Transactions on Engineering and Computing Sciences - Vol. 12, No. 06
Publication Date: December 25, 2024
DOI:10.14738/tmlai.1206.18078.
Achatbi, I. (2024). Leveraging Machine Learning Techniques to Forecast and Enhance Supplier Reliability in Supply Chain Management. Transactions on Engineering and Computing Sciences, 12(06). 54-63.

Leveraging Machine Learning Techniques to Forecast and Enhance Supplier Reliability in Supply Chain Management

Iman Achatbi
National School of Applied Sciences – Tangier

ABSTRACT

In today's competitive market, selecting reliable suppliers is crucial to ensure supply chain efficiency. In fact, anticipating supplier behavior plays a vital role in effectively managing the risk of disruptions, enabling companies to develop proactive strategies to mitigate potential supply chain interruptions. The complexity of supplier management has driven companies to adopt artificial intelligence (AI) and machine learning, to enhance decision-making in the upstream supply chain. By analyzing historical data, machine learning models help predict risks and improve supplier reliability. These predictive capabilities allow businesses to identify vulnerabilities early, ensuring better risk preparedness and supply chain resilience. This article examines the application of AI to tackle supplier selection challenges, emphasizing its role in transforming supply chains into agile, data-driven, and predictive systems while addressing the critical need to manage disruption risks effectively.

Keywords: Artificial intelligence, Machine learning, Supplier reliability, Logistic Regression.

INTRODUCTION

The advent of Logistics 4.0—the integration of advanced digital technologies into supply chain management—has revolutionized the industrial landscape. By incorporating automation, IoT (Internet of Things), and data analytics, Logistics 4.0 has enabled companies to optimize operations, enhance visibility, and improve decision-making processes. However, despite these advancements, the reliability of suppliers remains a persistent challenge, particularly in upstream logistics, where unpredictable supplier behavior can disrupt production schedules, increase costs, and reduce customer satisfaction.

The issue of unreliable suppliers is compounded by their tendency to deliver late, fail to meet quality standards, or provide inconsistent communication, leading to cascading effects across the supply chain. Anticipating supplier behavior, therefore, is critical to mitigating risks, ensuring production continuity, and maintaining competitiveness. The ability to foresee a supplier's reliability—or lack thereof—is no longer just an operational need but a strategic imperative in modern supply chains. In this context, the use of Artificial Intelligence (AI) in supply chain management has emerged as a transformative solution. AI-powered tools enable organizations to analyze large datasets, uncover hidden patterns, and make informed predictions about supplier performance. Specifically, in upstream logistics, AI offers the potential to evaluate supplier reliability based on historical data, enabling proactive decisionmaking and fostering a resilient supply network.

The primary objective of this article is to develop a **machine learning (ML) model** tailored to assist production facilities in predicting supplier behavior. By leveraging advanced algorithms, the proposed model aims to provide actionable insights, allowing factories to classify suppliers as reliable or unreliable and make informed decisions about sourcing strategies. This proactive approach not only enhances operational efficiency but also builds a foundation for sustainable and resilient supply chain ecosystems.

By addressing these challenges, this research contributes to the growing body of knowledge on leveraging AI in supply chain management and highlights the critical importance of supplier reliability in achieving Logistics 4.0 goals.

LITERATURE REVIEW

The integration of artificial intelligence (AI) and machine learning (ML) into the supply chain has gained significant attention in recent years. These technologies have demonstrated immense potential in optimizing various aspects of supply chain management, such as demand forecasting, inventory optimization, and route planning. However, despite the increasing focus on AI-driven solutions, the application of these tools in supplier selection remains relatively underexplored. According to the study conducted by (Naqvi, 2021), several studies have addressed supplier selection using traditional methods such as Multi-Criteria Decision-Making (MCDM) techniques, the most popular techniques in this field are: Fuzzy TOPSIS (Achatbi, 2020), Fuzzy multi objective programming (Babic, 2014), Stochastic programming (Manerba, 2018), and Mixed-integer linear programming (Sodenkamp, 2016).

