****

**Enhancing Supplier Selection in Supply Chain Management: A Comparative Study of Machine Learning Algorithms**

Mohammad Hossein Nazari

S23120465

January 2024

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science in Big Data Analytics

School of Computing, Engineering and the Built Environment

Birmingham City University

**Acknowledgements**

I would like to express my profound gratitude to my supervisor, Professor Dr. Atif, for his exceptional guidance, patience, and expertise which have been instrumental to the success of my research. His mentorship has truly enriched my work, providing a foundation of academic rigor and intellectual growth.

I am deeply thankful to DPC Company for their invaluable assistance and support throughout this research. The insights, resources, and collaboration provided by the team were vital in the practical elaboration and grounding of my work.

Finally, I owe a debt of heartfelt thanks to my parents. Their unwavering love, encouragement, and sacrifices have not only shaped my character but have also underscored my academic endeavors. It is with their support that I have been able to pursue my passions and reach this milestone.

**Abstract**

In the rapidly evolving domain of supply chain management, the critical task of supplier selection has undergone significant transformation with the integration of machine learning (ML). This article scrutinizes the application of two prominent ML algorithms, Random Forest and XGBoost, in streamlining supplier selection processes. Traditional selection methods, often fraught with subjectivity and inefficiency, are giving way to novel, data-driven approaches capable of predictive analytics and empirical decision-making. This study outlines the key benefits of leveraging ML, such as enhanced performance evaluation, sustainability scoring, risk assessment, and criteria optimization, automating the supplier selection process while ensuring adaptability and alignment with strategic corporate goals.

To evaluate these ML models' effectiveness, we implemented a structured methodology encompassing rigorous data preprocessing and exploratory data analysis (EDA). Using a synthesized dataset representative of real-world supplier performance metrics, we employed k-fold cross-validation and a suite of performance metrics, including precision, recall, F1-score, and ROC-AUC. Our analysis not only explored predictive accuracy and computational efficiency but also considered the interpretability and feature significance through feature importance analysis.

Our findings reveal XGBoost's superior efficiency in training times without compromising prediction parity, accompanied by higher performance scores across accuracy, precision, recall, and F1 measure. These insights offer a nuanced understanding of how ML can fortify supplier engagements and enhance procurement strategies, providing supply chain professionals with a comparative analysis to select the most apt ML approach tailored to their specific requirements.

This article contributes to both academic research and practical application by: elucidating the complexities of ensemble learning methodologies, highlighting data-driven decision-making intricacies, and aiding practitioners in ML model selection for supplier evaluation. The comprehensive analysis presented in this study lays the groundwork for future research and the ongoing digital transformation in supply chain management.

**Table of Contents**

1. **Introduction -------------------------------------------------------------------- 6**
2. **Background -------------------------------------------------------------------- 8**

* What is supply chain management
* Supplier Selection
* Supplier Selection Process
* What is ML

1. **Literature Review --------------------------------------------------------------- 11**

* How can machine learning help
* Applied approaches
* Aim and Scope of the Article
* Methodological Approach
* Contribution of the Article

1. **Methodology ----------------------------------------------------------------- 15**

* Random Forest
* XGBoost
* Comparative Analysis
* Data Preprocessing and Exploratory Data Analysis (EDA)
* Model Evaluation Criterion
* Performance Metrics
* Feature Importance Analysis

1. **Comparative Analysis ------------------------------------------------------ 21**

* Comparison of Model Training Times
* Comparison of Model Prediction Times
* Mean Performance Score Comparison

1. **Conclusion ------------------------------------------------------------------ 24**
2. **Future Work ------------------------------------------------------------------- 25**
3. **References --------------------------------------------------------------------- 26**

**Table of Figures**

* Fig 1. Supplier Selection Steps -------------------------------------------- 8
* Fig 2. Supplier Selection Process ---------------------------------------- 9
* Fig 3. Dataset Information ------------------------------------------------ 16
* Fig 4. Correlation Matrix --------------------------------------------------- 17
* Fig 5. Distribution of Method of Payment ----------------------------- 18
* Fig 6. Distribution of Conditions ----------------------------------------- 18
* Fig 7. Distribution of Quality ---------------------------------------------- 19
* Fig 8. Distribution of Supplier Location -------------------------------- 19
* Fig 9. Distribution of Financial Condition ------------------------------ 19
* Fig 10. Distribution of Serviceability ------------------------------------ 20
* Fig 11. Comparison of Model Training Times ------------------------ 21
* Fig 12. Comparison of Model Prediction Times --------------------- 22
* Fig 13. Mean Performance Score Comparison --------------------- 23

**Introduction**

In the realm of supply chain management, the process of selecting suppliers stands as a pivotal element that profoundly influences the operational efficiency and overall success of businesses. Traditional approaches to supplier selection often involve manual assessments, a methodology susceptible to subjectivity, time-consuming evaluations, and inherent biases. In recent times, the integration of machine learning (ML) has emerged as a transformative force, reshaping how organizations identify and engage with suppliers.

The utilization of machine learning algorithms empowers businesses to harness historical data, supplier performance metrics, and diverse factors for making well-informed decisions regarding potential suppliers. This evolution in technology not only enhances accuracy but also accelerates the evaluation process, enabling organizations to analyse extensive datasets efficiently. The incorporation of machine learning into supplier selection processes provides businesses with the flexibility to adapt to dynamic market conditions, aligning strategic decisions with specific requirements and overarching objectives.

Key Benefits of Implementing Machine Learning in Supplier Selection:

**Data-Driven Decision-Making:**

The cornerstone of machine learning in supplier selection lies in its ability to conduct thorough analyses of large datasets, supplier profiles, and performance metrics. This data-driven methodology ensures that decisions are grounded in comprehensive and unbiased evaluations, minimizing reliance on subjective judgment.

**Predictive Analytics:**

Machine learning models excel in predicting supplier performance, reliability, and potential risks based on historical patterns and emerging trends. This predictive capability empowers organizations to anticipate and address potential issues proactively, facilitating superior risk management and mitigation strategies.

