SUPPLIER PERFORMANCE EVALUATION PREDICTIVE MODEL FOR DIRECT MATERIAL USING MACHINE LEARNING APPROACH IN SEMICONDUCTOR MANUFACTURING

S.H. Yee[1,2], S.A. Asmai[1*], Z. Abal Abas[1], S. Ahmad[1], A.S. Shibghatullah[3], D. Petrovic[4]

[1]Fakulti Teknologi Maklumat dan Komunikasi,
Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.

[2]STMicroelectronics Sdn. Bhd., Tanjong Agas Industrial Area P.O. Box 28 Muar, Johor, 84007 Malaysia.

[3]College of Computing & Informatics (CCI), Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia.

[4]Nottingham Business School, Nottingham Trent University, Nottingham, NG1 4FQ, United Kingdom.

*Corresponding Author's Email: azirah@utem.edu.my

Article History: Received 7 January 2024; Revised 16 June 2024; Accepted 4 July 2024

ABSTRACT: In semiconductor manufacturing, evaluating supplier performance for direct materials is often unreliable and biased, failing to accurately represent suppliers' true performance. The objective of this paper is to present a data-driven Supplier Performance Evaluation (SPE) predictive model for direct material in semiconductor manufacturing. By using multiple machine learning techniques, the model provides unbiased evaluations of supplier performance. The model uses six machine learning methods: Logistic Regression, Support Vector Machine, Naïve Bayes, Generalized Linear Model, Decision Tree, and Random Forest. . The results show that Logistic Regression outperforms the other techniques with regards to analyzing both data from incoming material checks and the assembly in-process. The AUC-ROC value is 0.993 from Logistic Regression, proving that the model can identify material withdrawal trends effectively. In conclusion, the resulting model can enhance monitoring, risk management, and proactive supplier management, which leads to an efficient supply chain.

KEYWORDS: *Supplier Performance Evaluation; Supply Chain Management; Semiconductor Manufacturing; Machine Learning, Logistic Regression.*

1.0 INTRODUCTION

Supplier Performance Evaluation (SPE) is crucial to determine supplier performance with regards to complying with contract specifications of product, and service level agreements as well as achieving Key Performance Indicators (KPIs). It can assist the organization to fulfill its objectives by establishing explicit requirements for suppliers and promoting transparency, so enabling suppliers to comprehend and pursue excellence. SPE enhances supplier quality and experience by the monitoring and evaluation of corrective actions taken by providers, including response times to complaints. This leads to a reduction in unnecessary costs through better delivery performance and tracking of product quality. Supplier Performance Evaluation (SPE) offers material providers valuable insights into operational efficiency, capacity for growth, and chances for optimization in several areas such as production schedules, technical issues, supply chains, and quality management [1].

Efficient supply chain management is crucial in semiconductor manufacture to fulfil client requirements. Although the industry has experienced substantial expansion, there is a lack of comprehensive quantitative data of SPE Conventional survey-based SPE methodologies can lead to selection bias, which can have an impact on the reputation of suppliers and the decisions made about source selection [2]. To address these challenges, the research proposes a new SPE model that functions continually, offering more unbiased and dependable data. This methodology enhances the alignment with modern data analytics, leading to a more accurate and unbiased assessment of suppliers.

Paper remains arranged as follows: Section 2 covers literature on classical SPE, data-driven SPE, and supplier selection framework.

Section 3 describes methodology for studies. We discuss the results in Section 4. Section 5 presents SPE predictive analysis's key findings, draws conclusions, and offers practical advice for future study and practice.

2.0 RELATED STUDY

Supplier Performance Evaluation (SPE) is crucial for cost reduction and production optimization, as it monitors delivery performance and product quality. It provides insights, growth opportunities, and risk mitigation, essential for successful supply chain management. However, traditional SPE faces challenges like data instability and selection bias, leading to unreliable evaluations [2]. Inaccurate SPEs can harm supplier reputations and influence sourcing negatively. Low response rates due to spam-flagged surveys further hinder the process [2]. Research also examines SPE communication from signaling theory's perspective, emphasizing information sharing and its impact on buyer-supplier relationships [3]. Subjective judgment dominates, causing variation in evaluations [4].

