

Integrated machine learning model for managing customer-driven end-to-end supply chain uncertainty

Hanbing Xia ^{a,b}, Shehu Sani^b, Jelena Milisavljevic-Syed ^b and Konstantinos Salonitis^b

^aSchool of Business Administration, South China University of Technology, Guangzhou, People's Republic of China; ^bSustainable Manufacturing Systems Centre, Faculty of Engineering and Applied Science, Cranfield University, Cranfield, UK

ABSTRACT

Supply chain management is increasingly challenged by global disruptions, globalisation, and demand volatility, complicating end-to-end supply chain management in manufacturing. The shift from Industry 4.0 to Industry 5.0 emphasises sustainability, human-centricity, and resilience, highlighting the urgent need to address customer-driven end-to-end supply chain uncertainty. This uncertainty encompasses managing backorder risks and forecasting recycled end-of-life product quantities to promote circularity and sustainability. Current methods fall short in mitigating this uncertainty due to computational limitations. However, machine learning, a key advancement from Industry 4.0, offers a promising solution. Thus, an integrated machine learning model is proposed to mitigate customer-driven end-to-end supply chain uncertainty. This model combines a Bayesian-optimised light gradient-boosting machine (LGBM) to predict backorder risks and a k-nearest neighbour mega-trend diffusion (KNNMTD)-optimised Stacking ensemble model to forecast recycled end-of-life product quantities. Case studies show the Bayesian-optimised LGBM model outperforms benchmark models in accuracy, recall, and AUC (all > 0.8) with 125 s of operational time, while the KNNMTD-optimised Stacking model achieves a superior R² of 0.9515 compared to baseline models. This integrated model enhances prediction performance, generalisation, and supply chain resilience. It enables manufacturers and remanufacturers to optimise production and remanufacturing, allocate resources effectively, and promote a sustainable, circular value chain.

ARTICLE HISTORY

Received 18 May 2025
Accepted 13 March 2026

KEYWORDS

Sustainable supply chain; machine learning; predictive analysis; backorder risk; recycling quantity; uncertainty management

1. Introduction

Supply chain systems function as interconnected business networks that coordinate a series of operations aimed at ensuring the timely delivery of products and services to final customers (Crandall, Crandall, and Chen 2009; Handfield 2002). The end-to-end (E2E) supply chain system is formed by various elements, for instance suppliers, manufacturers, distributors, and customers, that interact between each other not in absolute terms but in a relative sense (Chilmon and Tipi 2020). However, supply chain uncertainty is an issue faced by manufacturing enterprises within value chain, deriving from some challenges, such as global disruptions, demand volatility, and the growing complexity of supply chain networks (Bhatnagar and Sohal 2005; Flynn, Koufteros, and Lu 2016). Internal organisation uncertainties, internal supply chain uncertainties, and external uncertainties are the three primary categories of supply chain uncertainty (Hichri 2016), which is a general term that encompasses risks and uncertainties that could arise at any stage of a global supply chain network (Simangunsong, Hendry,

and Stevenson 2012). These uncertainties may disrupt production schedules, complicate logistics, and exacerbate the difficulty of balancing supply with demand, potentially underscoring the challenges of managing an E2E supply chain that spans from the procurement of raw materials to the delivery of products. Hence, it is necessary to manage the E2E supply chain uncertainty.

A major change in manufacturing supply chains is represented by the transition from Industry 4.0 to Industry 5.0 (Milisavljevic-Syed, Li, and Xia 2023). By focusing on sustainability, human-centricity, and resilience, Industry 5.0 expands on the technical improvements made by Industry 4.0 (Xia et al. 2025). More specifically, this progression allows forward supply chains to integrate more closely with reverse supply chain, enhancing sustainability through practices such as recycling, remanufacturing, and reusing end-of-life products. This integration helps in forming a circular supply chain, where materials are continuously cycled back into production processes, reducing waste and environmental impact. Additionally, the shift towards a human-centric

CONTACT Jelena Milisavljevic-Syed  jelenams@cranfield.ac.uk  Faculty of Engineering and Applied Sciences, Cranfield University, Room F9B, Building 50, Cranfield, Bedfordshire MK43 0AL, UK

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

approach enriches the understanding of the supply chain by emphasising a customer-driven perspective, which in turn enhances system adaptability and responsiveness to individual needs and broader societal challenges. Furthermore, a heightened focus on resilience underpins efforts to mitigate E2E supply chain uncertainty, ensuring that the supply chain could better withstand and adjust to various disruptions.

The customer-driven E2E supply chain uncertainties fall under the category of internal supply chain uncertainties, which are specified as end-customer demand and customer as a supplier (de Lima, Seuring, and Sauer 2022). On the one hand, the irregular orders from the end-customer potentially lead to backorder risk; on the other hand, in a relative sense, the end-customer can be seen as a supplier of reverse supply chain by providing end-of-life products. More specifically, firstly, backorder risk refers to the consequences arising from the inability to fulfil client orders, primarily because of inventory shortages or supply chain interruptions (Sani et al. 2023). By predicting the potential backorder risk, the enterprise can develop more sophisticated strategies to improve adaptability and flexibility and reduce the negative effects of end-customer driven uncertainty, such as decreased customer satisfaction and lost sales opportunities. Secondly, the management of end-of-life products presents significant challenges considering uncertainty surrounding the quality, amount, and return timing (Xia, Han, and Milisavljevic-Syed 2023b). The end-customer driven uncertainty can be also managed by accurately predicting the recycling quantity, thereby facilitating the development of sustainable reverse supply chain strategies and remanufacturing planning. Therefore, focusing on customer-driven E2E supply chain uncertainty is essential for promoting circularity in manufacturing processes, thus supporting sustainable practices and enhancing supply chain resilience against disruptions.

Existing studies primarily address either backorder risk or recycling quantity, often in isolation, without integrating the multiple customer-driven uncertainty factors present across the E2E supply chain. Conventional prediction methods for managing backorder risk and end-of-life product quantities have their limitations. Specifically, traditional methods for managing backorders risks, such as fuzzy modelling approaches (Kazemi, Ehsani, and Jaber 2010), rely heavily on manual monitoring and intervention, addressing shortages reactively after they occur, which leads to delays, increased costs, and customer dissatisfaction (Maitra and Kundu 2023). Moreover, conventional methods for forecasting end-of-life product quantities, like grey model (Zhou and Ma 2020) and population balance model (Xu, Yano, and Sakai 2018), struggle to capture the nonlinear relationship between

multiple factors and fail to provide stable predictions under variable conditions. These traditional methods are inadequate due to their reactive nature, limited data utilisation, and inefficiency in handling the complexities of modern supply chains. Unlike traditional methods, machine learning-based models are scalable and adaptable, enabling them to effectively handle the complexities of multi-tiered supply chains and dynamically respond to fluctuations in demand (Christopher 2022). As a key technology in Industry 4.0, machine learning can effectively capture complex, nonlinear patterns in both back-order risk prediction and recycling quantity forecasting, leading to improved predictive accuracy and better decision-making (Supply Chain Brain 2022). However, existing machine learning studies often focus on a single aspect of supply chain uncertainty, without addressing the integrated nature of both forward and reverse supply chain uncertainties within a unified model.

To manage the customer-driven E2E supply chain uncertainty, an integrated machine learning model is proposed based on the combined forward and reverse supply chain processes. More specifically, the proposed model combines a Bayesian-optimised light gradient-boosting machine (LGBM) to predict backorder risks and a *k*-nearest neighbor mega-trend diffusion (KNNMTD)-optimised Stacking ensemble model to estimate recycled end-of-life product quantities. The proposed integrated model aims to enhance prediction performance and generalisation capability in managing customer-driven supply chain uncertainties, which potentially enhances supply chain resilience and aligns supply with demand. Additionally, by optimising both production and remanufacturing processes, this proposed model supports more effective resource allocation and operational management, thereby promoting a sustainable and circular value chain. Specifically, the empirical results demonstrate that the Bayesian-optimised LGBM model outperforms conventional machine learning models, such as random forest, in terms of accuracy, recall, and AUC. Furthermore, the KNNMTD-optimised Stacking ensemble model demonstrates superior predictive precision compared to single baseline forecasting models, validating the effectiveness of the proposed integrated approach in handling both backorder risks and recycling quantity forecasting.

The rest of this paper is as follows. A comprehensive literature, focusing on machine learning methods for predicting backorder risk and the of end-of-life product quantity, is presented in Section 2. The proposed hybrid Bayesian-optimised LGBM model and KNNMTD-optimised Stacking ensemble model are detailed in Section 3. Then, the empirical studies of the proposed model and their results are illustrated in

Table 1. Comparison of conventional and machine learning approaches for backorder risk and recycling quantity prediction.