**Efficiency and Automation:**

The automation of supplier selection processes through machine learning streamlines the manual workload associated with evaluating numerous suppliers. This heightened efficiency leads to quicker decision-making, enabling businesses to respond promptly to dynamic market changes and evolving demands.

**Customization and Flexibility:**

Machine learning models can be tailored to consider specific criteria, industry standards, and organizational priorities. This adaptability ensures that supplier selection processes align seamlessly with the unique needs and strategic goals of the business.

**Continuous Improvement:**

The learning capabilities inherent in ML algorithms contribute to continuous improvement. These algorithms evolve over time, assimilating new data and insights to enhance accuracy and relevancy in supplier selection models.

**Cost Savings:**

The efficient application of machine learning in supplier selection can result in cost savings through optimized procurement processes, improved negotiation strategies, and minimized impact from supply chain disruptions.

The incorporation of machine learning into supplier selection processes signifies a monumental advancement in procurement strategies. By leveraging data-driven decision-making, predictive analytics, and automation, businesses can fortify their competitive edge, mitigate risks, and build resilient and efficient supply chains. As organizations embrace digital transformation, the role of machine learning in supplier selection is poised to become even more influential in shaping the future landscape of supply chain management.

**Background:**

**What is supply chain management?**

Supply chain management is the coordination of goods , information, and finances from suppliers to consumers, integrating procurement, production, distribution, and logistics to move products effectively and efficiently. A review of 309 academic papers shows a focus on quantitative models for managing sustainable supply chains, particularly environmental aspects, with limited research on the social  
dimensions of sustainability. The review calls for more empirical work and a balanced approach to all sustainability facets in future supply chain management research.(Croom et al., 2000)

**Supplier Selection:**

Choosing the right supplier is a crucial decision in the field of supply chain management, significantly impacting a company's competitiveness. The selection process involves evaluating various alternative suppliers based on a range of criteria, encompassing both qualitative and quantitative aspects. Supplier selection is identified as a key component in supply chain management, recognized as a multiple criteria decision-making (MCDM) problem. Effectively choosing a supplier can lead to cost reduction, improved profits, shorter product lead times, heightened customer satisfaction, and increased competitiveness. Despite its importance, there is no standardized approach to supplier selection, emphasizing the need for a context-specific application. Making the wrong choice can result in substantial losses for the supply chain and directly affect overall company performance. In today's dynamic environment, where supplier selection criteria evolve, the process involves three significant steps: criteria identification, questionnaire survey study with criteria analysis and weight determination, and implementation of a multi-criteria decision-making method to select the optimal supplier. The emphasis has shifted from price alone to considering factors such as quality, delivery performance, cost, and capability based on the specific purchasing situation. .(Taherdoost & Brard, 2019)

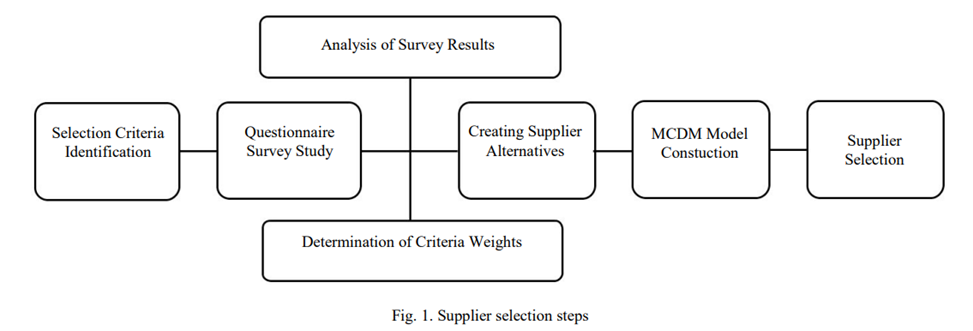


Fig1. Supplier Selection Steps (Taherdoost & Brard, 2019)

**Supplier Selection Process:**

The supplier selection process, as outlined by (Taherdoost & Brard, 2019) .is a critical organizational function overseen by the procurement department. The purchasing manager holds the responsibility of developing and implementing an effective process to identify qualified suppliers for business awards.this process demands a significant commitment of resources. Weele further highlights that the supplier selection process is a vital component of the overall purchasing process, commencing with market research after defining functional or technical specifications. The process comprises four key steps, detailed below and illustrated in Figure 2:

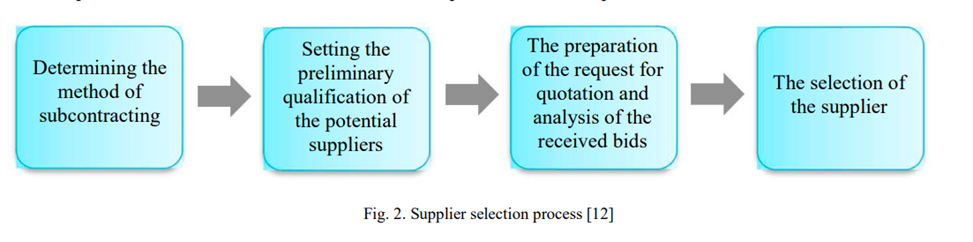


Fig2. Supplier Selection process (Taherdoost & Brard, 2019)

**A. Determining Subcontracting Method:**

The initial step involves deciding whether to opt for turnkey (performing the entire assignment) or partial subcontracting. This decision also includes determining the pricing method before awarding the work.

**B. Setting Preliminary Qualifications and Bidder’s List:**

Establishing the preliminary qualifications of potential suppliers and creating a list of bidders are crucial steps in the supplier selection process.

**C. Request for Quotation Preparation and Bid Analysis:**

This step involves preparing the request for quotation and subsequently analyzing the received bids to make informed decisions.

**D. Supplier Selection:**

The selection of the supplier stands out as the most crucial step in the purchasing process, serving as the foundation for various other activities. In situations where the number of approved suppliers is insufficient, conducting thorough supply market research becomes imperative for discovering new suppliers.