Therefore, reliable SPE tools are vital for supply chain success, and extensive research has explored their development methodologies. Methodologies such as Multi-Criteria Decision Making (MCDM) and Data Envelopment Analysis (DEA) are used to measure efficiency [5]. Khan et al. [6] propose an integrated model that combines MCDM, Fuzzy-Shannon Entropy (SE), and Fuzzy-Inference System (IS) to address conflicts in supplier selection criteria. Petrovic et al. [7] proposed new concept optimization model based on Fuzzy scenario in Supply Network. Liou et al. [8] suggest a hybrid approach integrating MCDM, SVM, FBWM, and FTOPSIS for practical and reliable evaluations, despite limitations like averaging multi-year data. Comprehensive empirical investigations are needed to validate these frameworks [6][8][9].

Supplier evaluation, which includes considerations of cost, quality, and delivery time [10], faces challenges due to the sensitivity of past performance data [2]. Emerging technologies like machine learning mitigate decision-making uncertainties [11], but sorting and analyzing big data present hurdles [7]. The development of data-driven Supplier Performance Evaluation (SPE) model holds promise for enhancing supplier management practices [12]. Additionally, the evolution of data science has fostered various tools for big data analytics, facilitating multi-criteria decision making in supplier evaluation [13]. By incorporating real-time manufacturing data, machine learning models offer objective supplier assessments, reflecting genuine operational impacts. Descriptive analytics provide retrospective insights, while predictive modeling forecasts future outcomes, benefiting sectors such as supply chain management and healthcare [11][14]. Machine learning's crucial role in industrial applications, particularly in production planning and operations management, is evident in recent literature [15].

Recent years have seen increased use of logistic regression (LR) in SPE. LR, a statistical and machine learning technique, is adept at binary classification, predicting the probability of a binary outcome based on predictor variables. Unlike linear regression, LR outputs probabilities mapped to discrete classes, making it ideal for event likelihood predictions like email spam detection or tumor diagnosis. Despite its simplicity, LR excels in decision-making, enhancing efficiency, and mitigating supply chain risks. Its application in managing customer data for churn prediction is well-documented [16]. Additionally, Cavalcante et al. [17] demonstrated LR's utility in estimating event probabilities, outlining supplier risk profiles. This predictive modelling tool aids in managing large-scale customer data and provides insights crucial for strategic decision-making in various industries.

Support Vector Machine (SVM), a potent machine learning algorithm, is utilized for classification and regression tasks. Operating on the principle of maximizing the margin between the decision boundary and the nearest data points from different classes, SVMs ensure robust models with reduced misclassification risks, especially effective in high-dimensional spaces and when classes are separable. They create both linear and non-linear decision boundaries using kernel functions, accommodating data that isn't linearly separable. However, SVMs may face challenges with larger or noisier datasets. Renowned for binary classification, SVMs rely on kernel functions and optimal linear lines or hyperplanes to divide outputs, striving for efficient classification in high-dimensional spaces [18][19][20]. Research endeavors include integrating convolutional neural networks for improved feature extraction in image classification [21].

Naive Bayes classification, grounded in Bayes' Theorem and the assumption of feature independence, is widely used despite its "naive" assumption's deviation from real-world scenarios. Successful in applications like document classification and spam filtering, Naive Bayes classifiers excel in scalability and efficiency, particularly beneficial for large datasets. Despite its simplicity, it outperforms more complex techniques and handles numerous features while remaining unaffected by irrelevant ones. However, it may struggle with unseen categories in test data, hindering predictions. Naïve Bayes has been applied in various fields such as Covid-19 risk prediction, lung adenocarcinoma projection, and fake news detection. Despite its simplicity, improving accuracy often requires hybrid techniques [21][22][23].