	Conventional Methods	Machine Learning Methods
Backorder Risk Prediction	<ul style="list-style-type: none"> - Economic order quantity model (Björk 2009) - Fuzzy models (Kazemi, Ehsani, and Jaber 2010) - Inventory management models (Feng et al. 2012) 	<ul style="list-style-type: none"> - RF and gradient boosting techniques (Hajek and Abedin 2020) - RF and gradient boosting techniques (Islam and Amin 2020) - LGBM, RF, NN, XGBoost, SVM, BB, KNN, and LR (Ntakolia et al. 2021) - ANN (Shajalal, Hajek, and Abedin 2023) - ANN (Jassim et al. 2023) - XGBoost (Zhang et al. 2022) - SVR (Xia, Han, and Milisavljevic-Syed 2023a) - KNN (Abbasi and El Hanandeh 2016) - DT (Cha et al. 2017) - GBRT (Lu et al. 2022) - Ensemble Model (Al-Omran and Khan 2024) - RF (Mudannayake et al. 2022).
End-of-life Product Quantities Prediction	<ul style="list-style-type: none"> - Grey model (Zhou and Ma 2020) - Population balance model (Xu, Yano, and Sakai 2018) - Structural equation models (Agrawal and Singh 2019) 	

Section 4. In Section 5, an in-depth analysis of the results is discussed. Finally, a conclusion is drawn, and future research is highlighted.

2. Literature review

The machine learning models applied to manage backorder risk within supply chain operations are reviewed. Then, existing machine learning-based models for estimating end-of-life product quantities are reviewed. The comparison of conventional and machine learning approaches for backorder risk and recycling quantity prediction is presented in Table 1.

2.1. Machine learning-based predictive methods for backorder risk

Backorder risk is a significant challenge in forward supply chain management, driven by factors such as uncertain demand, fluctuating lead times, suboptimal inventory policies, and other operational disruptions. High demand variability can lead to frequent stockouts and subsequent backorders, especially when forecasts fail to capture sudden spikes (Lee 2004). Similarly, longer and more unpredictable lead times, for instance, during demand spikes or supply disruptions, significantly increase the likelihood of backorders (Disney and Towill 2003). Poor inventory management, including inaccurate demand forecasting, suboptimal replenishment strategies, and insufficient safety stocks, further contributes to backorder risk (Nahmias and Olsen 2015). Additionally, unreliable suppliers with inconsistent delivery performance exacerbate backorder occurrences by failing to meet expected supply commitments (Monczka et al. 2020).

Predicting product backorders has attracted growing interest in recent years due to its critical financial implications for firms (Islam and Amin 2020). To address backorder-related uncertainties, conventional research often relies on optimisation or fuzzy modelling approaches. For example, Björk (2009) and Kazemi,

Ehsani, and Jaber (2010) extended the classical economic order quantity models by incorporating fuzzy demand and lead time parameters, demonstrating that accounting for uncertainty can improve decision-making in inventory systems with backorders. Feng et al. (2012) further highlighted the challenge of forecasting line-replaceable unit parts demand under backorder conditions, quantifying the impact of demand volatility and inventory cost uncertainties on backorder risk. These foundational studies underscore the importance of explicitly modelling uncertainty in backorder management, setting the stage for more data-driven predictive methods.

With the rise of data-driven analytics and industry 4.0 technology, machine learning techniques have been increasingly adopted to tackle supply chain risks, including backorder risk. Supply chain risk management often involves rapid and adaptive decision-making, while AI technologies enable better risk identification and mitigation by revealing hidden patterns in complex operational datasets (Baryannis et al. 2019), reflecting that data-driven machine learning approaches can strengthen supply chain resilience against disruptions. A variety of machine learning paradigms have been investigated for different supply chain risk problems. For instance, Ojha et al. (2018) employed a Bayesian network model to capture risk propagation in multi-echelon supply networks, providing a holistic view of network vulnerability and resilience. A recent review summarises reinforcement learning (RL) algorithms for inventory management (Rolf et al. 2023), but most RL-based supply chain studies use simplified or simulated data, highlighting the need to transition towards real-world, industry-scale problem setting to fully realise RL's potential in mitigating risks. In parallel, an adaptive federated learning system for information sharing is proposed, which allows multiple supply chain firms to jointly train risk prediction models without exchanging raw data (Zheng, Ivanov, and Brintrup 2025). Another emerging direction is the use of casual machine learning for proactive risk management. Wyrembek, Baryannis, and Brintrup (2025) argued that

casual machine learning can uncover cause-and-effect relationships for targeted risk interventions. Such capabilities are crucial for moving supply chain risk management towards more prescriptive and human-centric decision support, as expected in the transition from Industry 4.0 to Industry 5.0.

For backorder-specific studies, a broad array of machine learning applications in supply chain risk management has emerged in the literature, further justifying the contribution of this research. Ntakolia et al. (2021) conducted a comparative analysis of eight widely used classifiers, including LGBM, random forest (RF), neural networks (NN), extreme gradient boosting (XGBoost), support vector machine (SVM), balanced bagging (BB), k-nearest neighbors (KNN), and logistic regression (LR), to predict backorders. Their results demonstrated the promise of machine learning for this task, though the study did not adequately tackle the common challenge of class-imbalanced datasets in backorder data, which often complicate real-world applications. Similarly, Islam and Amin (2020) employed distributed RF and gradient boosting techniques for backorder prediction and reported improved predictive performance; however, they also faced limitations in handling highly imbalanced training data. Other researchers have explored deep learning and ensemble approaches. For example, a deep neural network-based approach was introduced to predict backorder, aiming to enhance supplier operational efficiency by proactively identifying likely backorder events (Shajalal, Hajek, and Abedin 2023). On the other hand, an ensemble-based backorder prediction system was advocated, combining RF and gradient tree boosting methods with an under-sampling strategy, which proved effective in mitigating noisy data and minimising overfitting, thereby improving the reliability of predictions (Hajek and Abedin 2020).

In summary, this growing body of studies underscores that machine learning-based methods have great potential to improve supply chain performance by predicting and mitigating risks such as backorders. Prior studies have demonstrated machine learning techniques can be used to anticipate supply chain delivery delays and other disruptions, thereby helping firms anticipate supply chain delivery delays and prevent stockout situations. However, despite highlighting the potential of these techniques for improving backorder risk prediction, challenges remain, including severe class imbalance in data, limited generalisation to new conditions, and the need for rapid model training and updating under operational time constraints.

2.2. Machine learning methods for estimating end-of-life product quantities

Developing a predictive model for recycled end-of-life product quantities is essential to mitigating uncertainties in the reverse supply chain. Several conventional predictive methods have been proposed for estimating end-of-life product quantities, including the grey model (Zhou and Ma 2020), population balance model (Xu, Yano, and Sakai 2018), and structural equation models (Agrawal and Singh 2019), among others. However, these methods struggle to capture the multifaced and nonlinear relationship of socioeconomic factors, resulting in unstable prediction accuracy. Machine learning techniques, which excel at analysing unstable nonlinear datasets, are commonly used to predict end-of-life product quantity. These include artificial neural network (ANN), extreme gradient boosting (XGBoost), support vector regression (SVR), decision tree (DT), k-nearest neighbor (KNN), gradient boosting regression tree (GBRT), as well as a variety of ensemble models.

Artificial Neural Network (ANN). ANN mimics the human brain by creating networks of input, hidden, and output neurons (Puntarić et al. 2022). There are several different ANN designs that are extensively utilised to anticipate end-of-life product quantities. These include recurrent neural networks (Li and Ma 2019), long short-term memory networks (Jassim et al. 2023), deep neural networks (Akanbi et al. 2020; Altin, Budak, and Özcan 2023), feedforward neural networks (Galang and Ballesteros 2018), backpropagation neural networks (Oguz-Ekim 2021), adaptive neuro-fuzzy inference systems (Jafari and Mousavi 2024), Elman neural networks (Meza et al. 2019), and nonlinear autoregressive models (Rashid et al. 2022). Strong prediction capabilities in end-of-life product forecasting have also been demonstrated using multi-layer perceptrons (MLP), an advanced form of feedforward neural networks characterised by deeper and more complex architectures (Coskuner et al. 2020). A range of AI techniques, such as Bayesian optimisation (Hoy et al. 2022), genetic algorithms (Shamshiry et al. 2014), and particle swarm optimisation (Wang et al. 2022) have been utilised to enhance the performance of ANN models. Supporting methods like principal component analysis (Liu et al. 2022), gamma test (Noori, Karbassi, and Sabahi 2010), grey models (Tian et al. 2023), and structural breaks (Adamović et al. 2017) also play critical roles in refining the predictive accuracy of these ANN models.

Extreme Gradient Boosting (XGBoost). XGBoost constructs trees based on feature splitting and continuously fits residuals from previous predictions to enhance model performance through iterative updates.

The XGBoost model has been demonstrated to be effective in forecasting the creation of municipal solid waste (MSW) in China across different shared socioeconomic pathways (Zhang et al. 2022).