**What is ML?**

Machine learning is a subset of artificial intelligence that involves the development of algorithms and statistical models that enable computers to perform tasks without explicit instructions, relying instead on patterns and inference. It is categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning uses labeled data to predict outcomes, commonly applied in classification and regression tasks. Unsupervised learning, on the other hand, deals with unlabeled data, helping to discover the underlying structure of the data for clustering, compression, and feature extraction. Semi-supervised learning is a mix of the two, using both labeled and unlabeled data to improve learning accuracy.

Reinforcement learning is distinct in that it focuses on making sequences of decisions by interacting with a dynamic environment to achieve a goal, learning from the consequences of actions rather than from direct data labeling. It is particularly useful in fields such as robotics, gaming, and navigation, where the system must make a series of decisions that lead to a cumulative reward.

The document you provided discusses the recent progress in machine learning, particularly deep learning and reinforcement learning, and their implications for process systems engineering. Deep learning, a subset of machine learning, involves training neural networks with multiple layers to perform hierarchical feature learning, which has shown significant success in areas like image and speech recognition. Reinforcement learning has been highlighted through the success of AlphaGo, which uses a combination of Monte Carlo tree search and deep neural networks to play the game of Go at a high level of proficiency. (Lee et al., 2018)

**Literature Review**

**How can machine learning help:**

Machine Learning (ML) can significantly enhance the supplier selection process by providing advanced data analysis capabilities. Here is how ML can help, referencing the paper provided:

**Performance Evaluation:** ML can aid in the analysis of supplier performance by assessing historical data on quality, delivery, and cost, thus predicting future performance trends (Hsu & Hu, 2009)

**Sustainability Scoring:** Algorithms can quantify suppliers’ environmental and social performance by examining factors like carbon emissions, waste management, and labor practices, which aligns with the sustainability criteria emphasized in the paper (Sarkis &Dhavale, 2015)

**Risk Assessment**: ML models like Decision Trees or Neural Networks can help in identifying high-risk suppliers by correlating factors such as location, financial stability, and market dynamics with risk levels.

**Optimization of Selection Criteria:** By analysing diverse datasets, ML can facilitate the identification of the most impactful criteria for supplier selection, which helps in the model development for evaluating sustainability performance as suggested by (Govindan et al.,2013).

**Facilitating Green Supply Chain:** ML techniques can facilitate the development of a green supply chain by ranking suppliers based on their adherence to ecological and sustainable practices (Kumar et al., 2016).

Using ML for supplier selection not only enhances the accuracy of the selection process but also helps in creating a supply chain that is resilient, sustainable, and aligned with the company's goals and regulatory demands. It streamlines the process, ensuring that complex data is manageable and the insights derived from it are actionable, as has been advocated in the research conducted by various scholars outlined within the provided text.

**Applied approaches:**

Leveraging contemporary methodologies to enhance supplier selection processes, a groundbreaking study by (Ali et al., 2023) examined the application of machine learning to classify supplier evaluation criteria. Their research elaborated on a decision support system that uses Random Forest, an ensemble learning method, to assess and rank suppliers by importance based on various selection criteria.

The Random Forest algorithm, known for its generalizing capabilities, synthesizes predictions from multiple decision trees to procure a more stable and accurate classification. Ali et al.'s study not only identifies a comprehensive list of important criteria but also validates their applicability across diverse organizational contexts. This approach exemplifies the successful integration of machine learning algorithms into operational practices, markedly improving accuracy and f-score by 3.89% and 5.17%, respectively, after refining the criteria set.

Their findings point to Quality, On-Time Delivery, Material Price, and Information Sharing as top-tier criteria, while underlining Transportation Cost as a previously undervalued yet crucial factor. The application of this system enables managers to streamline the supplier selection process by prioritizing key performance measures, supporting the development of a resilient and efficient supply chain.

In conjunction with the decision support methodology highlighted by (Abdulla et al., 2023) , another pivotal study by Abdulla, Baryannis, and Badi (2023) introduces an innovative method that synergizes machine learning with the MARCOS method for supplier evaluation. The integrated system capitalizes on the interpretability of tree-based machine learning algorithms to deduce feature importances, subsequently assimilating these as weights within the MARCOS (Measurement of Alternatives and Ranking according to COmpromise Solution) framework.

This sophisticated amalgamation addresses common multi-criteria decision-making (MCDM) complications, specifically the challenge of assigning precise evaluation criteria weights. Abdulla and colleagues’ research advances the field by offering a dynamic method that adjusts criteria weights in real-time, enhancing the precision of the supplier selection process.

Employing this novel method within the pragmatic context of the oil and gas industry, the study shows that the model aligns well with selections made by human experts and performs comparably to existing multi-criteria decision-making procedures. The validated model is a testament to the benefits of integrating machine learning insights with traditional MCDM approaches, potentially revolutionizing supplier evaluation processes for more effective and efficient outcomes.

**Aim and Scope of the Article:**

In the dynamic landscape of supply chain management, selecting the right suppliers is critical for ensuring the quality, cost-efficiency, and reliability of products and services. The aim of this article is to illuminate the application of advanced machine learning techniques for enhancing supplier selection processes.

Specifically, we endeavor to Present a comparative study between two powerful ranking-based models—Random Forest and XGBoost—and their applicability in evaluating and selecting suppliers.

Examine the effectiveness of these models in understanding complex, multidimensional supplier data and providing a quantitative basis for comparison.

Illustrate the process of employing these methods from data preprocessing to final ranking output and decision-making.

**Methodological Approach**

Our methodology takes a structured approach to implementing Random Forest and XGBoost models on a dataset meticulously constructed by a language model (LLM) to mirror real-world supplier metrics. The objective of this comparison is to evaluate the models’ accuracy in supplier ranking, as well as their interpretability, scalability, and responsiveness to feature variations.

To ensure a solid foundation for our analysis, we will explore the theoretical frameworks of each algorithm. This allows for a thorough comprehension of their intrinsic mechanisms, facilitating the application of such theories to the empirical dataset. The dataset, notably synthesized by the LLM to reflect realistic supplier scenarios, will be subjected to stringent cleaning and preparation. Such due diligence ensures that our models are trained on data that genuinely encapsulates authentic supplier performance and attributes.