A Generalized Linear Model (GLM) extends conventional linear regression to accommodate non-normally distributed response variables. It comprises three components: the random component defining the response variable's probability distribution, the systematic component representing explanatory variables as a linear predictor, and the link function linking the random and systematic components. This framework offers flexibility with non-normal response variables and diverse link functions, making GLMs adaptable to various data types and research contexts. Seungwook et al. [24] employed GLM-MANOVA to explore differences in quality management practices among supplier groups based on performance levels. Everardo et al. [25] highlighted GLM's usefulness, when coupled with an appropriate link function, as a supplementary tool for interferometric sensors, mitigating temperature cross-sensitivity issues.

Decision trees, employed for classification and regression, form tree-like structures based on feature values, recursively splitting data into subsets until criteria are met. Each path from root to leaf represents a decision rule leading to an outcome, aiding interpretability. However, they can overfit if overly complex. Advanced methods like Random Forests aggregate multiple decision tree outputs for predictions. Decision trees are non-parametric, suitable for both tasks, and structured for supervised learning. They serve as interpretable data arrangements, dividing input data into zones to predict the dependent variable [26] [27] [28].

Random Forest, an ensemble learning method, combines predictions from multiple decision trees for final outputs. It generates diverse trees through bagging subsets of the original data and random feature subsets at each node. Each tree is trained independently, reducing correlation and enhancing robustness. Outputs are determined by majority vote for classification or averaging for regression. Despite being resistant to overfitting and effective with outliers and unbalanced data, Random Forests may train slowly with many trees and struggle with time-series or extrapolation. The RF algorithm, detailed by Boschetti and Massaron [29], constructs an ensemble of decision trees on randomly selected data subsets. Each tree is built using a sample drawn with replacement from the training set, controlling sample size with a maximum parameter [30][31][32].

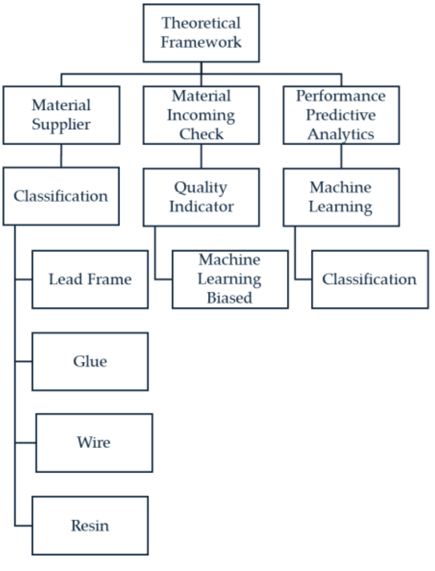

Figure 1: Research theoretical framework

Figure 1 shows the theoretical framework proposed in this research, concentrating on the quality and technical domains in SPE. For material suppliers, a classification technique has been employed to categorize suppliers into four types of direct material. For material incoming checks, a machine learning-based method will be applied to identify the most crucial quality indicator for both user and supplier. Subsequently, a classification technique under machine learning or data mining will be utilized in predictive analytics.

To select the appropriate supplier, it's vital to evaluate various criteria relevant to each supplier's attributes [6]. In the digital transformation era, big data enables companies to leverage supplier performance information for better sourcing decision-making. This chapter's key contributions lie in understanding the existing works on using relevant indicators and performance criteria. This understanding can help companies to use their time more effectively when analyzing and building databases, especially when dealing with quantitative data as in sustainability management [8].

3.0 RESEARCH METHODOLOGY

This study proposed the Supplier Performance Evaluation (SPE) predictive model, which is a new method that improved the accuracy of supplier data by directly incorporating information from production sources. This research utilized high-velocity, high-variety, and unbiased data strategically, employing modern analytical approaches based on machine learning. The proposed data-driven Supplier Performance Evaluation (SPE) model, seen in Figure 2, combines quantitative and qualitative data in a novel way, making it a significant development in supplier performance evaluations.