There is a significant gap in utilizing AI and ML methods for supplier selection, particularly within the context of circular supply chains. According to a review conducted by (Farshadfar, 2024), only a minority of published studies have explored supplier selection using these advanced technologies. This gap highlights an opportunity for innovative approaches to improve the reliability and efficiency of supplier evaluation processes. (Fallahpour, 2016) focused on supplier selection using hybrid AI-based models and introduced innovative methods combining 'data envelopment analysis' and 'artificial neural networks.' Additionally, it proposed a novel genetic programming (GP) approach to enhance decision-making efficiency in this context. (Kamalahmadi, 2016) developed a two-stage mixed-integer programming model to create flexible sourcing strategies that mitigate supply and environmental risks while minimizing total supply chain costs. (Tavana, 2016) proposed a Multi-Layer Perceptron (MLP) to predict and rank supplier performance. (Choy, 2003) introduced a combined ANN-based model designed to select and benchmark potential partners for Honeywell Consumer Products Limited in Hong Kong. (Sawik, 2011) analyzed supplier selection and order allocation decisions by considering factors such as the price and quality of purchased components, along with the reliability of deliveries, in scenarios involving both local and global disruptions. (Ruiz-Torres, 2013) examined scenarios involving multiple demand points and suppliers, each characterized by distinct costs and reliability levels. A decision-tree approach was employed to evaluate all possible contingencies in the event of failure by one or more suppliers.

Our work seeks to address this need by developing an AI-driven model to classify suppliers based on their reliability. By integrating multiple factors that influence supplier performance, such as delivery timelines, product quality, and after-sales service, we aim to create a comprehensive and scalable framework. This approach not only helps in identifying reliable suppliers but also aids in the formation of a robust database of trustworthy partners, enhancing decision-making for procurement teams. The result is a smarter, data-driven strategy for supplier management that aligns with the broader goals of modern supply chain innovation.

Problem Formulation

The goal of this study is to address the challenge of classifying suppliers into reliable and unreliable categories based on a set of criteria that describe both the supplier's characteristics and the attributes of the products supplied. A reliable classification system is essential for enhancing the efficiency and robustness of upstream logistics by predicting the behavior of suppliers, including those not previously engaged.
To achieve this, the study proposes to:

1. Identify key criteria (features) that influence supplier reliability, encompassing supplier-specific attributes (e.g., response time, after-sales service quality, financial stability) and product-specific metrics (e.g., delivery punctuality, quality compliance).
2. Determine the weight of each criterion to reflect its relative importance in assessing supplier reliability.
3. Develop a predictive model using the logistic regression algorithm to classify suppliers as reliable or unreliable, providing a straightforward yet effective approach to supplier evaluation.

Given the lack of a comprehensive real-world dataset, the study will create a synthetic dataset to simulate a realistic environment. This dataset will include records of orders made to various suppliers, alongside relevant supplier and product information. The generated data will serve as a testing ground for validating the model's accuracy and effectiveness in classifying suppliers and anticipating the behavior of new ones.

By tackling this problem, the research aims to contribute to the optimization of supplier selection processes, reduce risks associated with unreliable suppliers, and improve overall supply chain resilience.

PROBLEM SOLVING METHOD

Features Impacting Supplier Reliability

Features that impact supplier reliability can be categorized into different groups based on supplier attributes and the products or services they provide. In our study, we aim to classify the criteria impacting supplier reliability. These criteria can be grouped into three main categories:

- Product-Related Criteria:
  - Product Quality: Whether the supplied product meets the agreed-upon standards.
  - Order Accuracy: Whether the quantity delivered matches the quantity ordered.
  - Unit Price: The price of the product per unit as agreed in the order.
- Delivery-Related Criteria:

- Planned Delivery Date vs. Actual Delivery Date: The adherence of the supplier to the scheduled delivery time.
- Order Date: The date when the order was placed. o Delivery Delays: The time difference between the expected delivery date and the actual delivery date.

