures improves the accuracy of RF models by reducing noise and better structuring data distributions. Based on this principle, the current proposal integrates Random Forest models not only as a prediction tool, but also as the core of an intelligent architecture that will be fed by historical, operational, and real-time sensor data, thus optimizing the performance of the load-haul cycle.

Likewise, previous work in mining, such as that by [4] [9], validates the efficiency of ensemble learning models in complex mining scenarios. However, most of these applications have focused on geological or blasting variables, leaving a broad field of application in the operational control of mining fleets, where this study stands as an innovative and practical contribution.

The integration of this proposal into the state-of-the-art responds to the growing need for robust predictive systems that can adapt to dynamic conditions, generate early warnings, and support real-time decision-making. The aim is not only to advance the academic literature but also to offer a viable technological solution for modern mining, aligned with the principles of efficiency, sustainability, and digital transformation promoted by Mining 4.0.

### 3. Proposal

The contribution of this research is the proposal of a machine learning model to improve productivity in haulage operations within the mining cycle of an open-pit mine. This proposal seeks to establish a robust methodological architecture that allows for early prediction of mining truck productivity levels and appropriate action to optimize operational efficiency. The contribution aims at using machine learning technologies, integrated with real operational data, with the goal of reducing downtime and operational costs in transport.

The proposed methodology considers the following steps:

- 1) Collection of historical and real-time operational data.
- 2) Data preprocessing and structuring.
- 3) Selection of relevant variables.
- 4) Definition of the prediction model.
- 5) Model training.
- 6) Evaluation with statistical metrics.
- 7) Proposal for application in real environments.

The conceptual architecture diagram of the Machine Learning model for hauling is presented in **Figure 1** below, structured in three modules:

- **Module 1. Data Collection and Analysis**

This module will detail the operational data collection process. The primary source of information will be the mining operation's dispatch system, which records key variables of the hauling cycle, such as: Loading time, Waiting time, Travel time loaded and empty, Truck availability, Shift and time, Type of material, Distance traveled. The dataset will be cleaned, structured, and subjected to exploratory analysis techniques. A Gaussian mixture modeling (GMM) technique will then be applied to segment operational patterns and reduce the influence of outliers, thereby improving the quality of the dataset used for the prediction model.

- **Module 2. Predictive Modeling**

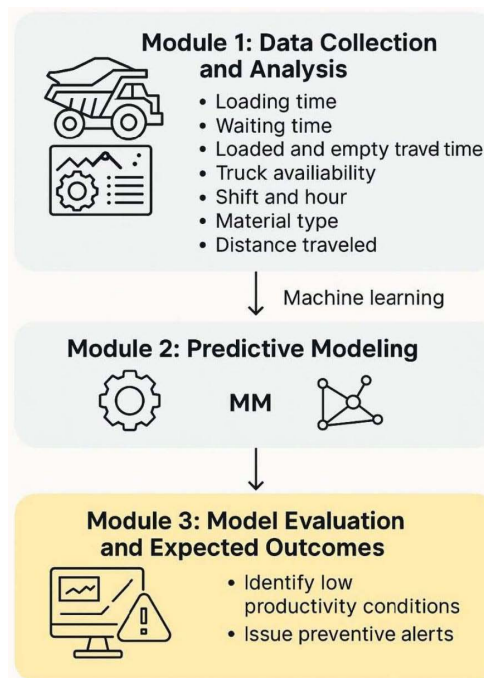
With the data prepared, the predictive model will be trained. The Random Forest Regressor algorithm is proposed due to its proven effectiveness in mining contexts with high variability. Furthermore, techniques such as gradient boosting and

XGBoost are considered viable alternatives for future comparisons. Training will include cross-validation and hyperparameter tuning using techniques such as RandomizedSearchCV.

The model will be designed to predict haulage productivity (expressed in tons per hour) from multiple independent variables mentioned in the previous module.

#### • **Module 3. Model Evaluation and Expected Results**

At this stage, the performance of the proposed model will be evaluated using metrics such as: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and  $R^2$  (Coefficient of determination). The model is expected to identify operating conditions that lead to low productivity and issue preemptive alerts to improve equipment allocation, route planning, and load distribution. In the future, once the model has been validated, it is planned to integrate it into the mine's operational control system, which will allow the issuance of performance reports by shift and haulage unit, with automatic recommendations for improving performance.



**Figure 1.** Diagram of the proposed conceptual architecture.

## 4. Expected Contributions

The main contribution of our work is the proposal of a machine learning model aimed at predicting and improving productivity in haulage operations at an open-pit mine. This model seeks to anticipate scenarios of poor operational performance by analyzing historical and real-time data, enabling more efficient and proactive decision-making by the operations team.

Unlike previous studies focused on blasting or mineral processing phases, our proposal focuses specifically on hauling, a critical and highly costly stage of the

mining cycle, with a significant impact on overall efficiency.

Currently, in the national mining sector, there is no integrated predictive tool that uses machine learning algorithms to optimize haulage. Therefore, this research proposes laying the groundwork for the development of a prototype system capable of generating operational alerts, identifying patterns of poor performance, and suggesting adjustments to fleet distribution and daily planning.