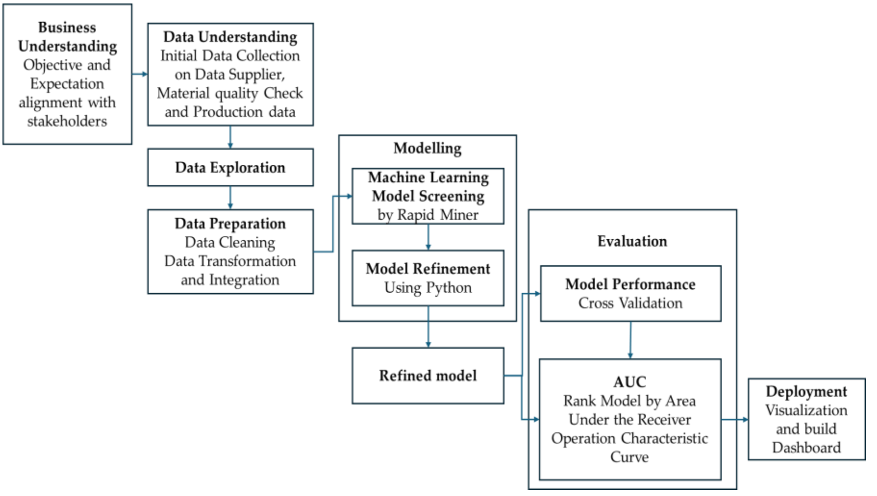

Figure 2: The Proposed Supplier Performance Evaluation Model

This integrated approach highlighted the importance of a clearly defined and repeatable process, combining the capabilities of machine learning techniques with a deep understanding of supplier dynamics. The primary objective was to enhance the efficiency of suppliers by using a comprehensive and data-centric approach that integrates quantitative metrics with qualitative insights.

Data preparation is an essential initial phase that involves the cleaning of unnecessary data, filling in missing information, and standardizing formats in datasets obtained from different platforms. Validation guarantees the precision of this preparation. Following the cleaning process, the data undergoes transformation and enrichment by incorporating pertinent details, such as the creation of new attributes like the nation of origin for suppliers as show in Figure 3.

A	B	C	D	E	F	G	H
STFabType	STFabDesc	STPlantName	STPlantCode	STMaterialGr	STMaterialCat	STMaterialCo	STMaterialDesc
Back-End	Muar	Muar	0959	Chemical Back	BE CHEMICAL	2CHC364M	BELT ADDITIVES (1 box =20 bottles)
Back-End	Muar	Muar	0959	Chemical Back	BE CHEMICAL	2CHC353M	HYDROGEN PEROXIDE
Back-End	Muar	Muar	0959	Chemical Back End		2MI0015M	PURE TIN BALL 19MM DIAMETER
Back-End	Muar	Muar	0959	Chemical Back End		2MI0015M	PURE TIN BALL 19MM DIAMETER
Back-End	Muar	Muar	0959	Chemical Back End		2MI0014M	PURE TIN (99.99%) BARS (0.5gm/pc)
Back-End	Muar	Muar	0959	Chemical Back	BE CHEMICAL	2CHC376M	PYRA TINLUX HS ADDITIVE 20LT/CAR
Back-End	Muar	Muar	0959	Chemical Back	BE CHEMICAL	2CHC361M	CONC. SOLDERLUX ADDITIVE (25LT/CARBOY)
Back-End	Muar	Muar	0959	Chemical Back	BE CHEMICAL	2CHC356M	METHANE SULFONIC ACID (25 LT/JERRY CAN)
Back-End	Malta	Malta	1054	Chemical Back	BE CHEMICAL	2MI00127	TIN ANODE BARS 90/10 12*25*100mm
Back-End	Shenzhen	Shenzhen	3068	Chemical Back	BE CHEMICAL	2XX8067Z	PLT SOLDERON ST-300 ADDITIV 20L
Back-End	Shenzhen	Shenzhen	3068	Chemical Back	BE CHEMICAL	2XX8051Z	PLT SOLDERON TIN CONC 300G/L (EM) PDR D
Back-End	Rennes	Rennes	73RO	Chemical Back	BE CHEMICAL	2CR20142	GALDEN D02 (5 kg drum)
Back-End	Shenzhen	Shenzhen	3068	Chemical Back	BE CHEMICAL	2XX8068Z	PLT SOLDERON RD CONCENTRATE /5L
Back-End	Rennes	Rennes	73RO	Chemical Back	BE CHEMICAL	2CR20143	PERFLUOROPOLYETHER GALDEN HT55 5KG
Back-End	Muar	Muar	0959	Chemical Back	FLUX	2CEE5000	FLUX WSD3810 150g cartridge

