Identifying Critical Factors for Supplier Selection using Machine Learning

Arun Kumar Mishra[#], Megha Sinha[*] , Sudhanshu Kumar Jha[$]

[#]*Department of Computer Science and Engineering,University College of Engineering and Technology(UCET)*
Vinoba Bhave University
Hazaribag-825301, Jharkhand, India

[*]*Department of CSE, Sarala Birla University, Ranchi-835103, Jharkhand, India*

#Department of Electronics and Communication, Faculty of Science, University of Allahabad, Prayagraj-211002, Uttar Pradesh, India

Abstract

A company's operational efficiency and competitive advantage are directly impacted by the supplier selection process, which is a crucial aspect of supply chain management. The purpose of this study is to use Python-based machine learning algorithms to determine the essential parameters for supplier selection. In order to identify the most important variables, a synthetic dataset was created that mimics real-world supplier metrics and ran it through the Random Forest Regressor, a machine learning model. The findings show that a number of important parameters, including cost, quality, and delivery time, affect supplier success. The results open the door for more data-driven decision-making in supplier selection procedures and offer procurement managers useful information.

.

Keywords: Machine Learning, Supply Chain Management, Supplier Selection, Python, Random Forest Regressor, Critical Factors.

1. Introduction

The performance of a supply chain's suppliers has a significant impact on its efficacy and efficiency. The selection of suppliers has a significant impact on delivery, quality, cost, and overall customer satisfaction (Chopra, 2016). Conventional supplier selection techniques frequently depend on basic standards and subjective assessment, which may not adequately account for the complexity of contemporary supply chains (Christopher, 2016). Big data and machine learning have made it feasible to evaluate big datasets and identify important variables that affect supplier performance. This study investigates the use of Python and machine learning techniques—more specifically, the Random Forest Regressor—to determine these parameters on a synthetic dataset.

2. Literature Review

With an emphasis on the effects of artificial intelligence (AI) on operational effectiveness, strategic innovation, and sustainability, the article by Eyo-Udo, 2024 offers a thorough analysis of the integration of AI into supply chain management (SCM). They suggested that Artificial intelligence (AI) improves decision-making, lowers costs, and optimizes resource allocation to dramatically increase supply chain efficiency. Further, Data privacy issues, moral dilemmas, and the shortage of qualified workers are only a few of the difficulties that impact supply chain management's use of AI. Also, with room for more innovation and resilience, the future of AI-enhanced supply chains seems bright, but only if current issues are resolved. Changalima et al. 2024 looked at how supplier selection and monitoring might improve Tanzania's public procurement efficiency in terms of cost savings. They found that the efficiency of public procurement in terms of cost reduction is positively and significantly predicted by supplier selection and supplier monitoring. The cost of procurement is reduced by 27.2% for every unit increase in supplier selection. Also, a significant positive predictor of lower procurement costs is supplier monitoring. Lean, agile, resilient, and green (LARG) criteria are the basis of a unique hybrid fuzzy decision-making framework that is proposed in the research conducted by Sheykhzadeh et al. 2024 for supplier selection in a pharmaceutical supply chain. Their paper also examines the significance of these criteria prior to and following the COVID-19 pandemic. Just-in-time delivery, lead time, and safety stock became more significant criteria for selecting suppliers after the pandemic, replacing quality, cooperation, and environmental considerations. Healthcare supply chains can more effectively adapt to market demands during disruptions by using supplier selection strategies that place a high priority on product availability and prompt response, such as keeping excess inventory on hand. Machine learning algorithms assist in supplier selection and procurement by analyzing historical procurement data, supplier performance metrics, and market conditions. Using supervised learning models, such as decision trees and random forests, suppliers can be ranked and assessed based on attributes such as quality, price, and delivery reliability.

One interesting use of machine learning is the prediction of supplier risks and disruptions (Ni et al., 2019). By analyzing data from multiple sources, including news articles, financial reports, and social media, machine learning models can identify potential risks and recommend alternative suppliers. Businesses can lower supply chain risks and ensure continuity by taking the initiative. The manufacturing and supply chain management industries have seen a substantial transformation with the advent of Industry 4.0, necessitating the adaptation of supplier selection processes to this quickly evolving technological landscape. Major areas of focus for the comprehensive review study include the function of multi-criteria decision-making (MCDM) techniques and recent advances in the context of Industry 4.0 supplier selection (Sushil Kumar Sahoo, Shankha Shubhra Goswami, and Halder 2024). It examined the opportunities and challenges presented by these developments in this new era of technological advancements like the Internet of Things, big data analytics, artificial intelligence, and advanced manufacturing techniques. It also highlights how crucial effective supplier selection is to achieving operational excellence and enhancing supply chain resilience. The study looked at the MCDM techniques' theoretical underpinnings and highlights how well-suited they were to deal with the complex, multidimensional requirements associated with supplier selection.

