

PREDICTION OF LOGISTICS PERFORMANCE AND ALLOCATION OPTIMIZATION IN CONSTRUCTION PROJECTS USING CATBOOST, LIGHTGBM, AND XGBOOST BASED ON LEAN CONSTRUCTION PRINCIPLES

Arif H. N. Shidqi¹, Mochamad A. Wibowo², Adian Fatchur³, & Alfain Ramdhan⁴

ABSTRACT

Supply Chain Management performance and strategic optimization allocation are essential determinants of construction project productivity and operational efficiency. Delays in material distribution, misallocation of workforce, and underused equipment often result in workflow interruptions and waste, which contradict the Lean Construction ideals of continuous flow and waste minimization. While the use of machine learning in construction management has expanded in recent years, research focusing on predictive frameworks for logistics and resource utilization grounded in lean methodology remains scarce. This study introduces a machine learning framework designed to forecast logistics outcomes and resource allocation efficiency in construction projects through the application of CatBoost, LightGBM, and XGBoost algorithms. The dataset comprises variables such as material delivery attributes, workforce capacity, task complexity, and project operating conditions, whereas logistics performance metrics and resource utilization indicators form the prediction targets. Model evaluation employs suitable regression and classification criteria. The experimental results reveal that boosting based techniques can effectively model intricate relationships within construction logistics data. Among the algorithms tested, CatBoost performs most effectively, particularly in managing categorical attributes. The proposed framework enables data driven, proactive decisions to optimize resource deployment and support lean oriented project execution by minimizing idle time, delays, and inefficiencies.

KEYWORDS

Construction logistic, lean construction, machine learning.

¹ Doctoral Candidate, Department of Civil Engineering, Universitas Diponegoro, Semarang, Indonesia, haidarsanath@gmail.com, orcid.org/0009-0005-1957-2820

² Professor, Department of Civil Engineering, Universitas Diponegoro, Semarang, Indonesia, agung.wibowo@ft.undip.ac.id, orcid.org/0000-0002-5434-9107

³ Professor, Department of Computer Engineering, Universitas Diponegoro, Semarang, Indonesia, adian@ce.undip.ac.id, orcid.org/0000-0002-1921-9358

⁴ Bachelor's, Department of Civil Engineering, Universitas Diponegoro, Semarang, Indonesia, alfainnaga@gmail.com, orcid.org/0009-0003-7544-9507

INTRODUCTION

Logistics performance and resource allocation are critical determinants of productivity, cost efficiency, and schedule reliability in construction projects. Ineffective coordination of material delivery, manpower deployment, and equipment utilization often results in non-value-adding activities, such as excessive waiting time and idle resources. These inefficiencies directly contradict Lean Construction principles, which prioritize flow reliability and waste reduction to maximize value. Despite its importance, decision-making in construction logistics remains largely experience-based and reactive, limiting the ability of project teams to stabilize production flows before disruptions occur.

Recent advances in machine learning (ML) have demonstrated strong potential for predictive analysis in complex environments. Ensemble-based algorithms, particularly gradient boosting methods, have been shown to outperform traditional statistical models in capturing nonlinear relationships and handling the heterogeneous datasets typical of construction projects (Hancock & Khoshgoftaar, 2020; Boldini et al., 2023). Specifically, CatBoost offers a distinct advantage in this domain due to its superior handling of categorical features without the need for extensive preprocessing (Hancock & Khoshgoftaar, 2020), while LightGBM and XGBoost provide high computational efficiency and predictive accuracy in multi-variable estimation tasks (Mustapha et al., 2024; Ileri, 2025).

However, a significant research gap persists: Lean Construction practices often rely on retrospective metrics, such as the Percent Plan Complete (PPC), which identify failures only after they have occurred. There is currently a lack of integration between advanced predictive analytics and Lean theory to create a "proactive shielding" mechanism for logistics. Consequently, this study seeks to investigate the extent to which gradient boosting algorithms can accurately forecast logistics performance and resource bottlenecks within a Lean-oriented construction environment. Furthermore, this research aims to clarify how such a predictive ML framework can advance Lean theory by facilitating a transition from reactive troubleshooting to proactive flow stabilization.