Figure 3: Supplier Dataset Sample

Data integration is the process of combining numerous sources of data into one dataset. During this process, textual values are converted into numeric values so that they can be used for computational reasons. The datasets are prepared for the f modelling phase by identifying the key attributes.

The modelling phase is the core component of this machine learning study. The evaluation involves six machine learning models: Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), Generalized Linear Model (GLM), Decision Tree (DT), and Random Forest (RF). The processed data will be divided into 80% for training and 20% for testing. The initial model screening will be performed to quickly and efficiently evaluate the models with minimal coding required. The most effective models will then be improved using Python for adjusting hyperparameters and optimizing performance.

During the evaluation step, the testing datasets are used to verify the accuracy of the trained models. The evaluation of improved models is conducted using the Area Under the Receiver Operating Characteristic Curve (AUC) metric, which quantifies a model's capacity to differentiate between positive and negative examples. The AUC metric is very valuable when dealing with imbalanced datasets and facilitates the comparison of different models. If there is potential for enhancing the outcomes, the process of data preparation is reexamined. A final model evaluation is conducted to confirm that the model is in line with the study objectives.

During the model deployment phase, it is crucial to have a well-structured deployment strategy that includes a thorough plan for monitoring and maintenance. Subsequently, the ultimate predictive model is employed in production for real-time implementation. This prediction model is specifically created to alleviate the workload of specialists by assessing various suppliers using a data-centric method, consequently improving the overall effectiveness and precision of Supplier Performance Evaluation.

4.0 RESULTS AND DISCUSSION

This research aimed to identify the best predictive model for evaluating supplier performance. RapidMiner was initially used to screen various models and top models were then refined with Python for greater control. Effectiveness was measured using Area Under the Receiver Operating Characteristic (AUC) scores across three semiconductor manufacturing databases such as SQM, MES, and SAP, developing a robust system for Supplier Performance Evaluation (SPE), enhancing supply chain efficiency and decision-making.

4.1 Model Screening

Six models were assessed in RapidMiner namely; Naive Bayes (NB), Generalized Linear Model (GLM), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). Based on the results in Table 1, the Logistic Regression model had the highest AUC score of 0.8301, with efficient processing times. Decision Tree was the fastest, while Naive Bayes showed limited predictive capability. Random Forest, though having a moderate AUC, was the most time-consuming. SVM had strong predictive power with a high AUC and low variability.

Table 1: Performance Results

Model	AUC	Standard Deviation	Gains	Total Time	Training Time (1000 Rows)	Scoring Time (1000 Rows)
Naïve Bayes	0.3118	0.0461	0.0	17532.0	99.5	786.28
Generalized Linear Model	0.8269	0.0231	0.0	10299.0	146.3	386.99
Logistic Regression	0.8301	0.0227	0.0	10014.0	284.5	159.20
Decision Tree	0.8126	0.0269	0.0	9504.0	53.4	172.38
Random Forest	0.6798	0.0510	0.0	32059.0	77.4	297.27
Support Vector Machine	0.8152	0.0193	-6.0	18646.0	94.2	129.29