Support Vector Regression (SVR). SVR employs a kernel function to map samples into high-dimensional space and uses a hyperplane for regression. SVR, enhanced through wavelet transform has been applied to predict weekly household waste generation (Abbasi et al. 2014). Similarly, a PSO-optimised SVR model is developed for estimating the quantity of ELVs in China (Xia, Han, and Milisavljevic-Syed 2023a). Additionally, a hybrid model integrating categorical principal components analysis and SVR is developed to forecast construction waste generation (Cha, Moon, and Kim 2022).

K-Nearest Neighbor (KNN). KNN is a nonparametric, instance-based lazy learning model that has a reputation for being resilient under noisy settings. A weighted KNN algorithm is effectively employed to predict Australia MSW generation (Abbasi and El Hanandeh 2016).

Decision Tree (DT). DT, a supervised machine learning algorithm composed of root, internal, and leaf nodes, is celebrated for its excellent interpretability. A new approach for forecasting construction waste generation is applied using the chi-squared automatic interaction detection method, a specific type of DT model (Cha et al. 2017).

Gradient Boosting Regression Tree (GBRT). Employing a boosting strategy, GBRT iteratively learns from multiple trees to make collective decisions. This method is employed for forecasting waste generation in Beijing and Shenzhen, China (Lu et al. 2022). In Addition, GBRT has been combined with ANN to forecast construction waste generation (Kontokosta et al. 2018).

Ensemble Model. Ensemble learning involve sophisticated ensemble learning techniques such as Boosting, Bagging, and Stacking, which are types of hybrid machine learning models (Tan et al. 2019). In these models, either homogeneous or heterogeneous algorithms may be iteratively applied to construct more robust predictive frameworks. Specifically, Bagging and Boosting techniques generally utilise identical base learners, which can increase model correlation and potentially lead to overfitting. In contrast, the relationships between predicted outputs and actual observations could be captured more thoroughly due to Stacking's incorporation of diverse base model types. Additionally, an ensemble voting regressor, combining RF, extreme random trees, elastic net, and XGBoost, has been employed to forecast healthcare waste generation in the Kingdom of Bahrain (Al-Omran and Khan 2024).

Random Forest (RF). As a classification and regression tree model, RF employs a Bagging integration

approach. RF has demonstrated the capability to achieve results comparable to the best deep learning models after a grid search in forecasting MSW generation (Mudanayake et al. 2022).

In summary, due to the underdeveloped nature of the reverse recovery industry (Xia et al. 2024), developing a predictive model for forecasting end-of-life product quantities needs to deal with small sample datasets. However, single machine learning models tend to exhibit poor generalisation in such scenarios. Moreover, unlike the Stacking ensemble model, other ensemble models (e.g. Bagging and Boosting) are prone to overfitting, as they rely on homogeneous algorithms as their base models.

In conclusion, advancements in data analytics have significantly elevated the role of machine learning techniques in supply chain management, particularly for predicting backorders risk and end-of-life product volumes. Nevertheless, there remain important gaps that this research aims to address. Existing studies focus on a single aspect of uncertainty, such as backorder risk or recycling volume, without integrating multiple customer-driven uncertainty factors across the E2E supply chain. Moreover, current prediction methods for backorder risk and end-of-life product volume have their own limitations. To overcome these challenges, this research proposes an integrated machine learning model that combines two complementary predictive techniques, namely a Bayesian-optimised LGBM model for backorder risk prediction, and a KNNMTD-optimised Stacking ensemble model for forecasting recycled end-of-life product quantities. By integrating these models, the approach concurrently addresses both forward supply risks associated with delivery delays and reverse supply uncertainties associated with recycling quantity for remanufacturing, representing a novel contribution. Unlike existing research, which typically addresses these issues separately, integrating backorder risk prediction and recycled end-of-life product volume prediction into a single model not only mitigates E2E supply chain uncertainty but also supports global optimisation and collaborative decision-making by offering a more thorough view of the supply chain. Specifically, the Bayesian hyperparameter optimisation of the LGBM ensures high predictive accuracy and efficiency, facilitating fast training and strong scalability for processing large, imbalanced datasets. The Stacking ensemble, enhanced by KNNMTD, improves generalisation in forecasting end-of-life product quantities by combining heterogeneous base learner with a data augmentation strategy, leveraging the strengths of multiple models to improve prediction performance and robustness.

3. Methodology

An integrated machine learning-based model based on the combined customer-driven forward and reverse supply chain processes within the value chain is proposed to manage the customer-driven E2E supply chain uncertainty (see Figure 1). The integrated approach consists of two sequential but complementary steps, each addressing distinct yet interconnected aspects of supply chain uncertainty. Firstly, a Bayesian-optimised LGBM is employed to predict backorder risks. This step targets forward supply chain uncertainty related to delivery delays, enabling timely and proactive measures to mitigate adverse impacts such as customer dissatisfaction. Secondly, following the backorder risks prediction, a KNNMTD-optimised Stacking ensemble model is developed to estimate recycled end-of-life product quantities. By predicting the expected return flow of products, this step focuses on reverse supply chain uncertainty, thereby facilitating more effective planning for remanufacturing and recycling processes and supporting circular economic practices.

Furthermore, optimising production schedules and resource distribution through this integrated approach helps balance new product manufacturing with recycling operations, reducing waste, enhancing supply chain resilience, and improving overall flexibility and

responsiveness. For instance, accurately forecasting recycled end-of-life product quantities can inform production schedules for new products, thereby reducing the backorder risk. Therefore, this research adopts a two-step integrated approach, specifically tailored to address each type of uncertainty, ultimately enhancing overall modelling effectiveness and practical applicability.

3.1. The hybrid Bayesian-optimised LGBM model

As shown in Figure 2, this research employs a Bayesian-optimised LGBM model to forecast the risk of supply chain delivery delays. The first step in the process is gathering relevant data on key factors influencing delivery delays within the supply chain. Following comprehensive data processing, the LGBM model is developed, and its hyperparameters are methodically adjusted using a Bayesian optimisation algorithm. Four statistical metrics: Accuracy, Recall, Precision, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are used to assess the model's predictive performance (Su et al. 2025). These metrics are selected for their collective ability to assess model performance in imbalanced and risk-sensitive scenarios such as delivery delay prediction. Accuracy gives an overall success rate, Precision reduces false alarms, Recall captures true delay cases, and AUC-ROC measures overall discrimination across thresholds. Together, they offer a comprehensive assessment of the model's predictive effectiveness.

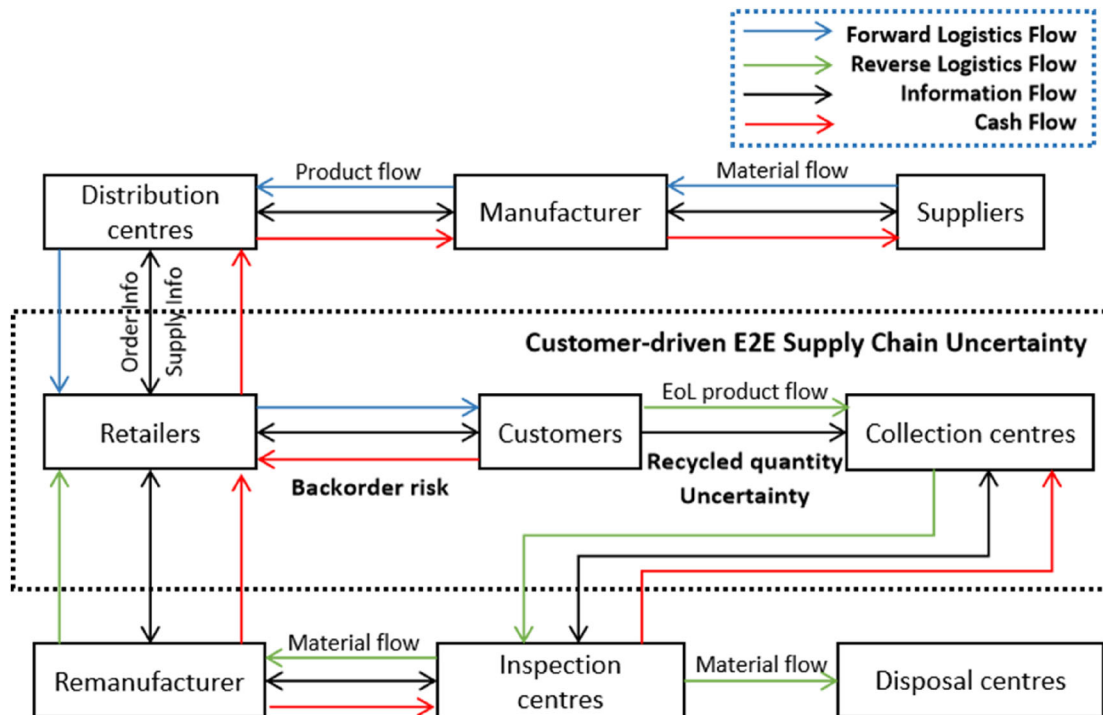


Figure 1. The customer-driven E2E supply chain uncertainty within the value chain.