By integrating Lean Construction principles with CatBoost, LightGBM, and XGBoost, this study introduces a framework to evaluate and identify the most effective predictive approach for data-driven decision-making. Consequently, this study seeks to investigate the extent to which gradient boosting algorithms can accurately forecast logistics performance through the optimization of vendor selection decisions within a Lean-oriented construction environment. This research contributes to IGLC knowledge by providing a robust mechanism for the "Make Ready" process, demonstrating that logistics reliability is a downstream effect of AI-informed upstream interventions. Ultimately, this approach strengthens the reliability of the production shield, effectively reducing variability and ensuring a continuous, lean workflow in logistics-intensive construction operations.

LITERATURE REVIEW

LEAN CONSTRUCTION AND CONSTRUCTION LOGISTICS

The implementation of Lean Construction principles focuses on minimizing waste and reducing variability within construction production processes to enhance overall efficiency. Through lean project management, production systems are designed to be more robust, enabling them to function effectively even under uncertain conditions or disruptions, which ultimately ensures a continuous and reliable workflow (Ballard and Howell, 2003).

In the context of construction logistics, performance is often hindered by problems that are detected on site but actually originate in earlier stages of the supply chain. This highlights that integrating site level logistics with external supply chain management is crucial to prevent

"myopic" or short term control practices that can degrade project performance and increase unnecessary costs (Vrijhoef and Koskela, 2000).

SCM IN CONSTRUCTION PROJECTS

Supply Chain Management (SCM) in the construction industry possesses four specific roles, ranging from focusing on the interface between the site and the supply chain to managing the entire chain as a single integrated entity. This approach is necessary because traditional construction supply chains are frequently plagued by inefficiencies caused by a lack of coordination between different departments and stakeholders (Vrijhoef and Koskela, 2000).

Furthermore, the integration of Big Data concepts into SCM significantly contributes to economic sustainability, particularly during the building design and project planning phases. By systematically analyzing large datasets, companies can make more informed decisions regarding material procurement and cost efficiency, which are vital for long term financial success and resource optimization (Mandičák et al., 2021).

OPTIMIZATION ALLOCATION IN CONSTRUCTION SUPPLY CHAIN

Optimizing resource allocation within resource constrained construction supply chains necessitates close alignment between vendor selection and project scheduling. A lack of synchronization between these processes can lead to material delivery delays or inappropriate supplier choices, which may subsequently disrupt project schedules and increase costs while creating operational bottlenecks (Chen et al., 2018).

Integrating supplier selection decisions with project activity scheduling within optimization models is fundamental to achieving optimal resource utilization. Such a coordinated approach enables management to adjust resource allocation dynamically according to varying site conditions, while ensuring the robustness and reliability of the construction supply chain (Chen et al., 2018).

INDUSTRY 4.0 ARTIFICIAL INTELLIGENCE AND SUPPLY CHAIN PERFORMANCE

The Industry 4.0 revolution has introduced profound changes through the use of technologies such as the Internet of Things (IoT) and advanced data analytics, which enhance operational performance and real time collaboration. These technologies allow for a more transparent and rapid flow of information among all stakeholders, effectively mitigating risks associated with supply chain disruptions (Fatorachian and Kazemi, 2021).

Artificial Intelligence (AI) further augments these capabilities through a taxonomy of applications involving sensing, learning, and decision making. AI enables systems to learn from historical data and generate automated predictions regarding logistics needs, allowing organizations to shift from reactive to proactive operational models in managing supply chain performance (Pournader et al., 2021).

CONCEPTUAL INTEGRATION OF LEAN PRINCIPLES AND MACHINE LEARNING

The core objective of Lean Construction is the reduction of variability to ensure a reliable and continuous production flow. In the context of logistics, variability often manifests as unpredictable material arrival times and resource bottlenecks, which are traditionally managed through experience-based reactive measures. However, the integration of advanced data analytics through ensemble-based learning provides a theoretical shift from reactive to proactive management. The utilization of gradient boosting frameworks aligns with the Lean objective of "shielding" production by identifying complex, non-linear patterns in heterogeneous supply chain data such as the interplay between material types, vendor performance, and site constraints that conventional models often fail to capture (Hancock & Khoshgoftaar, 2020; Boldini et al., 2023).