Subsequently, we will undertake a methodical routine of feature engineering, model training and refinement, and assessment of model effectiveness. These procedures will involve a suite of metrics tailored to address the nuanced demands of the algorithm’s classification capabilities and the intended output of supplier ranking.

In the quest for optimal performance, leveraging the linguistic prowess of LLM can be theorized as follows: by further enriching feature extraction through advanced text analysis and semantic comprehension, the LLM can potentially uncover subtleties and insights from qualitative data such as customer reviews or supplier descriptions. These nuances may otherwise remain elusive to conventional analysis. The enriched features, theoretically, could enable the models to capture a more nuanced representation of supplier profiles, thereby potentially enhancing prediction outcomes and the utility of the rankings.

**Contribution of the Article**

This article seeks to enhance both academic knowledge and pragmatic approaches to supplier selection by accomplishing the following objectives:

1. Illustrating the intricacies of two prominent methodologies—Random Forest and XGBoost—within the ensemble learning framework.

2. Shedding light on the data-driven intricacies that underpin decision-making in supply chain management.

3. Providing a comparative examination that empowers practitioners to identify the most suited machine learning technique for their unique supplier assessment needs.

Ultimately, the insights yielded by this research are intended to direct supply chain professionals towards adopting advanced analytical methods, thereby fortifying supplier engagements and refining procurement processes.

**Methodology**

**Random Forest**

The Random Forest algorithm, introduced by Leo Breiman in 2001, represents an ensemble learning method typically used for classification and regression tasks. The model’s essence lies in constructing a multitude of decision trees during training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees (Breiman, 2001).Such an ensemble approach effectively reduces the risk of overfitting, a common drawback of single decision trees, hence improving the model’s generalizability to new data.

With its built-in feature selection mechanism and adaptability, Random Forest can handle a wide array of data types and relationships. A key advantage of Random Forest is its capability to rank the importance of variables in a regression or classification problem in a natural way (Gini importance), which simplifies the model interpretation and understanding of the data (Breiman, 2001).

**XGBoost**

On the other hand, XGBoost (Extreme Gradient Boosting) is an efficient and scalable implementation of gradient boosting machines, a class of algorithms that have proven successful in a wide range of practical applications. As described by Chen and Guestrin (2016), XGBoost significantly improves over traditional gradient boosting in terms of computation speed and model performance. This is achieved through several algorithmic enhancements, such as a novel tree learning algorithm that efficiently handles sparse data and an effective regularization term to prevent overfitting (Chen & Guestrin, 2016).

XGBoost also offers several features that make it robust, such as built-in cross-validation, missing value handling, and the ability to optimize on a custom loss function. Furthermore, it leverages numerous hardware optimizations and highly efficient data structures to achieve state-of-the-art results on many machine learning benchmarks (Chen &Guestrin, 2016).

**Data Preprocessing and Exploratory Data Analysis (EDA):**

The dataset, replete with attributes like supplier quality, quantity, conditions, payment methods, and various qualitative factors, underwent rigorous preprocessing. Categorical variables were transformed via one-hot encoding, while numerical variables were standardized, ensuring a uniform scale across features.

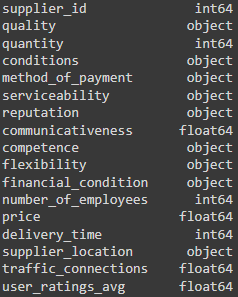
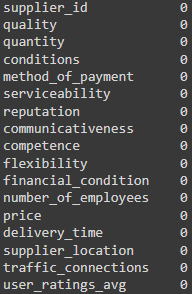
 

Fig3. dataset information

In the initial phase of our study, we conducted thorough data preprocessing to ensure the quality and consistency of the dataset, which encompassed diverse measures such as supplier quality, quantity, conditions, payment methods, and various qualitative descriptors. Categorical variables underwent one-hot encoding to facilitate their incorporation into the machine learning models. Numerical variables were standardized to a unified scale to avoid any dominance of large-value variables over small-value ones, ensuring an unbiased input for model training.

During the exploratory data analysis, we constructed a correlation matrix to discern the association between the user average rating and other supplier performance metrics. This analysis proved insightful for feature selection and understanding the dynamics within the data.