- Supplier-Specific Criteria:
  - Response Time: The time the supplier takes to respond to inquiries or requests for information.
  - After-Sales Service: The quality and reliability of support provided postdelivery.
  - Annual Revenue: The supplier's financial capacity as indicated by their turnover.

By analyzing these diverse criteria, we aim to better understand the factors influencing supplier reliability and use them to build a predictive model for supplier classification.

Algorithm of Logistic Regression

In our work, we use Logistic regression wish is a machine learning algorithm for binary classification problems. It predicts the probability of an observation belonging to one of two categories, in our case reliable or unreliable supplier. Logistic regression uses a linear combination of input features but applies a sigmoid function to map the output to a probability between 0 and 1:

$$h_\theta(x) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

Where:

- $z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$ (linear combination of inputs)
- σ(z): Sigmoid function
- θ: Model coefficients (weights)

Then after, the predicted probability ($h_\theta(x)$) is used to classify:

- $h_\theta(x) \geq 0.5$: Classify as 1
- $h_\theta(x) < 0.5$: Classify as 0

For the cost function, logistic regression minimizes the log loss function below:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log(1 - h_\theta\left(x^{(i)}\right))]$$

Where:

- $y^{(i)}$:Actual label for the i^{th} observation (0 or 1).
- $h_\theta(x^{(i)})$: Predicted probability for the i^{th} observation.
- m: Number of observations.

This function ensures that predictions closer to the true labels result in a lower cost.

IMPLEMENTATION AND RESULTS

Data Generation

The performance of a machine learning model cannot achieve its intended goal without the availability of data for testing. Therefore, in this study, we address the challenge by generating a synthetic dataset using Python. This approach is necessary because manufacturing plant data is currently unavailable.

The objective is to create a dataset that records orders placed with various suppliers, along with detailed information about these suppliers. This synthetic dataset will be used to evaluate the proposed algorithm and validate its effectiveness in classifying suppliers based on their reliability. To generate the dataset, we proceed as follow:

- Randomly assign attributes such as order date, quantity ordered, price, and delivery times.
- Compute delivery delays (actual delivery date vs. planned delivery date).
- Track the conformity of delivered products to quality standards.
- We define a function <classer_fiabilite> to assign the attribute Reliability_Supplier based on Average Delivery Delay, Conformity Rate and after sales service.

Here is a visualization of features distribution in our generated dataset (figure 1):