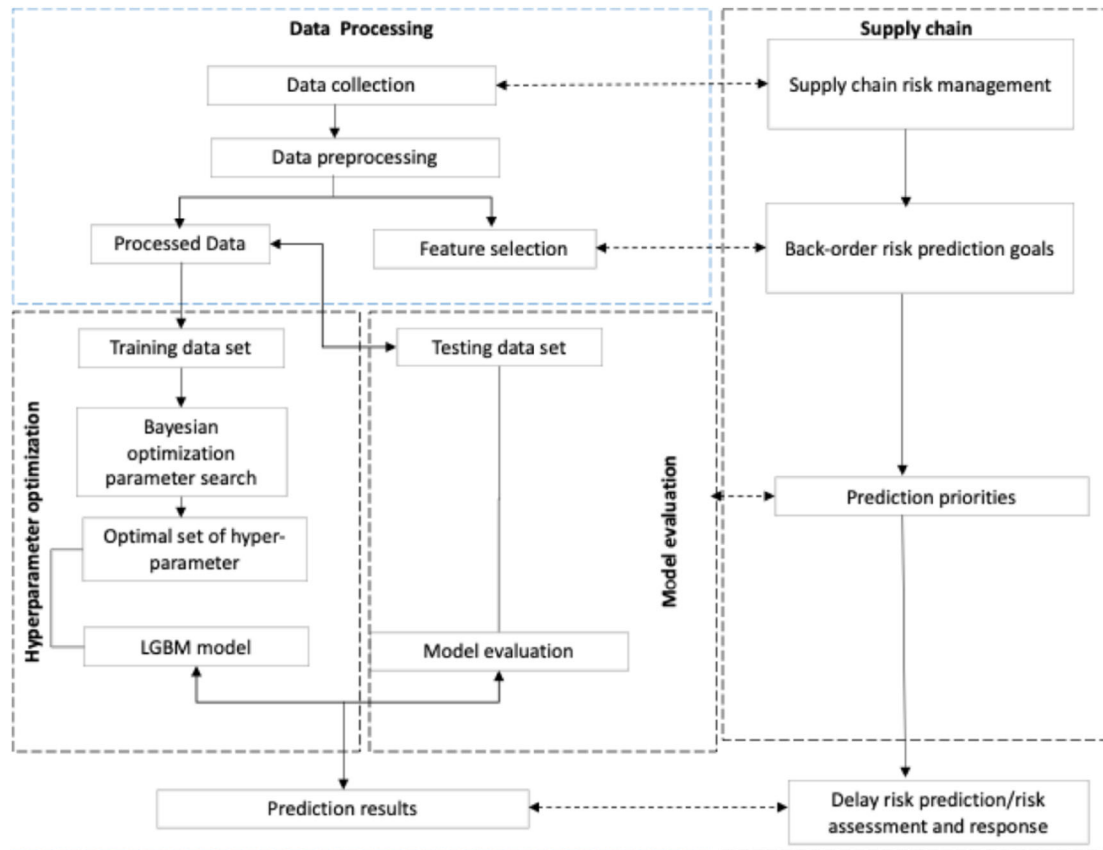


Figure 2. A paradigm for supply chain delivery risk prediction.

3.1.1. Data preprocessing

The quality and dependability of downstream analytical outcomes are greatly enhanced by data preprocessing, which is a fundamental component of data evaluation and extraction (Xu et al. 2015). The data processing procedure in this research involves several key steps, including data standardisation, removing outliers, and integrating datasets. To ensure that data preprocessing does not compromise data quality, missing values are carefully handled by inputting appropriate values. To facilitate more efficient and consistent analysis in subsequent modelling stages, category string values, namely ‘Yes’ and ‘No’ are converted into binary representations (0 and 1).

To address class imbalance, random under-sampling is applied to the training set, where the majority class samples are randomly reduced to create a more balanced dataset (Rafrastara et al. 2023). By removing a portion of the majority class, under-sampling prevents the model from being biased towards the majority class, thereby improving its generalisation ability (Ramyaichitra and Manikandan 2014). This method ensures that the data retains its authenticity without introducing excessive noise, which is crucial for improving prediction performance and preventing overfitting (Tanha et al. 2020).

3.1.2. Model development

LGBM Model. The Gradient Boosting Decision Tree (GBDT) framework is implemented open-sourcely by the LGBM (Chen et al. 2019). While traditional GBDT model are often limited by low computational efficiency and poor scalability when applied to high-dimensional, large-scale datasets, LGBM addresses these limitations through a series of algorithmic optimizations (Al Daoud 2019). Specifically, LGBM enhances training efficiency by incorporating two core techniques: the Gradient-based One-Side Sampling algorithm, which selectively samples data points with large gradients to maintain accuracy while reducing computational cost; the histogram-based decision tree algorithm, which accelerates the process of finding optimal split points.

Bayesian Optimisation. Choosing appropriate hyperparameters is critical to model performance, as well-tuned parameters can substantially improve predictive accuracy and generalisation capability (Frazier 2018). Bayesian optimisation offers an efficient strategy for hyperparameter tuning by leveraging prior evaluations to inform the selection of subsequent parameter configurations. This sequential, probabilistic approach enables the algorithm to explore the parameter space more effectively, requiring fewer iterations to identify near-optimal

Table 2. Confusion matrix.

Actual Outcome	Predicted Outcome	
	P	N
N	FP	TN
P	TP	FN

* N: Negative; P: Positive.

solutions. To estimate delivery delay risks, the optimal hyperparameter configuration of LGBM model is found using Bayesian optimisation in this research (Wang et al. 2013).

3.1.3. Model assessment

To assess the binary classification model's performance, overall classification accuracy and error rate are employed as fundamental metrics. In order to provide a more comprehensive assessment, Precision and Recall are also included as evaluation indicators. The confusion matrix is used to calculate these metrics. As presented in Table 2, False Negative (FN) occurs when the actual class is positive but the model predicts negative; False Positive (FP) occurs when the actual class is negative but the prediction is positive; True Negative (TN) is the number of instances correctly predicted as negative; and True Positive (TP) is the number of instances correctly predicted as positive (Satapathy et al. 2015).

The accuracy rate represents the percentage of accurately classified instances among all samples, as defined in Equation (1).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (1)$$

The precision rate indicates the percentage of true positive predictions among all instances classified as positive by the model, as expressed in Equation (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The recall quantifies the percentage of true positive instances that the model accurately recognises, according to Equation (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Furthermore, the AUC-ROC, commonly abbreviated as AUC, is used to evaluate the model's performance. AUC is a comprehensive metric that quantifies how well the classifier can distinguish between positive and negative classes. It enables performance assessment across various classification thresholds and is particularly valuable for evaluating model robustness in the presence of imbalanced datasets.

3.2. The hybrid KNNMTD-optimised stacking ensemble model

The proposed hybrid KNNMTD-optimised Stacking ensemble model to forecast recycled end-of-life product quantities is described in Figure 3. Initially, data on influencing factors for end-of-life products are collected, standardised through z-score standardisation, and augmented through a method known as KNNMTD. Then, the Stacking ensemble model is built, with each base model's parameters optimised using a combination of grid search (GS) and time series cross validation (TSCV). Finally, the statistical metrics, namely mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R^2), are used to assess model's prediction performance.

3.2.1. Data preprocessing

The data processing procedure in this research involves several key steps, including data standardisation, outlier removal, and dataset integration. The specific data standardisation process is outlined below. The data values x and y are transformed into x' and y' using z-score standardisation to eliminate the impact of the varying attributes, as detailed in Eqs. (4) and (5) (Xia, Han, and Milisavljevic-Syed 2023a). To ensure that data preprocessing does not compromise data quality, missing values are carefully handled by inputting appropriate values.

$$x' = \frac{x_i - \mu_x}{\sigma_x} \quad (4)$$

$$y' = \frac{y_i - \mu_y}{\sigma_y} \quad (5)$$

Where μ represents the mean value, and σ denotes the standard deviation of the corresponding data distribution.