3. Present Work

First of all, a synthetic dataset was created in order to mimic actual circumstances. A supplier is represented by each of the 1,000 records in the dataset, which includes information on cost, quality, delivery time, innovation, flexibility, risk, sustainability, profit impact, and dependability. The numpy and pandas packages for Python were used to construct these features, guaranteeing a wide variety of values to represent various supplier attributes. The attributes descriptions are as follows:

a) Supplier_ID: A distinct number between 1 and 1000 that is assigned to every supplier.
b) Cost: A variable cost, between 50 and 500, related to every supplier.
c) Quality: A number between 1 and 10 indicating each supplier's quality.
d) Delivery_Time: The number of days (from 1 to 30) for the delivery.
e) Innovation: A number between 0 and 1 that indicates how innovative the provider is.
f) Flexibility: A number between 0 and 1 that represents the supplier's degree of flexibility.
g) Risk: A number between 0 and 1 that indicates the possible dangers connected to each provider.
h) Sustainability: A number between 0 and 1 that represents the supplier's dedication to

sustainable practices.

i) Profit_Impact: A number between 0 and 1 that indicates how much of an influence the supplier has on profitability.

j) Supply_Risk: A number between 0 and 1 that represents the chance of a disruption in supplies.

k) Strategic_Importance: A number between 0 and 1 that indicates how important a provider is strategically.

l) Supplier_Reliability: A zero-to-one reliability score assigned to each supplier.

m) Score: A number between 0 and 100 representing each supplier's overall performance.

The sample dataset has been shown in Figure 1.

	Supplier_ID	Cost	...	Supplier_Reliability	Score
0	1	218.543053	...	0.373641	53.303056
1	2	477.821438	...	0.332912	13.789882
2	3	379.397274	...	0.176154	59.124291
3	4	319.396318	...	0.607267	31.478562
4	5	120.208388	...	0.476624	5.234877

Figure 1. Sample Dataset

Then, the target variable (Score) and features were separated out of the dataset. Prior to modeling, the features were standardized using sklearn's StandardScaler to make sure they are on a similar scale. Random forest regression is an essential tool in data science. Our powerful machine-learning technology helps us analyze large, complicated datasets and produce accurate forecasts. A Random Forest Regression Model is a single model that combines several decision trees. Each tree in the forest uses a different subset of the available data to inform its own prediction. The final input prediction is based on the weighted average, or average, of all the predictions given by each individual tree ("Random Forest Regression — How It Helps in Predictive Analytics?," 2023). Because of the Random Forest Regressor's resilience when dealing with non-linear data and capacity to prioritize features, it was chosen. To verify the performance of the model, the dataset was divided into training (70%) and testing (30%) sets. The workflow has been shown in Figure 2.

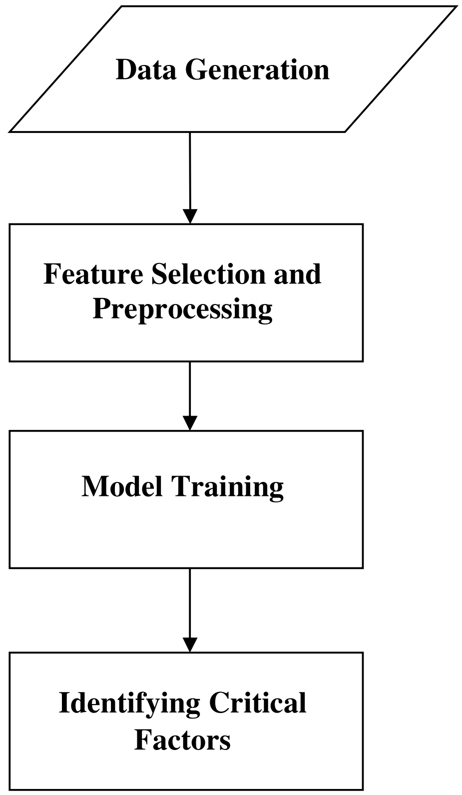

Figure 2. Workflow

4. Result and Discussions

After running the model on the test data, features and their importance was plotted and we got the graph as shown in the Figure 3.