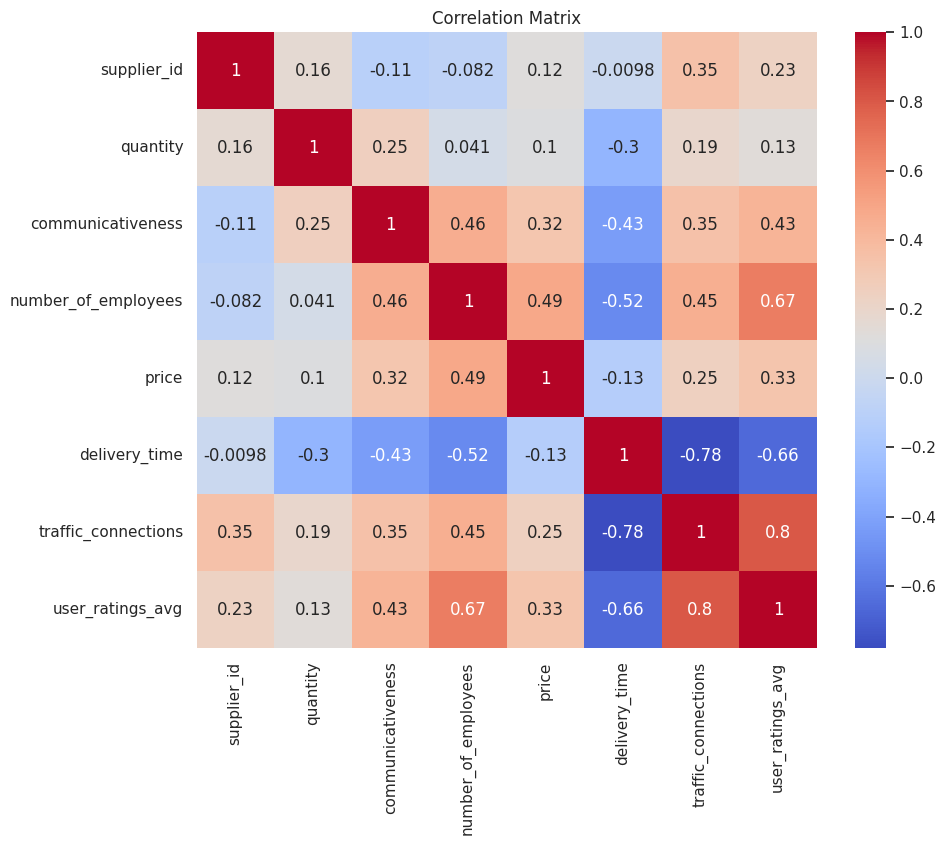


Fig4. Correlation Matrix

Key findings from our correlation analysis were as follows:

* **User Rating Average vs. Quantity:**A correlation coefficient of 0.13 nudged us to consider the subtle relationship between the scale of supplier operations and their perceived quality.
* **User Rating Average vs. Communicativeness:**A more pronounced correlation of 0.43 highlighted the value users place on supplier interactions, signalling a noteworthy predictor in the user rating milieu**.**
* **User Rating Average vs. Number of Employees:**The robust 0.67 coefficient underscored users’ preference for larger, possibly more robust, enterprise structures.
* **User Rating Average vs. Price:**With a positive 0.33 correlation, price emerged as a moderate yet tangible influencer of user ratings.
* **User Rating Average vs. Delivery Time:**A notable negative correlation of -0.66 keenly pointed to the criticality of swift delivery in achieving higher user satisfaction.
* **User Rating Average vs. Traffic Connection:**Standing out with a robust correlation of 0.8, the ease of traffic connection appeared as a dominant driver in user ratings, nuancing our model development priorities.

Visualizing the distribution of selected features:

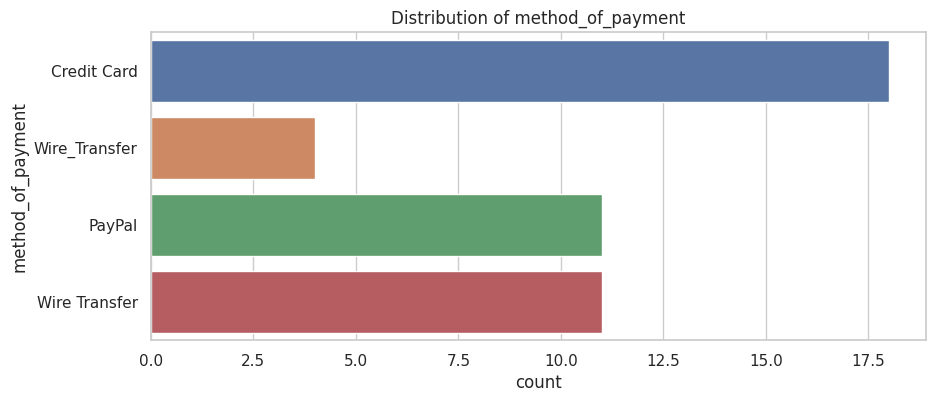


Fig5. Distribution of method\_of\_payment

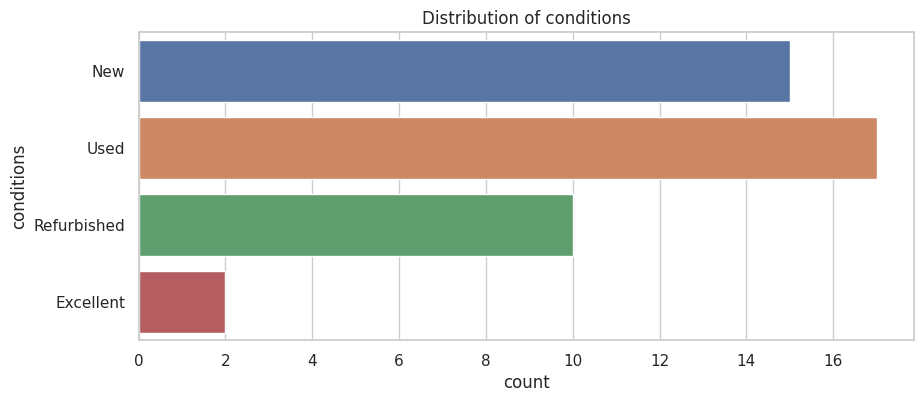


Fig6. Distribution of conditions

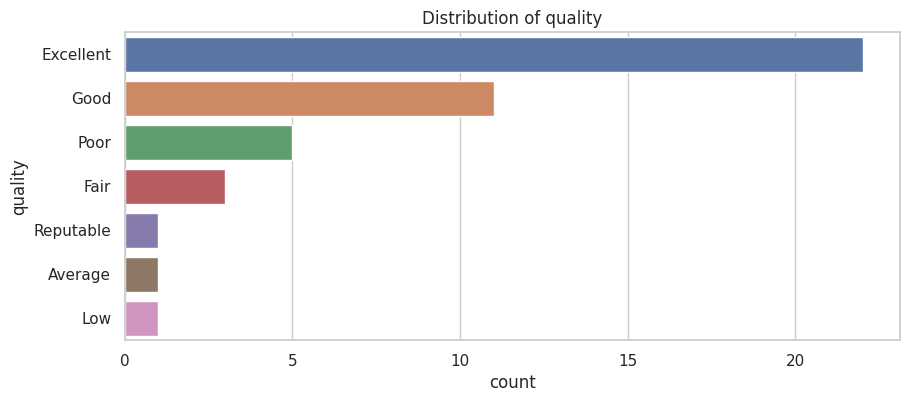


Fig7. Distribution of quality

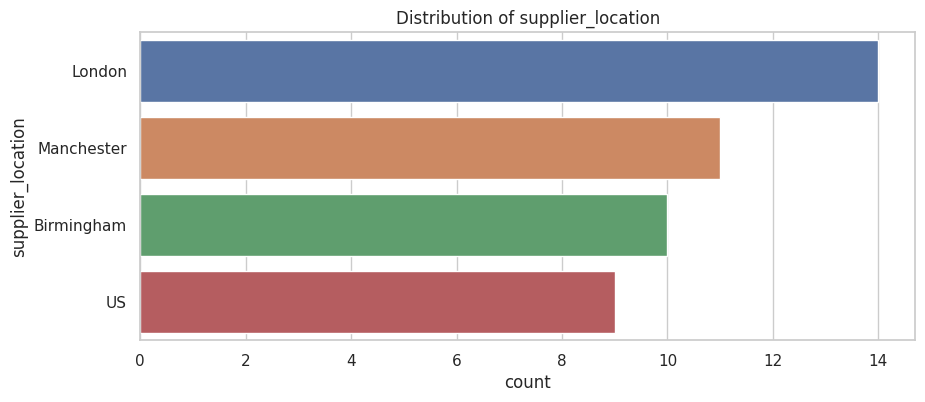


Fig8. Distribution of supplier\_location

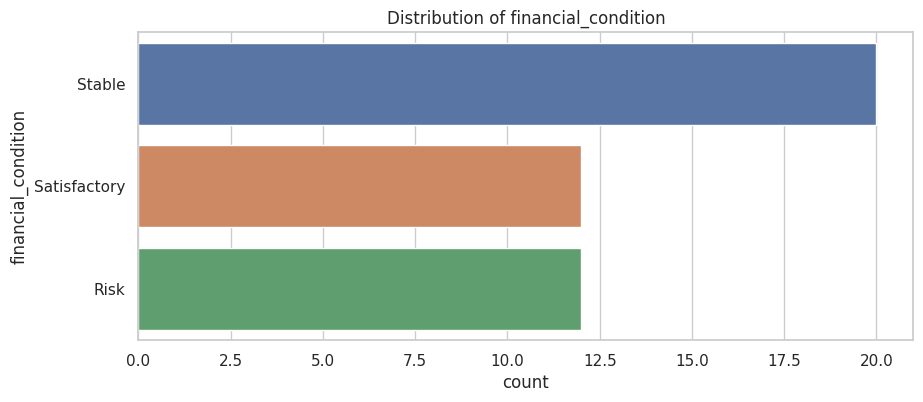


Fig9. Distribution of financial\_condition

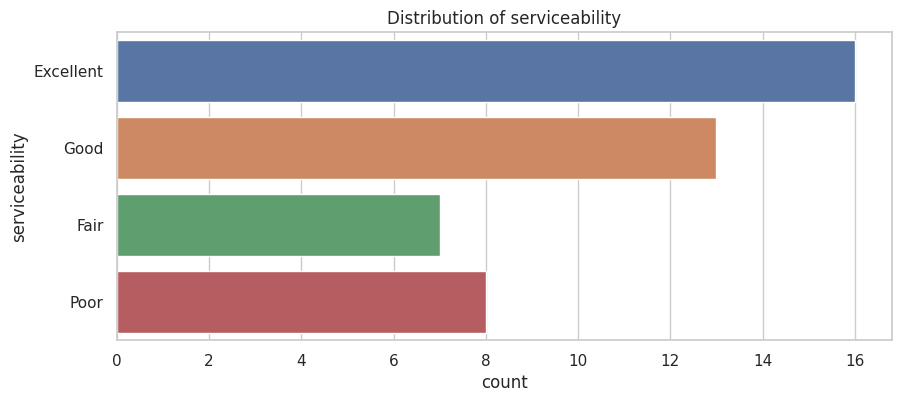


Fig10. Distribution of serviceability

These visualizations of feature distributions supplied a multi-angled perspective of the dataset that was instrumental to decipher the complex fabric of supplier evaluations; an understanding that bare numbers alone could hardly yield.

Incorporating these distribution observations, we tailored our data preprocessing approaches, such as feature scaling and transformation, to enhance the ensuing models’ predictive grounding. These calibrations were pivotal to generating models that are not only high-performing on the sanitized dataset but also robust and reliable when extrapolated to real-world scenarios.

By harnessing the power of our EDA discoveries, we ensured that the preparatory work—grounded in empirical data patterns—would feed into the feature engineering and model training processes. This foundational work crystallized into a well-informed blueprint for the remainder of our study, advancing the comparative analysis and optimization of the Random Forest and XGBoost models against a backdrop of rigorous evidence-based scrutiny.

**Comparative Analysis**

Considering the unique architectures and operations of Random Forest and XGBoost, our research aims to extensively compare their effectiveness in predicting supplier performance. We have methodically measured the comparative proficiency of these models through the following structured process:

**Model Evaluation Criterion:** Our comparative study employed the robust k-fold cross-validation technique, partitioning the data into ‘k’ segments for cyclic training and validating processes. This method provides a flexible and less biased estimate of model performance.

**Performance Metrics:** We benchmarked both models using standard performance indicators, such as precision, recall, F1-score, and ROC-AUC. These metrics provided a comprehensive view of each model’s accuracy, sensitivity, and specificity.

**Feature Importance Analysis:** Beyond performance metrics, the models’ interpretability was scrutinized by evaluating feature importance scores. We aimed to identify the most influential predictors that drive supplier performance forecasts.

Adding to the analysis, our findings incorporated a comparative perspective on computational efficiency, as follows:

**Comparison of Model Training Times:**

Results showed that Random Forest had a training time of 0.14 compared to XGBoost’s more efficient 0.04, indicating XGBoost’s superior speed in model training.