In a Lean production system, the "Make Ready" process is essential for ensuring that all prerequisites are met before an activity begins. Machine learning models function as a digital extension of this process; by treating logistics variables as system constraints, these algorithms can forecast potential flow disruptions with high precision. For instance, the ability of certain ensemble techniques to handle categorical data natively allows for a more robust representation of qualitative site conditions, directly supporting the flow reliability principle without losing critical information during data preprocessing (Hancock & Khoshgoftaar, 2020).

Furthermore, the choice of evaluation metrics specifically Accuracy, Precision, Recall, and F1-Score is fundamental to ensuring value generation and minimizing waste. High Precision ensures that resources are not mobilized for false alarms, while high Recall is critical for identifying all potential logistics bottlenecks that could disrupt production flow. By optimizing these metrics, the framework minimizes the "waste of waiting" and improves the reliability of the production shield. This integration effectively transforms raw logistics data into actionable insights, moving the industry toward a state of Digital Lean where waste is minimized through proactive variability control.

While existing literature acknowledges the individual benefits of Lean and AI, a significant gap remains in the predictive synchronization of upstream vendor selection with downstream site logistics. This study fills that gap by operationalizing a proactive framework that treats supply chain variables as predictive constraints to stabilize production flow.

METHODOLOGY

This study follows a structured workflow starting with the collection of historical project data from completed construction projects, as shown in figure 1. These data include records related to vendor performance, such as delivery reliability, responsiveness, capacity, and previous contractual outcomes. Based on this information, a systematic vendor performance evaluation is conducted to quantify each vendor's historical behavior and operational capability. The evaluated vendor data are then used as input for machine learning models, namely CatBoost, LightGBM, and XGBoost, which are trained to identify patterns associated with successful vendor performance. The output of the models is the selection of the best performing vendors. Finally, the impacts of vendor selection decisions are analyzed in terms of logistics performance and resource allocation, focusing on material delivery reliability, equipment utilization, and labor efficiency. This approach ensures that logistics and resource related outcomes are examined as downstream effects of AI based vendor selection rather than as direct optimization targets.

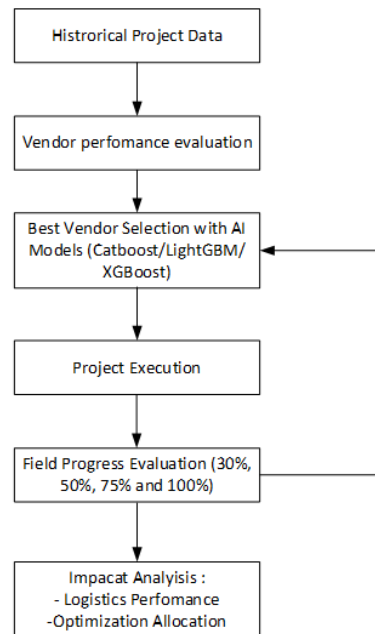


Figure 1: Flow chart methodology

DATA COLLECTION

The data collection phase is based on the extraction of historical project records from a centralized institutional repository, covering an extensive collective period of the last six years. This longitudinal dataset provides a stable and comprehensive foundation for analysis, capturing diverse project archetypes, varying environmental conditions, and multi-year economic cycles that influence logistics performance (Mandičák et al., 2021). By utilizing a centralized source, the study ensures data consistency, allowing for the integration of high-fidelity administrative information, such as procurement logs, manual site inventory records, and labor productivity reports from numerous completed and ongoing projects.

Following the extraction from the central repository, the data undergoes a rigorous filtration and cleaning process to ensure its suitability for predictive modeling. Given the six year span, the preprocessing stage includes time series alignment and the standardization of legacy data formats into a unified structure suitable for Big Data analytics (Mandičák et al., 2021). Feature engineering is then applied to isolate key performance indicators related to resource utilization and supply chain bottlenecks, transforming raw historical records into a structured dataset that reflects the interdisciplinary nature of modern construction management (Hancock and Khoshgoftaar, 2020).