To counteract overfitting in small sample datasets, the KNNMTD method, a specialised data augmentation method, is employed (Sivakumar et al. 2022). By utilising this method, the sample size is increased, and the dataset is expanded, maximising the use of limited data and enhancing the machine learning model's generalisation capability. More specifically, the MTD technique is applied to estimate the diffusion dynamics within this subsample, while the KNN algorithm is used to identify the nearest data points to a given instance. The KNNMTD method is utilised to calculate sample variance, to adjust for data skewness, and to generate artificial virtual samples through a probabilistic framework based on predicted membership values. A new, augmented dataset is then constructed by merging the original observations with the synthetic

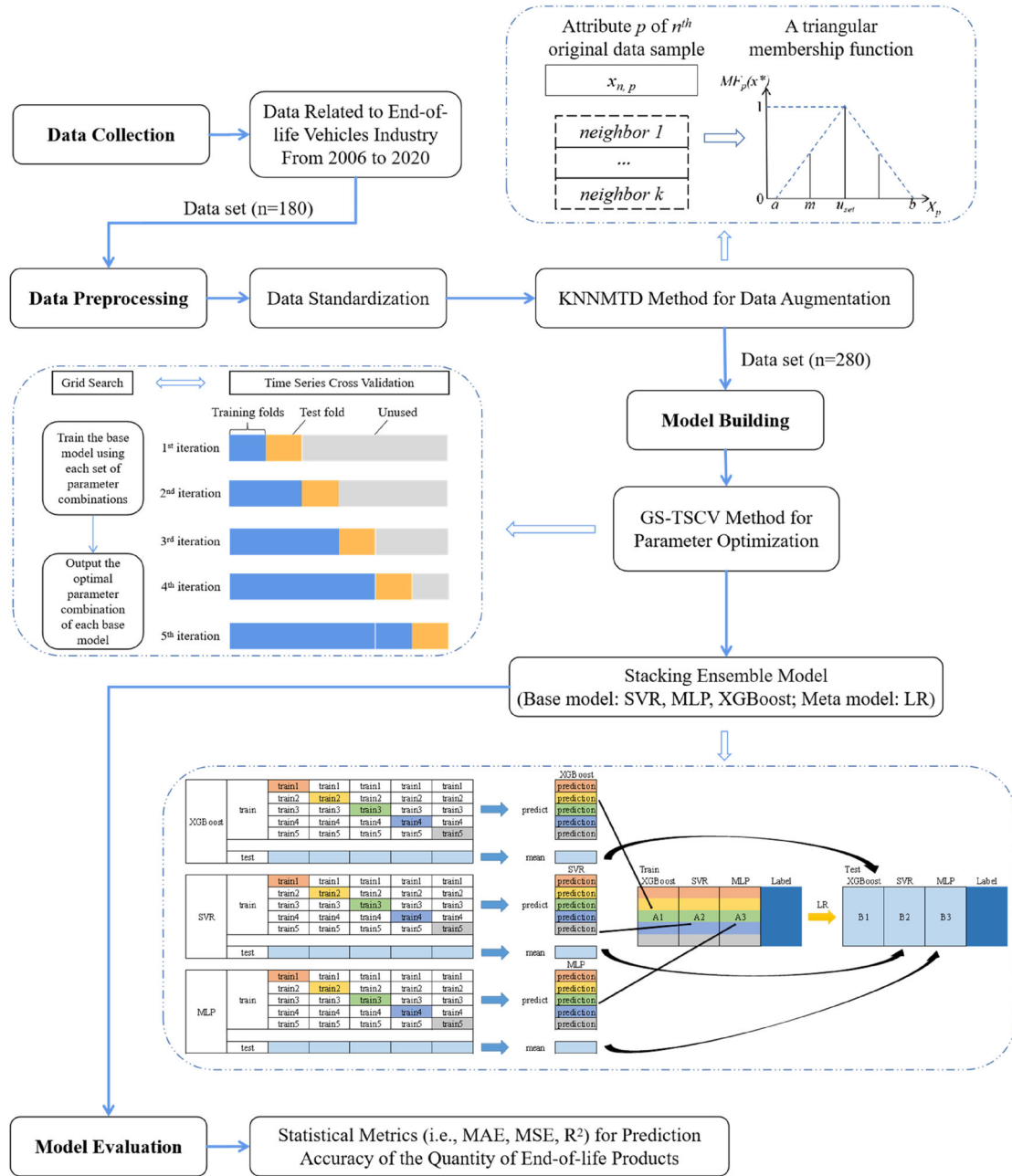


Figure 3. The flowchart of the hybrid KNNMTD-optimised Stacking ensemble model.

samples. To evaluate the similarity between the original and artificial samples, the pairwise correlation difference (PCD) is computed by the Frobenius norm, as shown in Equation (6).

$$PCD = \text{corr}(X_r) - \text{corr}(X_s)_F \quad (6)$$

Where $\text{corr}(X_r)$ and $\text{corr}(X_s)$ are the Pearson correlation matrices of the original sample matrices X_r and the artificial virtual sample matrices X_s , respectively.

3.2.2. Model development

By eliminating redundancy and optimising prediction accuracy, Stacking, a parallel ensemble learning strategy, can enhance the predictive capabilities beyond what a single machine learning model can achieve (Xia, Han, and Milisavljevic-Syed 2023b). The Stacking framework is composed of two hierarchical learning layers: the base learner and the meta-learner.

To provide preliminary predictions, multiple heterogeneous machine learning algorithms are trained on the

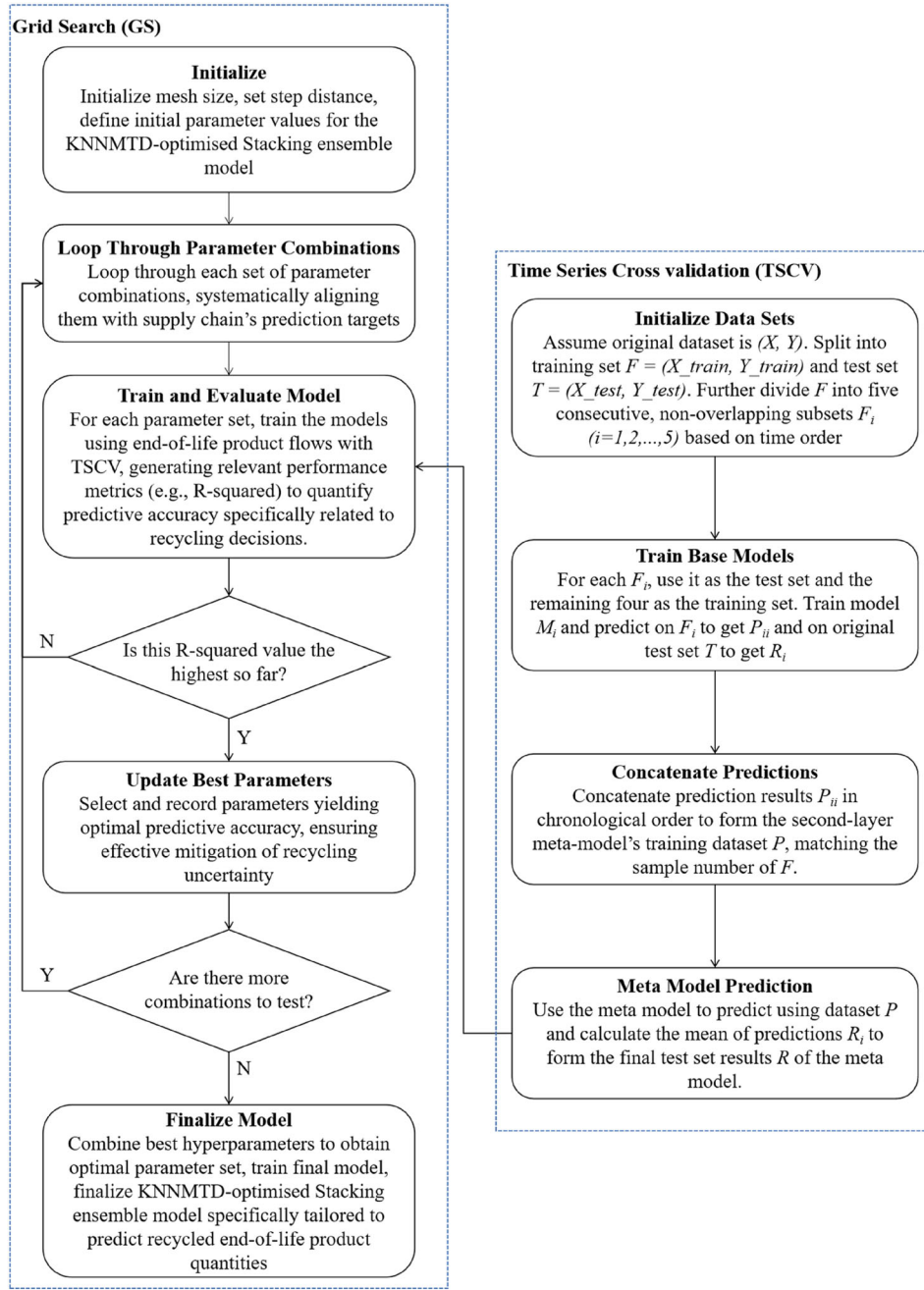


Figure 4. Detailed combination of GS and TSCV steps (a)(b).

entire training dataset in the base learner. Subsequently, in the meta-model layer, these predictions along with the true values are used for further training.