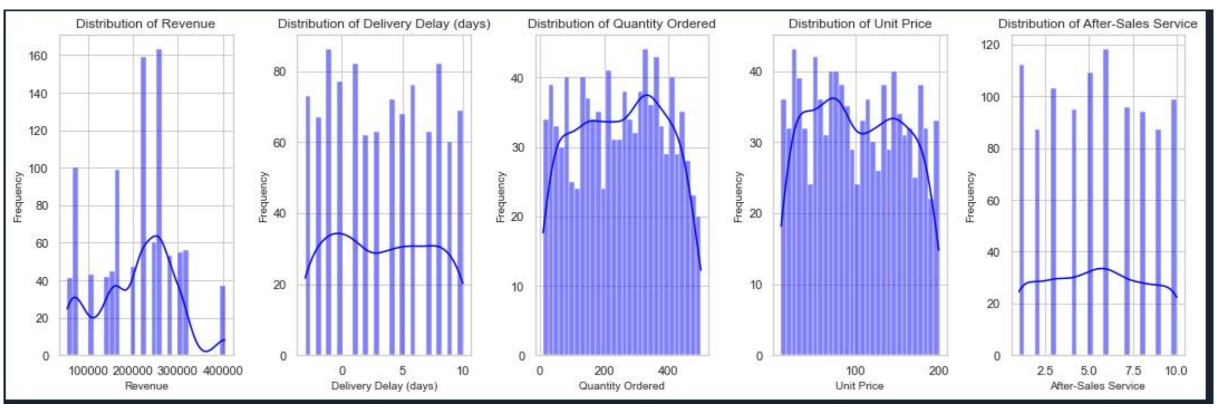

Figure 1: Features distribution

Machine Learning algorithms generally require numerical values as input for predictor variables. Therefore, **Label Encoding** is essential for transforming categorical labels into numerical representations. After normalizing the input data, machine learning algorithms often require feature scaling. This involves transforming the range of values from 0 to n into a more restricted range, typically between 0 and 1, using a **MinMaxScaler method**. The figure 2 below emphasize the heat Map of the correlation between the target variable and others attributes.