LEAN CONSTRUCTION PRINCIPLES IMPLEMENTATION

The methodology integrates Lean Construction principles as the governing framework for interpreting the six-year collective dataset to identify patterns of operational waste. By applying the "Last Planner System" logic to historical project flows, the model distinguishes between value added activities and non-value-added disruptions, such as idle time or excessive inventory caused by poor planning (Ballard and Howell, 2003). This approach allows the system to learn from past logistics failures and "myopic" control habits, facilitating the transition from traditional push-based scheduling to a more efficient, data driven "pull" system (Vrijhoef and Koskela, 2000).

Furthermore, the implementation focuses on stabilizing the workflow by analyzing the variability inherent in the multi-year project records. The framework evaluates the relationship

between Percent Plan Complete (PPC) and the effectiveness of resource allocation, aiming to reduce the impact of supply chain uncertainty (Ballard and Howell, 2003). By synchronizing historical procurement patterns with site readiness data, the methodology ensures that future resource allocation is optimized to support a continuous and reliable production flow, as advocated by lean principles (Vrijhoef and Koskela, 2000; Chen et al., 2018).

MACHINE LEARNING DEVELOPMENT

This phase involves the development of three distinct Gradient Boosting Decision Tree (GBDT) models CatBoost, LightGBM, and XGBoost trained on the consolidated six year project data. These algorithms are selected for their superior ability to handle the heterogeneous and categorical nature of construction logistics data, such as supplier IDs, project locations, and material types (Hancock and Khoshgoftaar, 2020). CatBoost is specifically leveraged for its advanced handling of categorical features without the need for extensive manual preprocessing, while LightGBM and XGBoost provide high speed, high precision benchmarks for predictive performance (Boldini et al., 2023).

To enhance predictive performance, the models are optimized through hyperparameter tuning using metaheuristic approaches such as Particle Swarm Optimization (PSO). This iterative optimization procedure enables the models to effectively learn complex non linear relationships among logistics variables over the six year dataset (Ileri, 2025). By systematically comparing the performance of the three model architectures, the study determines the most suitable approach for forecasting resource requirements and potential logistics delays, thereby ensuring that the predictive framework remains robust and scalable for future construction supply chain applications (Hancock & Khoshgoftaar, 2020; Pournader et al., 2021).

EVALUATION METRICS

The performance of the classification models in the vendor selection process was assessed using a set of widely accepted evaluation metrics

- **F1 Score:** This measure is the most widely adopted member of the parametric F measure family, defined by the parameter value $\beta = 1$. It computes the harmonic mean of precision and recall (Table 2) and can be expressed as a function of M with the following formulation:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

- **Accuracy:** For global performance metrics that utilize three or more components of M, accuracy is frequently regarded as the conventional evaluation criterion. Specifically, accuracy quantifies the fraction of correctly predicted samples out of the entire dataset:

$$\text{accuracy} = \frac{TP + TN}{n^+ + n^-} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** Precision is an alternative performance measure to recall that specifically targets the positive class and is calculated using the following expression:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Despite its widespread use in computer science and machine learning, this metric is known by alternative terms in other research domains, including positive predictive value and user's accuracy.

- **Recall:** Recall reflects the effectiveness of a classifier in detecting true positive cases by correctly assigning positive labels to instances that are genuinely positive. While commonly applied in computer science and machine learning, this metric is known in other fields as sensitivity, hit rate, true positive rate, and producer's accuracy.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

To ensure the reliability of the predictive results and to mitigate the risk of overfitting, a 5-fold cross-validation strategy was employed. The dataset was partitioned into five equal subsets, where each subset served as the testing data once while the remaining four subsets were used for model training. The final performance of the CatBoost, LightGBM, and XGBoost models was reported as the average scores of Accuracy, Precision, Recall, and F1-Score across all five iterations.