This proposal seeks to be implemented as a future pilot test in real-world operations, with the potential to be integrated into mining management systems and contribute to the advancement of digital mining, aligned with the principles of Mining 4.0.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Alarie, S. and Gamache, M. (2002) Overview of Solution Strategies Used in Truck Dispatching Systems for Open Pit Mines. *International Journal of Surface Mining, Reclamation and Environment*, **16**, 59-76. <https://doi.org/10.1076/ijsm.16.1.59.3408>
- [2] Chanda, E.K. and Gardiner, S. (2010) A Comparative Study of Truck Cycle Time Prediction Methods in Open-Pit Mining. *Engineering, Construction and Architectural Management*, **17**, 446-460. <https://doi.org/10.1108/09699981011074556>
- [3] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M. (2015) Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines. *Ore Geology Reviews*, **71**, 804-818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- [4] Ohadi, B., Sun, X., Esmaili, K. and Consens, M.P. (2020) Predicting Blast-Induced Outcomes Using Random Forest Models of Multi-Year Blasting Data from an Open Pit Mine. *Bulletin of Engineering Geology and the Environment*, **79**, 329-343. <https://doi.org/10.1007/s10064-019-01566-3>
- [5] Sun, X., Zhang, H., Tian, F. and Yang, L. (2018) The Use of a Machine Learning Method to Predict the Real-Time Link Travel Time of Open-Pit Trucks. *Mathematical Problems in Engineering*, **2018**, Article ID: 4368045. <https://doi.org/10.1155/2018/4368045>
- [6] Giesy, J.P., Anderson, J.C. and Wiseman, S.B. (2010) Alberta Oil Sands Development. *Proceedings of the National Academy of Sciences*, **107**, 951-952. <https://doi.org/10.1073/pnas.0912880107>
- [7] Cervantes, E.G., Upadhyay, S. and Askari-Nasab, H. (2018) Improvements to Production Planning in Oil Sands Mining through Analysis and Simulation of Truck Cycle Times. *Mining Optimization Laboratory*, Vol. 1, 142.
- [8] Liang, W., Luo, S., Zhao, G. and Wu, H. (2020) Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics*, **8**, Article 765. <https://doi.org/10.3390/math8050765>
- [9] Kaplan, U.E., Dagsan, Y. and Topal, E. (2021) Mineral Grade Estimation Using Gradient Boosting Regression Trees. *International Journal of Mining, Reclamation and Environment*, **35**, 728-742. <https://doi.org/10.1080/17480930.2021.1949863>
- [10] Sun, Y., Li, G., Zhang, N., Chang, Q., Xu, J. and Zhang, J. (2021) Development of

- Ensemble Learning Models to Evaluate the Strength of Coal-Grout Materials. *International Journal of Mining Science and Technology*, **31**, 153-162.  
<https://doi.org/10.1016/j.ijmst.2020.09.002>
- [11] Fan, C., Zhang, N., Jiang, B. and Liu, W.V. (2022) Preprocessing Large Datasets Using Gaussian Mixture Modeling to Improve Prediction Accuracy of Truck Productivity at Mine Sites. *Archives of Mining Sciences*, **67**, 661-680.
- [12] Ercelebi, S.G. and Bascetin, A. (2009) Optimization of Shovel-Truck System for Surface Mining. *Journal of the Southern African Institute of Mining and Metallurgy*, **109**, 433-439.
- [13] Dzakpata, I., Knights, P., Kizil, M.S., Nehring, M. and Aminossadati, S.M. (2016) Truck and Shovel Versus In-Pit Conveyor Systems: A Comparison of the Valuable Operating Time. In: Aziz, N. and Kininmonth, B., Eds., *Proceedings of the 16th Coal Operators' Conference*, University of Wollongong, 463-476.  
<https://hdl.handle.net/10779/uow.27685101.v1>
- [14] Soofastaei, A., Aminossadati, S.M., Kizil, M.S. and Knights, P. (2016) A Discrete-Event Model to Simulate the Effect of Truck Bunching Due to Payload Variance on Cycle Time, Hauled Mine Materials and Fuel Consumption. *International Journal of Mining Science and Technology*, **26**, 745-752.  
<https://doi.org/10.1016/j.ijmst.2016.05.047>
- [15] Ahmed, W., Muhammad, K. and Siddiqui, F.I. (2020) Predicting Calorific Value of Thar Lignite Deposit: A Comparison between Back-Propagation Neural Networks (BPNN), Gradient Boosting Trees (GBT), and Multiple Linear Regression (MLR). *Applied Artificial Intelligence*, **34**, 1124-1136.  
<https://doi.org/10.1080/08839514.2020.1824091>
- [16] Zhang, Y. and Haghani, A. (2015) A Gradient Boosting Method to Improve Travel Time Prediction. *Transportation Research Part C: Emerging Technologies*, **58**, 308-324. <https://doi.org/10.1016/j.trc.2015.02.019>
- [17] Li, K., Ma, Z., Robinson, D. and Ma, J. (2018) Identification of Typical Building Daily Electricity Usage Profiles Using Gaussian Mixture Model-Based Clustering and Hierarchical Clustering. *Applied Energy*, **231**, 331-342.  
<https://doi.org/10.1016/j.apenergy.2018.09.050>
- [18] Kuyuk, H.S., Yildirim, E., Dogan, E. and Horasan, G. (2012) Application of K-Means and Gaussian Mixture Model for Classification of Seismic Activities in Istanbul. *Non-linear Processes in Geophysics*, **19**, 411-419.  
<https://doi.org/10.5194/npg-19-411-2012>