4.2 Model Refinement

Logistic Regression, identified as the best performer, was further evaluated in Python. It demonstrated exceptional performance across all metrics, with an AUC-ROC value of 0.993 as shown in Figure 4. Precision, recall, and F1-score were all perfect, confirming its robustness for predicting supplier performance as shown in Figure 5.

	precision	recall	f1-score	support
Fail	0.99	1.00	1.00	132
Pass	1.00	1.00	1.00	437
accuracy			1.00	569
macro avg	1.00	1.00	1.00	569
weighted avg	1.00	1.00	1.00	569

Figure 4: Logistic Regression model output in Python

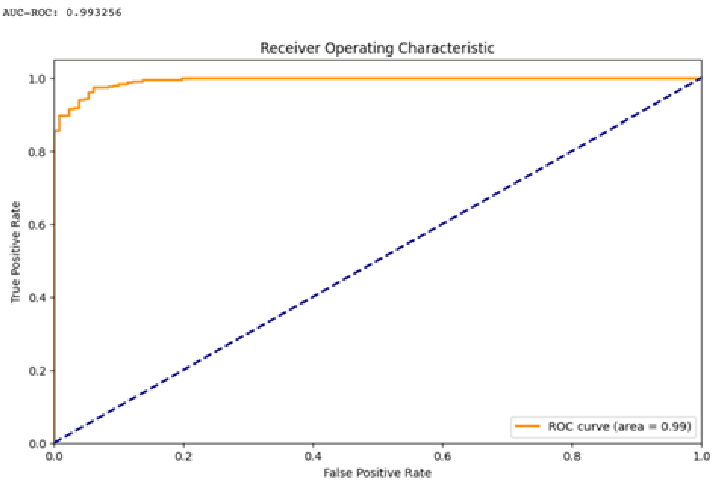

Figure 5: Logistic Regression model AUC and ROC plot in Python

In summary, Logistic Regression was the most effective model, with Decision Tree being the fastest. Naive Bayes showed the least accuracy, and Random Forest was computationally intensive. These findings validate Logistic Regression's suitability for supplier performance prediction.

5.0 CONCLUSION

This study systematically developed an intelligent Supplier Performance Evaluation (SPE) model for direct materials in semiconductor manufacturing using machine learning techniques. The five-phase process—business understanding, data understanding, data preparation, modelling, evaluation, and deployment—ensured a comprehensive and fair supplier assessment. We refined machine learning models, including Naive Bayes, Generalized Linear Model, Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine, using RapidMiner and Python, evaluating them with the AUC metric.

Despite constraints in IT resources and data structure, we also have effectively developed a data-driven supplier performance evaluation (SPE) dashboard, which has facilitated accurate and dependable assessments of suppliers. Logistic Regression was shown to be the most successful model for incoming material checks. Additionally, it was recommended for its efficiency in handling in-process control data. Furthermore, analysis of the patterns of material stock withdrawal provided valuable information for strategic planning and potential cost reductions in year-end operations.

ACKNOWLEDGMENTS

The authors express gratitude to Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka and the Center of Advanced Computer (C-ACT) for their support in this research.

AUTHOR CONTRIBUTIONS

S.H. Yee conducted the research, developed and implemented the machine learning models, and contributed to data analysis and writing the paper. **S.A. Asmai** provided overall guidance, supervision, and critical review of the research and paper. **Z. Abal Abas** offered insights into semiconductor manufacturing and supply chain management and assisted in data integration and paper review. **S. Ahmad** contributed expertise in machine learning techniques, assisted in model implementation, and helped analyze results. **A.S. Shibghatullah** (External Partner) assisted in providing insights, validating the model's applicability in practical scenarios, and contributing to manuscript preparation, formatting, and refinement. D. Petrovic (External Partner) contributed expertise in predictive modeling, evaluated model performance, and provided critical feedback on methodology and results.

CONFLICTS OF INTEREST

The manuscript has not been published elsewhere and is not under consideration by other journals. All authors have approved the review, agree with its submission and declare no conflict of interest on the manuscript.