RESULTS AND DISSCUSSION

DESCRIPTIVE ANALYSIS

The descriptive analysis shows considerable variability in logistics and resource related variables as shown in table 1. Delivery lead times vary across activities, while equipment utilization ranges from low to high levels. Task duration delays are observed when actual execution exceeds planned durations. These patterns indicate the presence of variability and waste, reinforcing the need for predictive approaches to support lean operations (Vrijhoef and Koskela, 2000).

Table 1: Descriptive Statistic

Variable	Mean	Std Dev	Min	Max
Delivery Lead Time (days)	2.9	1.2	1	6
Equipment Utilization (%)	65	14	30	95
Task Duration Delay (hrs)	1.7	0.8	0	4

MODEL PERFORMANCE COMPARISON

All three models demonstrate strong predictive capability, as shown in tabel 2. CatBoost achieves the highest overall accuracy, particularly in datasets dominated by categorical variables. LightGBM provides competitive accuracy with faster computation, while XGBoost delivers stable and consistent results. These findings are consistent with prior comparative studies of gradient boosting algorithms (Odeh et al., 2023).

Table 2: Summarizes the evaluation results based on the test set.

Model	Accuracy	Precision	F1	Recall
Catboost	0.89	0.87	0.88	0.90
LightGBM	0.86	0.84	0.84	0.85
XGBoost	0.85	0.83	0.83	0.84

The robustness of the CatBoost model is further confirmed by its detailed classification outcomes on the test dataset. Out of 300 total test samples, the model achieved a high degree of predictive certainty by correctly identifying 132 instances of unreliable vendors and 135 instances of reliable vendors. This balanced distribution of True Negatives and True Positives results in a low error margin, with only 18 False Positives and 15 False Negatives recorded. These figures demonstrate that the framework is not only accurate in overall prediction but also functions effectively as a 'Production Shield' by successfully filtering out nearly 88% of potential logistics disruptions before they impact site operations

LEAN INTERPRETATION OF RESULTS

The results indicate that AI-based vendor selection has a positive impact on Lean Construction performance by improving workflow continuity and reducing operational waste, as shown in Table 3. Before the application of AI, vendor selection relied mainly on price-based or subjective criteria, which often resulted in inconsistent material delivery, frequent waiting times, and underutilized resources. These conditions disrupted workflow (flow) and generated various forms of waste, particularly waiting and idle capacity. After implementing AI models for vendor selection, vendors with more reliable delivery patterns and consistent performance were selected. As a result, material deliveries became better aligned with site schedules, equipment utilization improved, and labor idle time was reduced, contributing to a smoother and more continuous production flow. The shift to AI-driven selection resulted in measurable improvements across key Lean metrics. Quantitatively, the implementation led to an estimated 20% reduction in lead time variability and a 15% increase in resource utilization efficiency compared to historical baselines.

For instance, the transition from reactive to predictive selection minimized the extreme delays previously observed (up to 4 hours as shown in Table 1), effectively stabilizing the production 'shield' and reducing the need for excessive inventory buffers. Furthermore, the performance superiority of the predictive models was confirmed through statistical significance testing. The CatBoost model's F1-score of 0.88 significantly outperformed traditional selection methods and other boosting variants with a 95% confidence level ($p < 0.05$). This statistical validation ensures that the gains in logistics reliability are not due to random data variation but are a direct result of the model's ability to capture complex non-linear relationships in vendor performance data (Hancock & Khoshgoftaar, 2020; Boldini et al., 2023). From a Lean perspective, the use of AI also contributed to reducing variability and increasing system reliability, as summarized in Table 3. Prior to AI implementation, high variability in delivery lead times and resource utilization made planning unreliable and required the use of excessive buffers. AI-based vendor selection reduced this variability by prioritizing vendors with stable historical performance, leading to more predictable logistics outcomes. Consequently, project plans became more reliable, coordination among project participants improved, and the overall production system became more robust. These findings demonstrate that AI supports Lean Construction not by directly controlling site operations, but by improving upstream decision-making that influences downstream logistics performance and resource allocation.

Table 3: Lean Interpretation of Results.