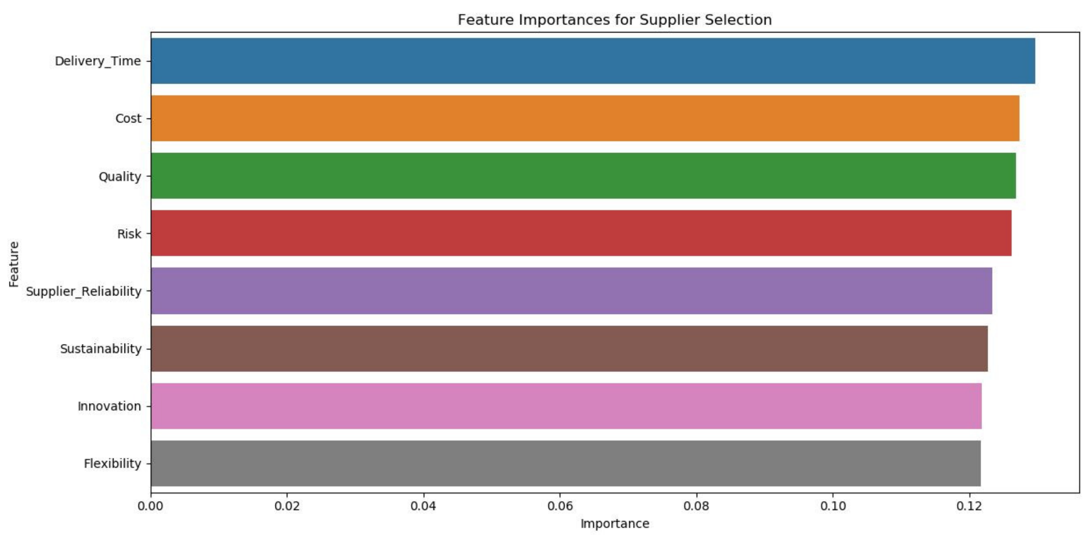

Figure 3. Feature Importance for Supplier Selection

It can be seen from the feature importance analysis that Delivery_Time, Cost and Quality were the top three factors influencing the supplier's overall performance score. The results underline the growing significance of delivery time, reliability and sustainability while confirming the value of conventional supplier selection criteria like cost and quality.

5. Conclusion

This study showed how to use machine learning and Python to find important variables while choosing a supplier. Delivery time, Quality and price emerged as the most crucial aspects according to the Random Forest model's successful ranking of the variables. With the use of this information, businesses may improve the performance of their supply chains and make better procurement decisions by streamlining their supplier selection procedures. The analysis was done on a synthetic dataset. It can be applied to the real datasets and comparative analysis of other machine learning algorithms may also be performed to choose the best model. The insights provided by the analysis might be useful for procurement managers while selecting the suppliers.

References

Changalima, I. A., Ismail, I. J., & Mchopa, A. D. (2024). Effects of supplier selection and supplier monitoring on public procurement efficiency in Tanzania: a cost-reduction perspective. *Vilakshan-XIMB Journal of Management*, *21*(1), 55-65.

Chopra, Sunil, and Peter Meindl. 2016. *Supply Chain Management : Strategy, Planning, and Operation*. 6th ed. Boston, Mass.: Pearson.

Christopher, Martin. 2016. *Logistics & Supply Chain Management*. 5th ed. Harlow, United Kingdom: Pearson Education.

Eyo-Udo, N. (2024). Leveraging artificial intelligence for enhanced supply chain optimization. *Open Access Research Journal of Multidisciplinary Studies*, *7*(2), 001-015.

Ni, Du, Zhi Xiao, and Ming K. Lim. 2019. "A Systematic Review of the Research Trends of Machine Learning in Supply Chain Management." *International Journal of Machine Learning and Cybernetics* 11 (December). https://doi.org/10.1007/s13042-019-01050-0.

Random Forest Regression — How it Helps in Predictive Analytics? (2023, December 26). *Medium*. https://medium.com/@byanalytixlabs/random-forest-regression-how-it-helps-in-predictive-analytics-01c31897c1d4

Sheykhzadeh, M., Ghasemi, R., Vandchali, H. R., Sepehri, A., & Torabi, S. A. (2024). A hybrid decision-making framework for a supplier selection problem based on lean, agile, resilience, and green criteria: A case study of a pharmaceutical industry. *Environment, Development and Sustainability*, 1-28.

Sushil Kumar Sahoo, Shankha Shubhra Goswami, and Rohit Halder. 2024. “Supplier Selection in the Age of Industry 4.0: A Review on MCDM Applications and Trends.” *Decision Making Advances* 2 (1): 32–47. https://doi.org/10.31181/dma21202420.