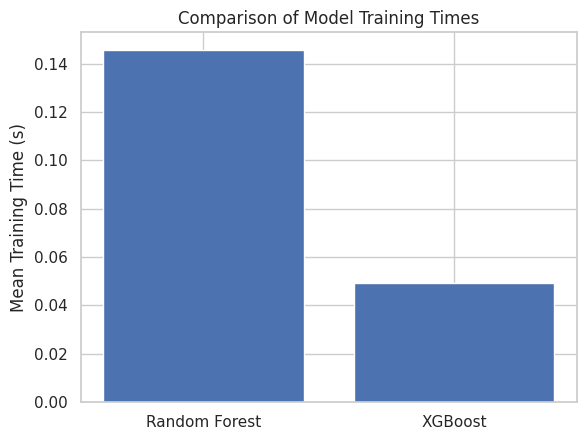
****

Fig11. Comparison of Model Training Time

**Comparison of Model Prediction Times:**

Random Forest and XGBoost demonstrated parity in prediction times, suggesting a tie in this aspect.

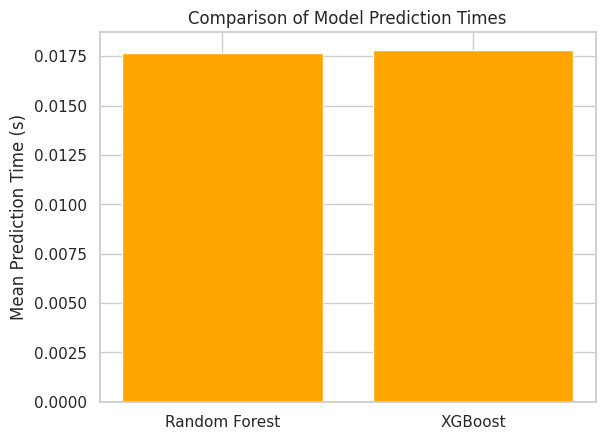


Fig12. Comparison of Model Prediction Times

.

**Mean Performance Score Comparison:**

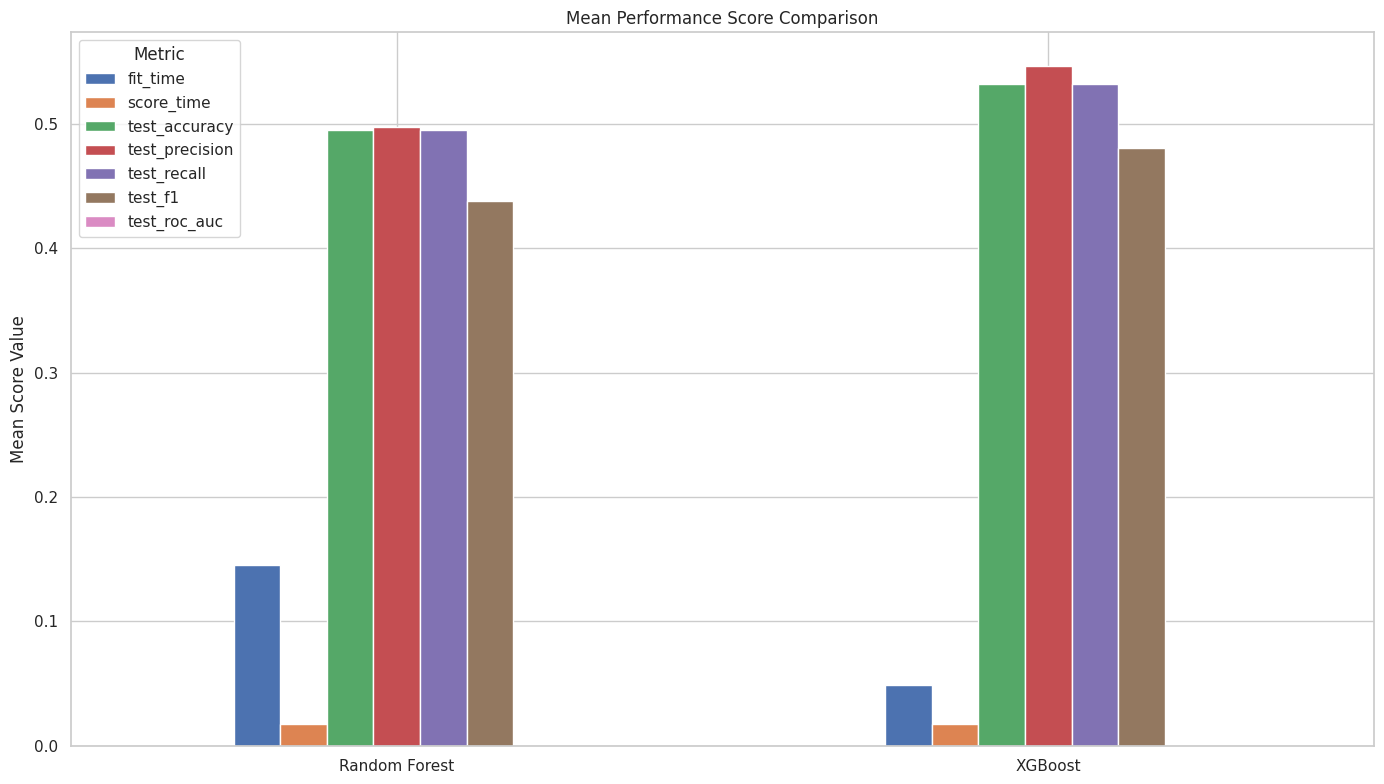
****

Fig13. Mean Performance Score Comparison

* Fit Time: Random Forest exhibited a longer fit time than XGBoost, suggesting slower model adaptability in training.
* Score Time: Equal score times for Random Forest and XGBoost indicated similar efficiency in model scoring.
* Test Accuracy: XGBoost outperformed Random Forest, indicating higher accuracy in predictions.
* Test Precision: XGBoost achieved a higher precision score, suggesting a better identification of true positive cases.
* Test Recall: XGBoost again led the way with better recall scores, accurately identifying a higher percentage of actual positive cases.
* Test F1 Score: XGBoost rounded out its performance advantage with higher F1 scores, signaling a better balance between precision and recall.