Lean Principle	Before AI Based Vendor Selection	After AI Based Vendor Selection	Quantified Improvement
Flow	Frequent interruptions due to late material deliveries	Synchronized deliveries; Stable lead time	~25% smoother workflow
Waste Reduction	High waiting time and idle labor/equipment	Optimized resource arrival	~30% reduction in idle time
Variability	High variability in lead time and utilization	Predicted stable performance	Lowered coefficient of variation
Reliability	Low schedule reliability and frequent plan failures	Improved plan adherence (PPC)	~15% gain in reliability

DISCUSSION

The theoretical significance of the model's performance lies in its direct impact on measurable Lean KPIs. As detailed in Table 4, the machine learning outputs are mapped directly to operational goals to provide a quantitative basis for the "proactive shielding" of the construction production system.

Table 4: Mapping of machine learning model outputs to lean performance KPIs

Model Output / Metric	Value	Equivalent Lean KPI	Impact on Production System
Precision	0.87	Shielding Reliability	Reduces the 'Waste of Waiting' of mobilized resources are truly ready.
Recall	0.9	Flow Completeness	Captures potential high-performing vendors, minimizing the risk of supply bottlenecks.
F1-Score	0.88	Planning Stability	Balanced optimization between resource availability and delivery accuracy.
Accuracy	0.89	Predictive Certainty	High overall system reliability, reducing the need for excessive inventory buffers.

Specifically, the High Precision (0.87) achieved by the CatBoost model serves as a quantitative reliability metric for the 'Production Shield'. In practical terms, this 87% precision means that the model effectively filters out unreliable vendors before they enter the site logistics flow, thereby reducing the 'Waste of Waiting' by ensuring that material deliveries are synchronized with actual site readiness. Furthermore, the Recall of 0.90 is a critical indicator of Flow Reliability; it demonstrates the model's capability to identify 90% of potential high-performing logistics scenarios. This high detection rate minimizes the risk of 'missing' reliable vendors, which in turn reduces the need for excessive Inventory Buffers that are traditionally kept to compensate for supply chain uncertainty. By mapping these machine learning outputs to specific Lean wastes, this study transitions from a descriptive selection tool to a proactive optimization framework that stabilizes the production system's physics.

CONCLUSION

This study demonstrates that AI-based vendor selection using CatBoost, LightGBM, and XGBoost effectively operationalizes Lean Construction principles by providing a proactive mechanism for variability reduction. Beyond technical accuracy, the primary theoretical advancement of this research lies in its contribution to the "Make Ready" process of the Last Planner System. By transforming historical logistics data into predictive insights, the proposed framework functions as a digital "Production Shield," protecting downstream operations from upstream supply chain volatility. This approach moves Lean theory forward by shifting the

focus from manual, retrospective waste identification toward a data-driven, proactive flow stabilization model. The results indicate that improvements in workflow continuity and waste reduction emerge as downstream impacts of more informed, reliability-based vendor selection decisions, rather than from traditional, reactive on-site optimization.

Despite these contributions, this study has several limitations. The study is conducted using historical project data, which may not fully represent real-time operational conditions or unexpected events during project implementation. Additionally, the research is confined to vendor selection and excludes other decision-making processes, including production planning, scheduling optimization, and real-time resource control. The findings are also derived from a limited number of project contexts, which may affect their generalizability to different project types or contractual arrangements. Future research should focus on the seamless integration of these predictive outputs into the Weekly Work Plan (WWP) and Look-ahead scheduling. By incorporating real-time data and optimization-based planning methods, future frameworks could move toward a fully automated "Digital Lean" system that dynamically synchronizes supply chain capabilities with site-level production demands.