Finding the optimal set of model parameters is essential to reducing the generalisation error of the model. An overly complex model may lead to overfitting, increasing the generalisation error, whereas an overly simplified model may underperform. To identify the optimal hyperparameter configuration for the Stacking ensemble,

a hybrid approach combining GS and TSCV is adopted. Figure 4 shows the detailed steps of this combination.

3.2.3. Model assessment

Three statistical metrics are applied to assess the prediction performance of the hybrid KNNMTD-optimised Stacking ensemble model, including MAE, MSE, and R^2 , as outlined in Equations (7)–(9) (Xia, Han, and

Milisavljevic-Syed 2023a).

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (8)$$

$$R^2(y_i, \hat{y}_i) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (9)$$

Where y_i , $\bar{y}_i$, and $\hat{y}_i$ denote the original, mean, and predicted values of variable Y, respectively.

4. Case study

4.1. Case study for the hybrid Bayesian-optimised LGBM model

Within supply chain management, backorders constitute a crucial risk to the timely fulfilment of deliveries, arising from factors such as limitations in material transportation capabilities, unforeseen external events, deficiencies in supplier evaluation processes, and supplier management inefficiencies. The occurrence of backorders can lead to considerable financial losses due to delays in products delivery. Consequently, backorder data is employed in this research to verify the effectiveness of the Bayesian-optimised LGBM method. An actual, imbalanced historical dataset spanning eight weeks, is utilised (Predictive Backorder Competition 2025). Table 3 provides a detailed description of the features of the dataset. The original dataset consists of 13,937 positive instances and 1,915,998 negative instances, reflecting a significant class imbalance. The data preprocessing steps for this dataset, including handling missing values, random under-sampling, and converting categorical data, have been previously detailed. After applying random under-sampling, the positive instances are reduced to 13,937 samples, and the negative instances are reduced to 584,558 samples. It is worth noting that while the post-sampling ratio is approximately 1:42, this level of imbalance was selected to prevent severe information loss. While a perfectly balanced distribution (1:1) is often theoretically ideal, it is not always optimal for real-world applications where the minority class represents a rare event (Weiss and Provost 2003). In this specific scenario, achieving a strictly balanced (1:1) distribution would require discarding over 97% of the negative instances. Such extreme under-sampling would lead to severe information loss regarding the 'normal' state of the supply chain, potentially causing the model to overfit to the minority class and generating excessive false positives (He and Garcia 2009). Therefore, a ratio of 1:42 is selected

Table 3. Definition of features.

Features	Definition
PPAP risk	Risk related to the production part approval process
Sales forecasts	Sales forecasts for the next 3, 6, and 9 months
Operational constraints	Operational constraint
Historical sales	Sales volumes over the past 1, 3, 6, and 9 months
Risk indicators	Risk indicators
Supplier performance	Supplier performance averages over 6 and 12 months
Potential issues	Identified potential issue with the item
Minimum stock level	Minimum recommended stock level
Inventory level	Current national inventory
Overdue parts	Overdue parts from the supplier
Lead time	Delivery lead time
In-transit quantity	Stock quantity currently in transit
Overdue local stock	Quantity of overdue local stock
Backorder status	Product backorder status

as a strategic trade-off. It significantly reduces the original extreme imbalance (1:137) to a range where the Bayesian-optimised LGBM, which is robust to mild class imbalance, can effectively learn the decision boundary without sacrificing the rich information contained in the majority class data.

4.1.1. Model development

This research compares seven machine learning models: LGBM, logical regression (LR), XGBoost, k-nearest neighbour (KNN), random forest (RF), naive bayes (NB), and decision tree (DT) in order to evaluate the efficacy, resilience, and predictive performance of the proposed model in managing large-scale and high-dimensional datasets. These models are chosen because they are widely used in real-world scenarios and are often used in the body of current research. Figure 5 shows each model's performance based on the preprocessed training and testing dataset, ensuring that the class imbalance and other data quality issues have been adequately addressed.

With only 0.5 accuracy on the training set, it is clear that the GNB model demonstrates the lowest classification accuracy when confronted with extremely imbalanced data. This outcome is explained by the Naive Bayesian model's inherent strength to handle small sample sizes, which is diminished when applied to large-scale datasets and results in poor generalisation performance. These observations align with findings reported in previous studies (Huang and Li 2011; Tzanos, Kachris, and Soudris 2019). Although RF and XGBoost attain high classification accuracy, their practicality for real-time operational deployment is limited by their substantial memory and compute resource requirements. Conversely, the LGBM model demonstrates strong predictive performance combined with superior computational efficiency and lower memory consumption (Al Daoud 2019). Consequently, the LGBM model is identified as achieving outstanding overall performance.



Figure 5. Model performance on the training set (a) and testing set (b).

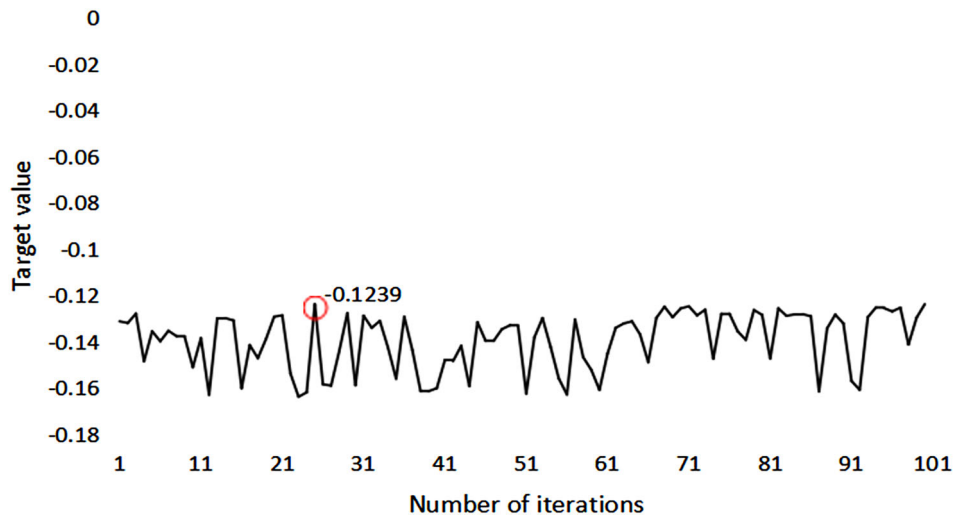


Figure 6. Iterations to reach minimum training MSE.

Through continuous iterative optimisation utilising Bayesian optimisation, an optimal hyperparameter set is identified to further improve the LGBM model's performance. The MSE obtained from 5-fold cross-validation, evaluated within a predefined range of hyperparameters, is the goal function for the optimisation process. The optimisation procedure is designed to include 50 random initial iterations and 100 controlled iterations in order to strike the appropriate balance between optimisation efficiency and prediction accuracy. Figure 6 displays the training set's MSE curve in relation to the target variable. Notably, the Bayesian optimisation process achieves its optimal value, -0.1239 , at the 25th iteration. It is important to highlight that, although the number of iterations appears relatively modest, further iterations cannot yield appreciable improvements, indicating that the selected iteration count is suitable for the specific characteristics of this problem.

The LGBM model's performance with and without Bayesian optimisation, is compared and presented in Figure 7. According to the results, the Bayesian-optimised LGBM (BO-LGBM) model significantly outperforms the baseline LGBM model in terms of backorder risk prediction, reaching better accuracy, recall, and AUC values all exceeding 0.8.

Figure 8 compares the operation times across all evaluated models, highlighting the BO-LGBM model's notable efficiency with an operational time of just 125 s, the shortest among all. Despite RF showing a high AUC score, its significantly longer running time and higher memory usage make it less suitable for agile business environments. In addition, despite XGBoost is a very effective implementation of the GBDT, its performance remains inferior to that of BO-LGBM.

The BO-LGBM model demonstrates strong predictive capability for backorder risk, particularly in contexts characterised by large sample sizes, high-dimensional

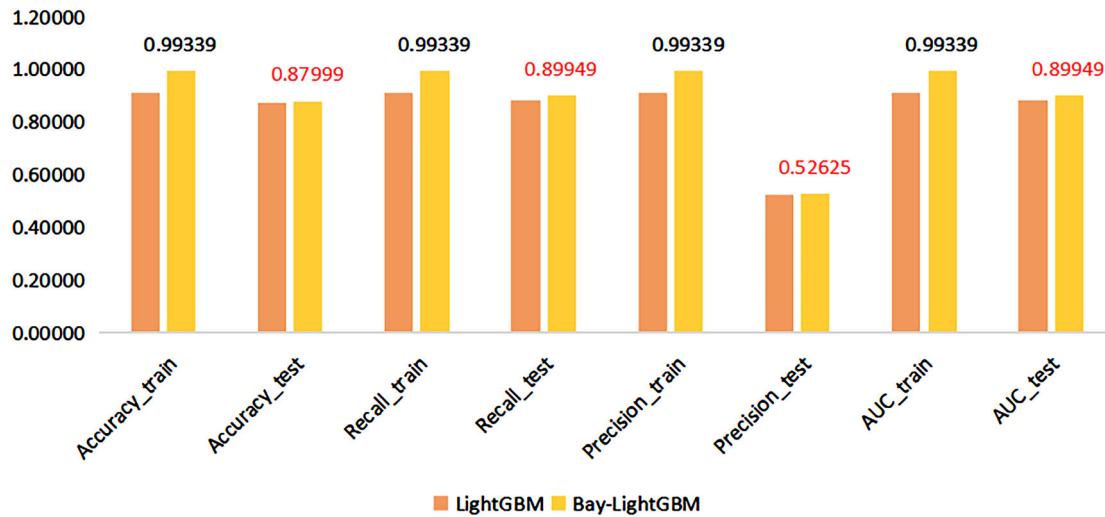


Figure 7. Comparison of LGBM and BO-LGBM performance.