REFERENCES

[1] K. Encinas, J. Schwarzkopf, M. Mueller and C. Hofmann-stoelting, "Cleaner Logistics and Supply Chain Explanatory factors for variation in supplier sustainability performance in the automotive sector – A quantitative analysis", *Cleaner Logistics and Supply Chain*, vol. 5, pp. 1-12, 2022.

[2] T.G. Hawkins, M. J Gravier, and W. A.Muir, W. A., "The role of supplier performance evaluations in mitigating risk: Assessing evaluation processes and behaviors", *Industrial Marketing Management*, vol 87, pp. 2–17, 2020.

[3] V. Maestrini, P. Maccarrone, F. Caniato, and D.Luzzini, "Supplier performance measurement systems: Communication and reaction modes", *Industrial Marketing Management*, vol.74, pp. 298–308, 2018.

[4] G. Vörösmarty and I. Dobos, "A literature review of sustainable supplier evaluation with Data Envelopment Analysis", *Journal of Cleaner Production*, vol. 264, pp. 1-10, 2020.

[5] M. Giannakis, R. Dubey, I. Vlachos, and Y. Ju, "Supplier sustainability performance evaluation using the analytic network process", *Journal of Cleaner Production*, vol. 247, pp. 1-12, 2020.

[6] S. A. Khan, S. Kusi-Sarpong, F.K. Arhin, and H. Kusi-Sarpong, "Supplier sustainability performance evaluation and selection: A framework and methodology", *Journal of Cleaner Production*, vol 205, pp. 964–979, 2018.

[7] D. Petrovic, M. Kalata, and J. Luo, "A fuzzy scenario-based optimisation of supply network cost, robustness and shortages". *Computers & Industrial Engineering,* vol. 160, pp. 1-16, 2021.

[8] J. J. H. Liou, M. H. Chang, H. W Lo, and M. H Hsu, "Application of an MCDM model with data mining techniques for green supplier evaluation and selection". *Applied Soft Computing*, vol. 109. 2021

[9] I. Abdullah, W.H. Wan Mahmood, H.F. Md Fauadi, M.N. Ab Rahman and S.B. Mohamed, "Sustainable manufacturing practices in Malaysian palm oil mills: Priority and current performance", *Journal of Manufacturing Technology Management*, vol. 28, no, 3, pp.278-298, 2017.

[10] C. Babbar, and S. H. Amin, "A multi-objective mathematical model integrating environmental concerns for supplier selection and order allocation based on fuzzy QFD in beverages industry", *Expert Systems with Applications*, vol. 92, pp. 27–38, 2018.

[11] A. Gunasekaran, M. Kumar Tiwari, R. Dubey and Fosso Wamba, "Big data and predictive analytics applications in supply chain management", *Computers and Industrial Engineering*, vol. 101, pp. 525–527, 2016

[12] S. Islam, S.H Amin, and L. J. Wardley, "Machine learning and optimization models for supplier selection and order allocation planning", *International Journal of Production Economics*, vol. 242, pp. 1-13, 2021.

[13] B. Liu, R. Huang, Y. Xiao, J. Liu, K. Wang, L. Li and Q. Chen, "Adaptive robust adaboost-based twin support vector machine with universum data". *Information Sciences*, vol. 609, pp. 1334-1352. 2022.

[14] P. T. Ogink, O. Q. Groot, B. J. J Bindels, and D. G. Tobert, "The use of machine learning prediction models in spinal surgical outcome: An overview of current development and external validation studies", *Seminars in Spine Surgery,* vol. 33, no. 2, pp. 1-8, 2021.

[15] M. Bertolini, D. Mezzogori, M. Neroni and F. Zammori, "Machine Learning for industrial applications: A comprehensive literature review", *Expert Systems with Applications*, vol. 175, pp. 1-29, 2021.

[16] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees", *European Journal of Operational Research*, vol. 269(2), pp. 760–772. 2018.