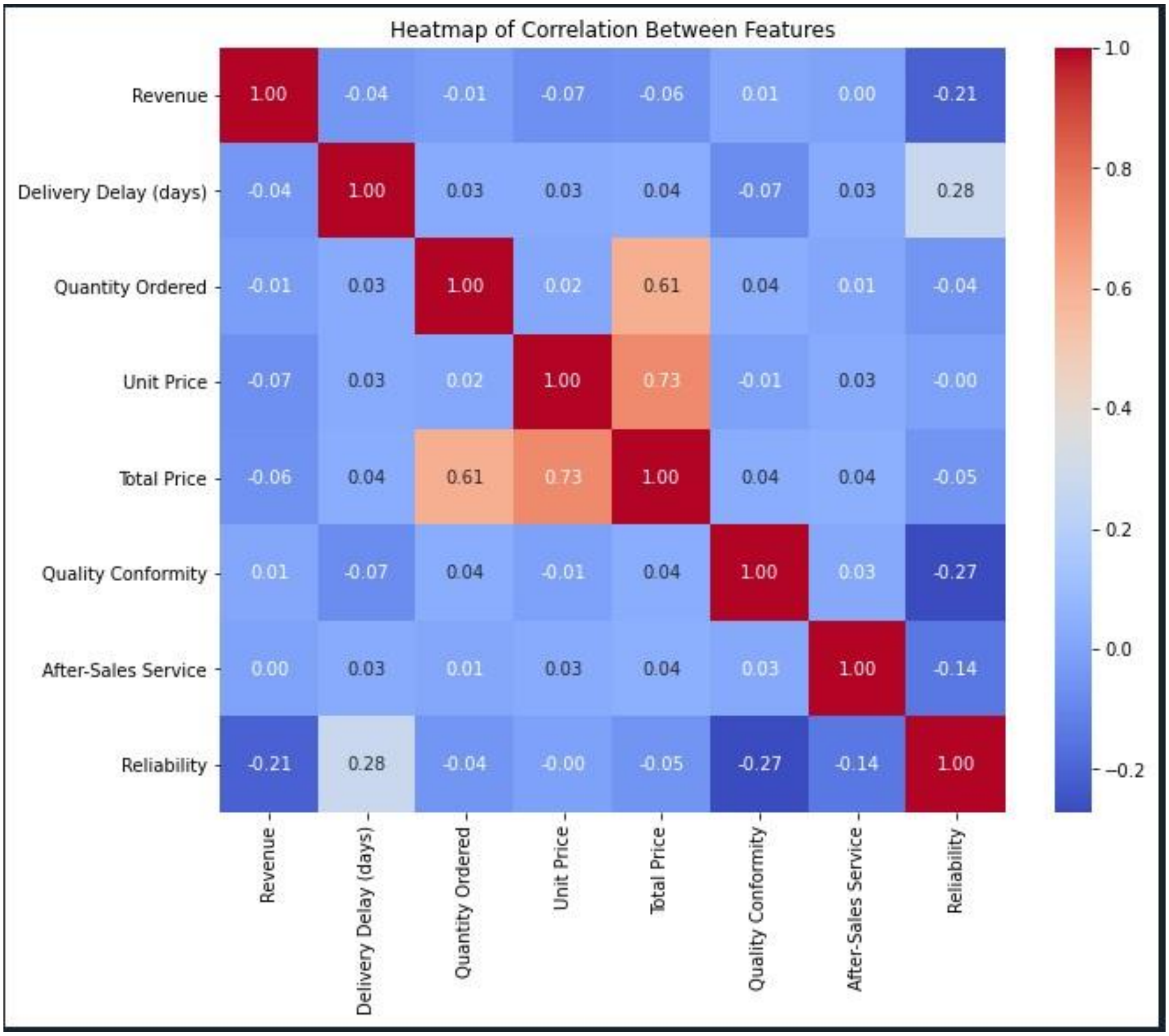

Figure 2: Heat map of the Correlation between the target attribute and the other attributes

Performance Measures:

To effectively evaluate the performance of a machine learning algorithm, it is essential to define a set of metrics tailored to the problem being addressed. These metrics provide insights into how well the model is performing and whether it meets the desired objectives. Commonly used metrics include:

Accuracy:

It is a straightforward metric to measure the proportion of correct predictions made by the model out of all predictions. In our dataset, the value of accuracy is: accuracy=0.87

Confusion Matrix:

To provide more detailed breakdown of the model's performance, we use the confusion matrix to show the counts of true positives TP, true negatives TN, false positives FP, and false negatives FN as shown in figure 3. This breakdown helps identify specific areas where the model may be making errors.

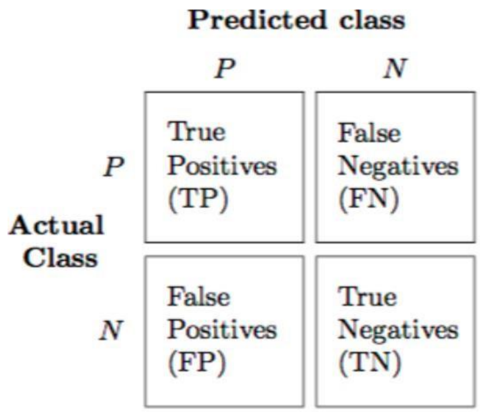

Figure 3: The Confusion Matrix

The confusion matrix for our generated dataset is:

$$\begin{pmatrix} 170 & 0 \\ 26 & 4 \end{pmatrix}$$

Classification Report:
The classification report includes more detailed metrics, such as:

- **Precision**: The proportion of true positive predictions among all positive predictions.
- **Recall (Sensitivity)**: The proportion of true positives correctly identified out of all actual positive cases.
- **F1-Score**: The harmonic mean of precision and recall, offering a balance between the two.

The table below shows the classification report of our Model:

Prediction	Precision	Recall	F1-score	Support
0	0.87	1.00	0.93	170
1	1.00	0.13	0.24	30

AUC - ROC curve is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represents degree or measure of separability. AUC - ROC curve is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represents degree or measure of separability. The figure below shows the value of AUC which is **AUC = 0.86**, this indicates that the model has an good ability to distinguish between the two classes (Reliable and Unreliable suppliers).