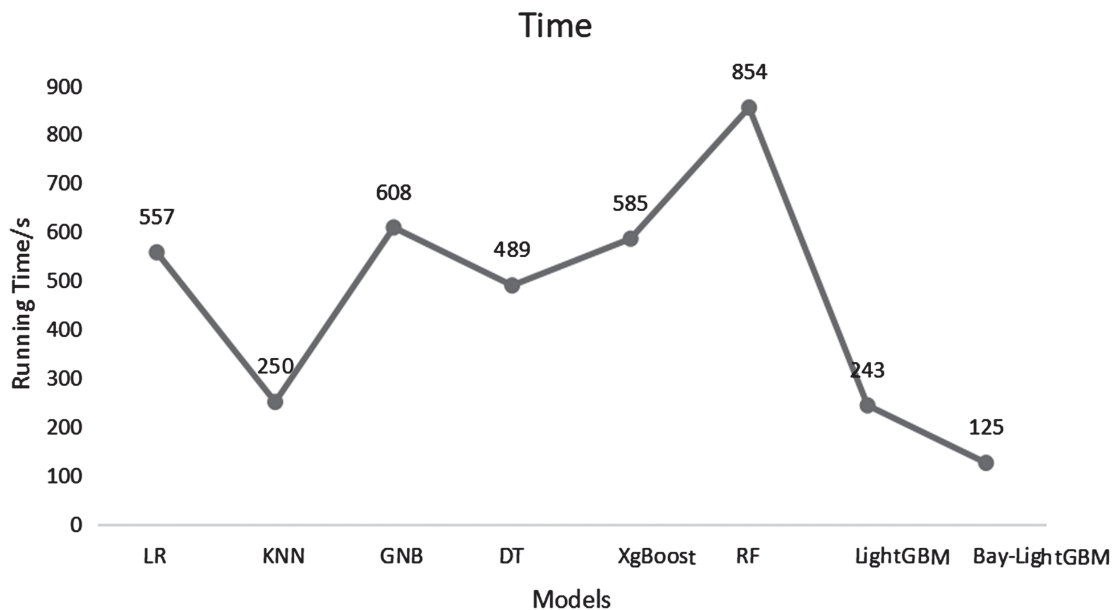


Figure 8. Operation time comparison.

feature spaces, and imbalanced datasets. These results imply that the BO-LGBM model provides substantial benefits by enabling businesses to proactively manage backorder risks through accurate and efficient predictions. The model's enhanced reliability and quick runtime support real-time strategic inventory adjustments, improved resource allocation, and supplier management, thereby reducing potential disruptions and enhancing customer satisfaction. Ultimately, this facilitates more resilient and cost-effective supply chain operations, especially critical in environments characterised by high uncertainty and rapid changes.

4.2. Case study for the hybrid KNNMTD-optimised stacking ensemble model

The empirical analysis is conducted using data from Chinese end-of-life vehicles (ELVs) sector spanning 2006 to 2020 in order to validate the effectiveness of the hybrid KNNMTD-optimised Stacking ensemble model. Initially, seven individual machine learning models, namely support vector regression (SVR), gradient boosting regression tree (GBRT), extreme gradient boosting (XGBoost), decision tree (DT), random forest (RF), and Multi-layer perceptrons (MLP), are trained to identify

Table 4. Definition of features.

Features	Definition	Data Sources
Licensed drivers	Number of licensed vehicle drivers	(NBS China National Bureau of Statistics 2025)
Urban income	Per capita urban income	
Passenger turnover	Passenger turnover	
Highway miles	Total highway miles	
Population	Population size	(AAM China Association of Automobile Manufacturers 2024)
ELVs enterprises	Number of ELVs enterprises	
Auto production	Number of automobiles produced	(NRRA China National Resources Recycling Association 2025)
Recycled prices	ELVs recycled material prices	

base learners with strong predictive performance and minimal correlation. Grid search combined with time series cross-validation (TSCV) is applied to optimise each base model's hyperparameters. Subsequently, the selected optimised base learners are combined with a meta-learner to construct the proposed Stacking ensemble.

4.2.1. Data collection

Eight socioeconomic factors influencing recycled ELVs quantities have been identified based on prior studies (Xia, Han, and Milisavljevic-Syed 2022; Xia, Han, and Milisavljevic-Syed 2023a). Table 4 provides a detailed description of the features of the dataset. Data related to these factors are collected from authoritative sources, including the China Resources Recycling Association (NRRA), the China National Bureau of Statistics (NBS), and the China Association of Automobile Manufacturers (AAM).

4.2.2. Data augmentation

To enhance the generalisation ability and predictive performance of the machine learning models for small datasets, a data augmentation method called KNNMTD method is used to generate artificial virtual samples. A detailed explanation of the KNNMTD method is provided in a previous study (Sivakumar et al. 2022). The number of these samples is set at 100 to avoid the production of irrational virtual samples (Li et al. 2013). The PCD is calculated for varying values of k ranging from 3 to 10, with the optimal value determined to be 4. These artificial samples are then integrated with the original data to construct an augmented dataset. The original dataset and the new dataset are used to train seven distinct machine learning models (i.e. DT, GBRT, MLP, RF, LGBM, XGBoost, SVR). Model performance is evaluated based on MAE, MSE, and R^2 , with the results presented in Figure 9. As depicted, the KNNMTD method considerably improves prediction accuracy across all evaluated metrics, underscoring its effectiveness for small datasets.

4.2.3. Stacking ensemble model

To determine the optimal parameter combinations for each machine learning models using the KNNMTD method, a combination of grid search (GS) and time series cross-validation (TSCV) is employed. Table 5 provides specifics on the optimal hyperparameters.

During the construction of the base-learner layer of the Stacking ensemble, it is crucial to select heterogeneous single machine learning models that demonstrate superior learning performance. This selection is important because a lower correlation among the base models typically results in reduced variance within the Stacking model. Given that models such as RF, XGBoost, GBRT, DT, and LGBM are all tree-based and share similar learning mechanisms, this could lead to higher correlation among them. To address this, the base-learner layer of the Stacking ensemble comprises SVR, MLP, and five independently developed tree-based models, thereby ensuring model diversity and robustness.

Additionally, the meta-learner layer of the Stacking ensemble, it is essential to employ a model with robustness and strong generalisation capability to mitigate overfitting and enhance prediction accuracy. Thus, linear regression is chosen as the meta-learner. Five distinct Stacking models are formulated, namely Stacking 1 (SVR, MLP, and XGBoost), Stacking 2 (SVR, MLP, and LGBM), Stacking 3 (SVR, MLP, and RF), Stacking 4 (SVR, MLP, and DT), and Stacking 5 (SVR, MLP, and GBRT). The predictive performance of these ensemble models is depicted as Figure 10. Consequently, after evaluating the prediction performance and model diversity, the Stacking 1 model, integrating SVR, MLP, and XGBoost as base learners with linear regression as the meta-learner, exhibits the highest predictive accuracy, achieving the lowest MAE and MSE, along with the highest R^2 value.