These results are indicative of the strengths and weaknesses inherent to each model. The data point towards XGBoost offering a more robust performance, particularly in terms of precision and accuracy, and efficiency in model training—while maintaining comparable prediction times to Random Forest. This comparison is anticipated to guide practitioners in making an informed choice about which machine learning model to utilize for supplier performance prediction, balancing the need for accuracy with computational resources.

**Conclusion**

In conclusion, our investigation into the application of machine learning for supplier selection has yielded significant insights and actionable results. The meticulous data preprocessing and exploratory data analysis (EDA), which included the standardization of numerical variables and one-hot encoding of categorical ones, laid a strong foundation for building robust predictive models. The EDA in particular provided us with a deep understanding of the relationships between various supplier attributes and ratings, highlighting key drivers such as communicativeness, delivery time, and traffic connection which could significantly influence supplier evaluations.

Our comparative study of Random Forest and XGBoost models unveiled clear differences in performance and computational efficiency. The faster model training times and superior prediction metrics of XGBoost point towards its overall efficacy in handling the task of supplier performance prediction. However, the parallel of prediction times between the two models reaffirms that Random Forest remains a competitive option, especially when model interpretability and feature importance are pivotal to the analysis.

The extensive feature importance analysis refined our understanding of which variables most significantly impact supplier ratings, allowing us to fine-tune the models for better predictive accuracy. These tools and methodologies could form the cornerstone for developing a sophisticated recommender system. Using similar techniques, businesses could integrate supplier performance data, user ratings, and qualitative feedback to construct a system that not only ranks suppliers but also recommends the most suitable suppliers to companies based on their unique requirements and past performance metrics.

**Future Work**

As we explore the potential for applying advanced machine learning models to improve supplier ranking and selection, future work could expand on several fronts. One area is the continued enhancement of the predictive models by integrating multi-modal data sources, such as sentiment analysis from user text reviews or real-time market fluctuations. The ability of language models (like LLM) to parse and understand customer feedback can offer deeper qualitative insights, complementing the quantitative data and potentially improving the accuracy of supplier ratings.

Moreover, developing a recommender system that leverages machine learning could streamline the supplier selection process significantly. By adapting the methodology used in our comparative analysis, such an advanced system could take into account user ratings, textual feedback through LLM analysis, and historical performance data to provide personalized supplier recommendations to businesses. This system could utilize XGBoost for its predictive power or Random Forest for its interpretability, depending on the application context.

Overall, the incorporation of machine learning into supplier selection exemplifies how technology can enhance decision-making processes, making them more data-driven, efficient, and tailored to specific business goals. The promising outcomes of this study pave the way for innovative solutions in supply chain management, opening avenues for continuous improvement and strategic alignment in the domain of supplier selection and evaluation.

**References**

Abdulla, A., Baryannis, G., and Badi, I. An integrated machine learning and marcos method forsupplier evaluation and selection. Decision Analytics Journal, 9:100342, 2023. ISSN 2772-6622.doi: https://doi.org/10.1016/j.dajour.2023.100342.URL https://www.sciencedirect.com/science/article/pii/S2772662223001820.

Ali, M. R., Nipu, S. M. A., and Khan, S. A. A decision support system for classifying supplier selection criteria using machine learning and random forest approach.Decision Analytics Journal, 7:100238, 2023. ISSN 2772-6622. doi: https://doi.org/10.1016/j.dajour.2023.100238.URL https://www.sciencedirect.com/science/article/pii/S2772662223000784.

Breiman, L. Random forests. Machine Learning, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. pp. 785–794, 2016. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785.

Croom, S., Romano, P., and Giannakis, M. Supply chain management: an analytical framework for critical literature review. European Journal of Purchasing Supply Management, 6(1):67–83, 2000. ISSN 0969-7012. doi: <https://doi.org/10.1016/S0969-7012(99)00030-1>. URL https://www.sciencedirect.com/science/article/pii/S0969701299000301.

Govindan, K., Khodaverdi, R., and Jafarian, A. A fuzzy multi criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach. Journal of Cleaner Production, 47:345–354, 2013. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2012.04.014.URL <https://www.sciencedirect.com/science/article/pii/S0959652612002016>.

Kumar, V., & Reinartz, W. (2016). Creating Enduring Customer Value. Journal of Marketing, 80(6), 36-68. https://doi.org/10.1509/jm.15.0414

Cleaner Production: initiatives and challenges for a sustainable world. Hsu, C.-W. and Hu, A. H. Applying hazardous substance management to supplier selection using analytic network process. Journal of Cleaner Production, 17(2):255–264, 2009. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2008.05.004. URL <https://www.sciencedirect.com/science/article/pii/S0959652608001431>.

Lee, J. H., Shin, J., and Realff, M. J. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. Computers Chemical Engineering, 114:111–121, 2018. ISSN 0098-1354. doi:https://doi.org/10.1016/j.compchemeng.2017.10.008.URL https://www.sciencedirect.com/science/article/pii/S0098135417303538.

FOCAPO/CPC 2017.Manohar, H. L. and Kumar, R. G. Impact of green supply chain management attributes on sustainable supply chains. International Journal of Supply Chain and Operations Resilience, 2(4):291–314,2016. doi:10.1504/IJSCOR.2016.084030. URLhttps://www.inderscienceonline.com/doi/abs/10.1504/IJSCOR.2016.084030.

Sarkis, J. and Dhavale, D. G. Supplier selection for sustainable operations: A triple-bottom-line approach using a bayesian framework. International Journal of Production Economics, 166:177–191, 2015. ISSN 0925-5273. doi: <https://doi.org/10.1016/j.ijpe.2014.11>. 007. URL <https://www.sciencedirect.com/science/article/pii/S0925527314003570>.

Taherdoost, H. and Brard, A. Analyzing the process of supplier selection criteria and methods. Procedia Manufacturing, 32:1024–1034, 2019. ISSN 2351-9789. doi: <https://doi.org/10.1016/j.promfg.2019.02.317>. URL <https://www.sciencedirect.com/science/article/pii/S2351978919303555>. 12th International Conference Interdisciplinarity in Engineering, INTER-ENG 2018, 4–5 October 2018,Tirgu Mures, Romania.