[17] I. M Cavalcante, E. M Frazzon, F. A. Forcellini, D. Ivanov, "A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing", *International Journal of Information Management*, vol. 49, pp. 86–97. 2019.

[18] T. Michael, E. I. Aghimien, F. D. Agbajor, Z. Yang, L. Lu, A. R. Adeoye & B. Gopalun, "A review on the integrated optimization techniques and machine learning approaches for modeling, prediction, and decision making on integrated energy systems", *Renewable Energy,* vol. 194, pp. 1-28, 2022.

[19] Z. Wang, J. Li, G. Pandu and Z. Wu. "Machine learning aided multi-objective optimization and multi-criteria decision making: Framework and two applications in chemical engineering", *Computers & Chemical Engineering*, vol. 165, pp. 1-15, 2022

[20] Y. Dhanasekaran, and P. Murugesan, "Improved bias value and new membership function to enhance the performance of fuzzy support vector Machine", *Expert Systems with Applications,* vol. 208, pp. 1-22, 2022

[21] N. Deepa, J. S. Priya, and T. Devi, "Towards applying internet of things and machine learning for the risk prediction of COVID-19 in pandemic situation using Naive Bayes classifier for improving accuracy", Materials Today: Proceedings, vol. 62, no. 4795–4799, 2022

[22] Z. Ye, P. Song, D. Zheng, X. Zhang and J. Wu, "A Naive Bayes model on lung adenocarcinoma projection based on tumor microenvironment and weighted gene co- expression network analysis", *Infectious Disease Modelling*, vol. 7, pp. 498–509, 2022.

[23] P. Phoenix, R. Sudaryono, D.Suhartono. "Classifying Promotion Images Using Optical Character Recognition and Naïve Bayes Classifier", *Procedia Computer Science*, vol. 179, pp. 498–506, 2021.

[24] P. Seungwook, L. Janet, Hartley, W. Darryl, "Quality management practices and their relationship to buyer's supplier ratings: a study in the Korean automotive industry", *Journal of Operations Management*, vol. 19, no. 6, pp. 695-712, 2001.

[25] E. V. Rodriguez and A. D. G. Chavez, "Application of the generalized linear model to enable refractive index measurement with thermal sensitive interferometric sensors", *Optics Communications*, vol. 524, pp.

1-6, 2022.

[26] A. Yeboah-Ofori, S. Islam, S. W. Lee, Z. U. Shamszaman, K. Muhammad, M. Altaf, and M. S Al-Rakhami, "Cyber threat predictive analytics for improving cyber supply chain security", *IEEE Access*, vol. 9, pp. 94318-94337, 2021.

[27] R.C. Barros, A.C. De Carvalho, A.A. Freitas "*Automatic design of decision-tree induction algorithms*", New York: Springer Briefs in Computer Science, 2015.

[28] T. Hastie, R. Tibshirani, and J. Friedman, "*The Elements of Statistical Learning, Data Mining, Inference, and Prediction, vol. 2*", New York: Springer Series in Statistics, 2017.

[29] A. Boschetti, and L. Massaron, "Python Data Science Essentials: Become an Efficient Data Science Practitioner by Understanding Python's Key Concepts", *Birmingham: Packt Publishing,* 2016.

[30] Y. F. Zhao, J. Xie, and L. Sun, "On the data quality and imbalance in machine learning-based design and manufacturing—A systematic review," *Engineering*, vol. 39, pp. 1-54, 2024.

[31] L. Jin, X. Zhai, K. Wang, K. Zhang, D. Wu, A. Nazir, J. Jiang, and W.-H. Liao, "Big data, machine learning, and digital twin assisted additive manufacturing: A review," *Materials & Design*, vol. 244, pp 1-53. 2024.

[32] S. Ramlan, M.M. Fauadi, N.H Razali, and X. Hao, "Agent-based Chemical Mechanical Planarization Qualification for Semiconductor Wafer Fabrication", *Journal of Advanced Manufacturing Technology*, vol. 15 no. 3, pp. 41-53, 2021.