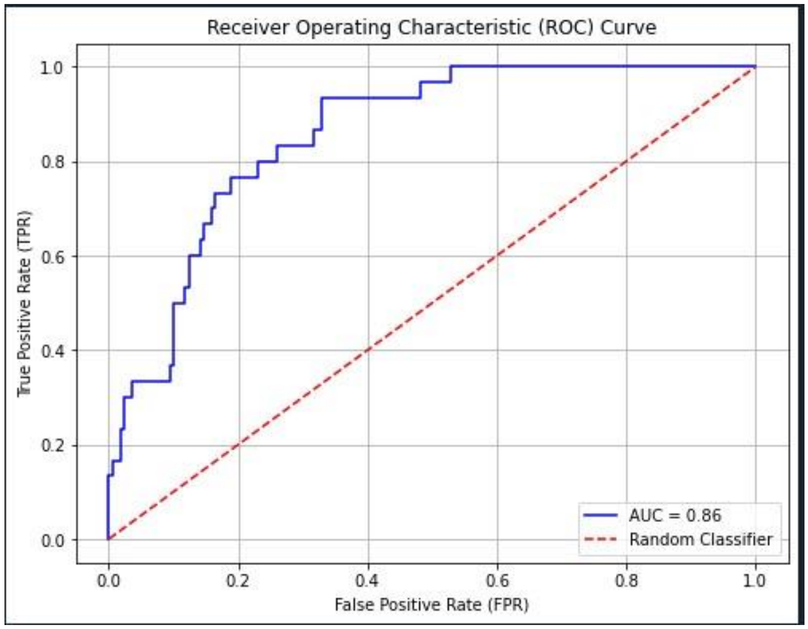

Figure 4: AUC-ROC Curve of Logistic Regression

COMPARISON AND DISCUSSION

Logistic regression is a model that establishes a **linear relationship** between the input features and the target variable. This linearity is reflected in the **coefficients** of the regression model, which serve as a direct indicator of the contribution of each feature to the outcome. While logistic regression is highly interpretable and effective for understanding relationships in data, it assumes linearity and may not fully capture the complexity of feature interactions or nonlinear effects. To gain a more comprehensive understanding of feature importance, it is valuable to compare the logistic regression coefficients with results from other methods especially nonlinear one. For this purpose, we apply the decision tree method to show the nonlinear relationship between features, this method measures how much each feature contributes to reducing impurity (e.g., Gini Index or Entropy) when splitting data in the tree.