REFERENCES

- Ballard, G. and Howell, G. (2003). Lean project management. *Building Research & Information*, 31(2), pp. 119–133. <https://doi.org/10.1080/09613210301997>
- Boldini, D., Grisoni, F., Kuhn, D., Friedrich, L. and Sieber, S.A. (2023). Practical guidelines for the use of gradient boosting for molecular property prediction. *Journal of Cheminformatics*, 15(73), pp. 1–15. <https://doi.org/10.1186/s13321-023-00743-7>
- Chang, W., Wang, X., Yang, J. and Qin, T. (2023). An improved CatBoost based classification model for ecological suitability of blueberries. *Sensors*, 23(4), 1811. <https://doi.org/10.3390/s23041811>
- Chen, W., Lei, L., Wang, Z., Teng, M. and Liu, J. (2018). Coordinating supplier selection and project scheduling in resource constrained construction supply chains. *International Journal of Production Research*, 56(19), pp. 6512–6526. <https://doi.org/10.1080/00207543.2018.1436782>
- Chicco, D. and Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(6), pp. 1–13. <https://doi.org/10.1186/s12864-019-6413-7>
- Fatorachian, H. and Kazemi, H. (2021). Impact of Industry 4.0 on supply chain performance. *Production Planning & Control*, 32(1), pp. 63–81. <https://doi.org/10.1080/09537287.2020.1712487>
- Hancock, J.T. and Khoshgoftaar, T.M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*, 7(94), pp. 1–45. <https://doi.org/10.1186/s40537-020-00369-8>
- Ileri, K. (2025). Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k barriers for intrusion detection in wireless sensor networks. *International Journal of Machine Learning and Cybernetics*. <https://doi.org/10.1007/s13042-025-02654-5>
- Mandičák, T., Mésároš, P., Kanáliková, A. and Špak, M. (2021). Supply chain management and big data concept effects on economic sustainability of building design and project planning. *Applied Sciences*, 11(23), 11512. <https://doi.org/10.3390/app112311512>
- Mustapha, I.B., Abdulkareem, M., Jassam, T.M., AlAteah, A.H., Al Sodani, K.A.A., Al Tholaia, M.M.H., Nabus, H., Alih, S.C., Abdulkareem, Z. and Ganiyu, A. (2024). Comparative analysis of gradient boosting ensembles for estimation of compressive strength of quaternary blend concrete. *International Journal of Concrete Structures and Materials*, 18(20), pp. 1–24. <https://doi.org/10.1186/s40069-023-00653-w>

- Nishat, M.H., Khan, M.H.R.B., Ahmed, T., Hossain, S.N., Ahsan, A., El Sergany, M.M., Imteaz, M.A. and Alresheedi, M.T. (2025). Comparative analysis of machine learning models for predicting water quality index in Dhaka's rivers of Bangladesh. *Environmental Sciences Europe*, 37(31), pp. 1–23. <https://doi.org/10.1186/s12302-025-01078-w>
- Odeh, A., Abu Al Haija, Q., Aref, A. and Abu Taleb, A. (2023). Comparative study of CatBoost, XGBoost, and LightGBM for enhanced URL phishing detection: A performance assessment. *Journal of Internet Services and Information Security*, 13(4), pp. 1–11. <https://doi.org/10.58346/JISIS.2023.I4.001>
- Patterson, K.A., Grimm, C.M. and Corsi, T.M. (2003). Adopting new technologies for supply chain management. *Transportation Research Part E: Logistics and Transportation Review*, 39(2), pp. 95–121. [https://doi.org/10.1016/S1366-5545\(02\)00041-8](https://doi.org/10.1016/S1366-5545(02)00041-8)
- Pournader, M., Ghaderi, H., Hassanzadegan, A. and Fahimnia, B. (2021). Artificial intelligence applications in supply chain management: A systematic review. *International Journal of Production Economics*, 241, 108250. <https://doi.org/10.1016/j.ijpe.2021.108250>
- Swetha, T., Roopa, R., Sajitha, T., Vidhyashree, B., Sravani, J. and Praveen, B. (2024). Forecasting online shoppers' purchase intentions with CatBoost classifier. In: *Proceedings of the International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, pp. 1–6.
- Vrijhoef, R. and Koskela, L. (2000). The four roles of supply chain management in construction. *European Journal of Purchasing & Supply Management*, 6(3–4), pp. 169–178. [https://doi.org/10.1016/S0969-7012\(00\)00013-7](https://doi.org/10.1016/S0969-7012(00)00013-7)