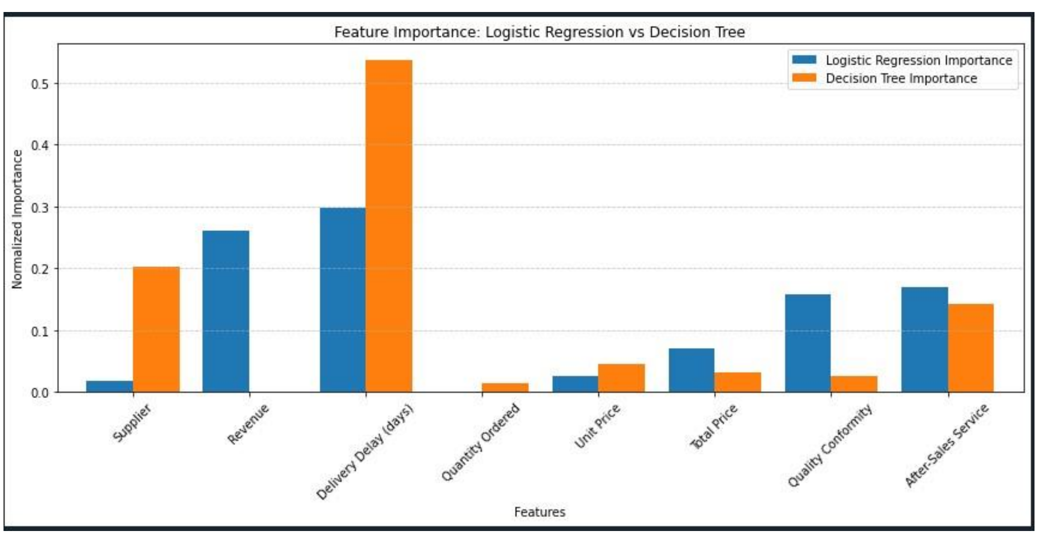

Figure 5: Feature importance logistic regression Vs decision Tree

As general observations, the dominant Features are delivery delay, Revenue and after-sales Service for both methods logistic regression and decision tree. For the decision tree model, the most important feature with approximately 50% importance goes to delivery delay. Regarding logistic regression, revenue has moderate importance comparing to decision tree with relatively lower weight. Minor features are Quantity Ordered, Unit Price, Total Price. These features have relatively low importance for both algorithms, indicating limited influence on the model's predictions. According to the comparison between the two algorithms, Logistic Regression distributes importance more evenly among features like Revenue, Delivery Delay, Quality Conformity, and After-Sales Service, and it suggests a more balanced reliance on multiple features. Decision Tree strongly favors Delivery Delay, emphasizing this feature almost exclusively, and likely captures non-linear relationships that make this feature critical in decision-making. Other features, such as Revenue and After-Sales Service, are secondary. To conclude, for the decision tree, the high dependency on Delivery Delay may indicate overfitting or strong nonlinear patterns in the data. Further tuning may help balance feature importance. Regarding Logistic Regression, the importance distribution suggests better generalization but may underutilize potentially powerful non-linear patterns in the data.

CONCLUSION

The objective of this article is to classify suppliers based on their reliability and to build a comprehensive database of dependable suppliers. To achieve this, we propose a machine learning-based prediction model designed to anticipate supplier behavior effectively. Recognizing the lack of available real-world data, we generated a test dataset using a tailored algorithm specifically developed for this purpose. This study is particularly valuable for companies aiming to proactively manage supplier relationships by predicting potential risks and ensuring more reliable supply chain operations. The proposed approach not only helps in optimizing supplier selection but also supports decision-making in scenarios of uncertainty.

Future research could expand this work by incorporating additional criteria related to environmental and social risks, which are increasingly critical in sustainable supply chain management. Such enhancements would improve the robustness of the model and provide a broader framework for evaluating suppliers in line with emerging global priorities.

References

Achatbi, I. A. (2020). Advanced system based on ontology and multi agent technology to handle upstream supply chain: intelligent negotiation protocol for supplier and transportation provider selection. Decision Science Letters, 9(3), 337-354.

Babic, Z. &. (2014). Multiproduct vendor selection with volume discounts as the fuzzy multiobjective programming problem. International Journal of Production Research, 52(14),4315–4331.

Choy, K. L. (2003). Design of an intelligent supplier relationship management system: a hybrid case based neural network approach. Expert systems with Applications, 24(2), 225-237.

Fallahpour, A. O. (2016). An integrated model for green supplier selection under fuzzy environment:

application of data envelopment analysis and genetic programming approach. Neural Computing and Applications, 27, 707-725.

Farshadfar, Z. M. (2024). Leveraging Machine Learning for Advancing Circular Supply Chains: A Systematic Literature Review. Logistics, 8(4), 108.

Kamalahmadi, M. &.-P. (2016). Developing a resilient supply chain through supplier flexibility and reliability assessment. International Journal of Production Research, 54(1), 302-321.

Manerba, D. M. (2018). The capacitated supplier selection problem with total quantity discount policy and activation costs under uncertainty. International Journal of Production Economics, 198, 119–132.

Naqvi, M. A. (2021). Supplier selection and order allocation: a literature review. Journal of Data, Information and Management, 3(2), 125-139.

Naz, F. A. (2022). Reviewing the applications of artificial intelligence in sustainable supply chains: Exploring research propositions for future directions. Business Strategy and the Environment, 31(5), 2400-2423.

Ruiz-Torres, A. J. (2013). Supplier selection model with contingency planning for supplier failures. Computers & Industrial Engineering, , 66(2), 374-382.

Sawik, T. (2011). Selection of supply portfolio under disruption risks. Omega, 39(2), 194-208.

Sodenkamp, M. T. (2016). Modeling synergies in multi-criteria supplier selection and order allocation: An application to commodity trading. European Journal of Operational Research, 254(3), 859–874.

Tavana, M. F.-A. (2016). A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection. Expert Systems with Applications, 61, 129-144.