These results highlight the Stacking 1 model's capability to accurately forecast recycled end-of-life product quantities, enabling better planning of resource allocation, operational scheduling, and inventory management. Accurate predictions facilitate timely and informed managerial decisions, minimising unnecessary inventory holding costs, optimiznsng the use of resources, and improving overall operational efficiency in remanufacturing processes. This proposed model, therefore,

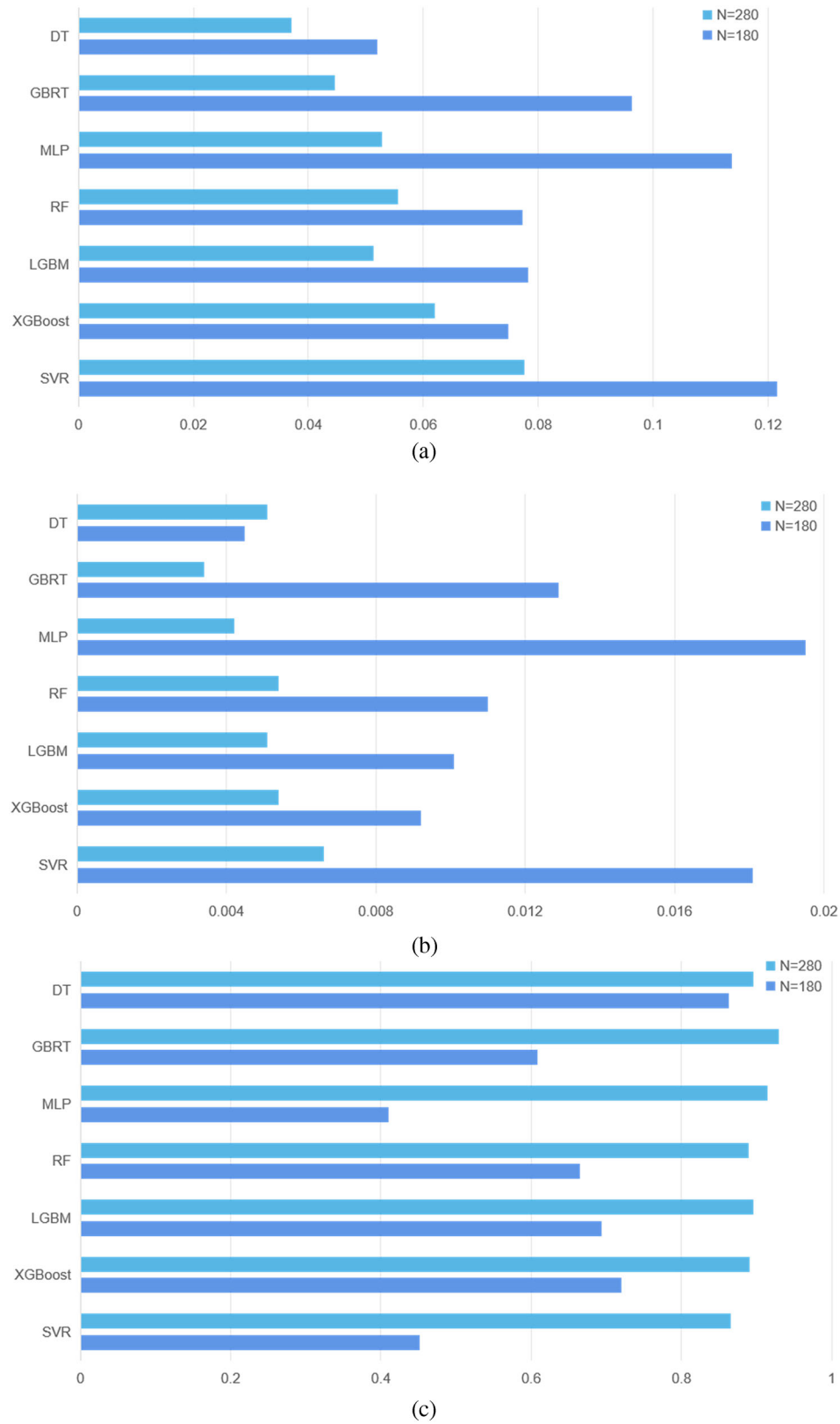
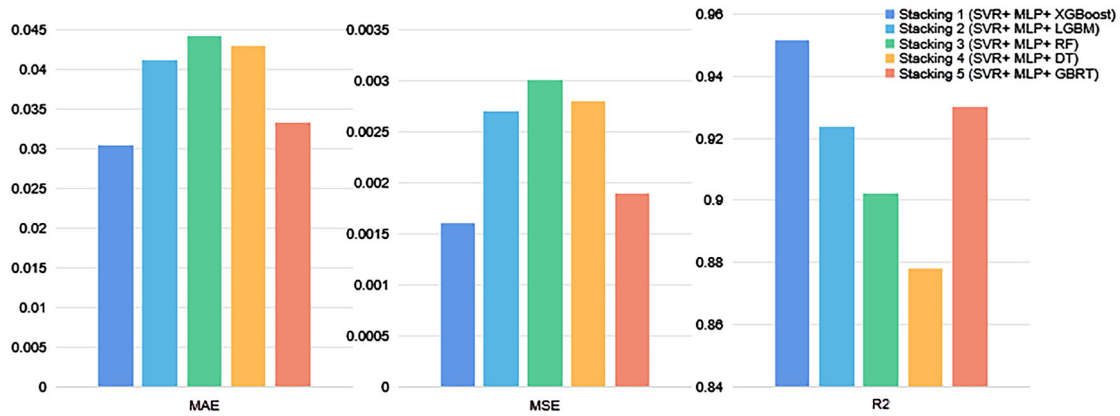


Figure 9. Evaluation metrics for single machine learning models: MAE (a), MSE (b), R^2 (c).

Table 5. Optimal hyperparameter setting.

Model	Parameter	Set	Model	Parameter	Set
SVR	Kernel	'RBF'	RF	n_estimators	20
	gamma	0.1		min_samples_split	3
	C	1.5		max_depth	5
	epsilon	0.1		min_samples_leaf	1
XGBoost	max_depth	5	MLP	max_features	3
	learning_rate	0.03		hidden_layer_sizes	40
	min_child_weight	10		alpha	0.001
	gamma	0		learning_rate_init	0.01
	colsample_bytree	1	GBRT	n_estimators	40
LGBM	subsample	1		max_features	0.5
	n_estimators	120		max_depth	5
	learning_rate	0.04	min_samples_leaf	min_samples_leaf	5
	max_depth	5		learning_rate	0.09
	colsample_bytree	1		subsample	0.5
	min_split_gain	0.01	DT	max_depth	5
	subsample	0.5		min_samples_split	3
	num_leaves	10		min_samples_leaf	1

**Figure 10.** Prediction results of various Stacking models.

not only supports sustainability but also enhances the economic performance and agility of remanufacturing operations.

5. Discussion

This research develops an integrated machine learning-based predictive framework to effectively address customer-driven E2E supply chain uncertainty by jointly modelling forward and reverse supply chains processes. This integrated approach fundamentally advances existing research by proposing a solution that combines two distinct models, namely a Bayesian-optimised LGBM for backorder prediction in the forward supply chain and a KNNMTD-optimised Stacking ensemble model for forecasting recycled end-of-life quantities in the reverse supply chain. While these models address different aspects of supply chain uncertainty, their combination within a unified predictive framework creates a more holistic solution. Previous studies (Al-Omran and Khan 2024; Jassim et al. 2023; Lu et al. 2022; Ntakolia et al. 2021; Shajalal, Hajek, and Abedin 2023) have typically treated

these domains independently, focusing on either backorder prediction or recycling quantity forecasting. Unlike prior studies that generally address these domains separately, this research aims to demonstrate the potential of an integrated solution. By leveraging insights from both forward and reverse supply chains, the proposed framework offers a comprehensive, data-driven approach that bridges the gap between conventional separate models.

5.1. Summary of key findings

In the forward supply chain, the Bayesian-optimised LGBM model is employed to predict backorder risks with high accuracy and efficiency. Traditional approaches to managing backorders typically rely on deterministic forecasting methods or rule-based systems, which often lack the capacity to generalise effectively in the face of real-world data irregularities such as high dimensionality, class imbalance, and non-stationarity (Islam and Amin 2020). The proposed model not only outperformed these conventional methods but also demonstrated a higher degree of robustness and adaptability when applied to complex and volatile datasets.

Bayesian optimisation played a pivotal role in fine-tuning the LGBM model's hyperparameters, ensuring optimal performance without excessive computational costs. This finding is aligned with the results of Eriksson et al. (2019), who highlight the benefits of automated hyperparameter tuning in enhancing the reliability of predictive models in supply chain environments. More importantly, the model's superior AUC and recall scores indicate its practical utility in minimising false negatives, an essential feature in risk management where failing to predict a backorder can result in severe operational disruptions and customer dissatisfaction. The LGBM model's ability to anticipate such risks supports proactive inventory planning, resource reallocation, and demand forecasting, reinforcing the idea that predictive analytics can transform reactive logistics operations into strategic decision-making systems (Oyewole et al. 2024). Moreover, this aligns with the emerging concept of supply chain resilience, which prioritises agility, visibility, and predictive capacity as key attributes for surviving disruptions, whether due to demand shocks, supply shortages, or global crises such as the COVID-19 pandemic (Ivanov 2020).

To effectively manage customer-driven E2E supply chain uncertainty within the reverse supply chain, accurately predicting recycled end-of-life product quantities is essential. Unlike traditional prediction methods, which typically show limited accuracy when handling nonlinear data (Agrawal and Singh 2019; Lin et al. 2018), a hybrid KNNMTD-optimised Stacking ensemble model is proposed to address this issue in this research. The proposed model's effectiveness is assessed using data from Chinese ELVs sector spanning the period from 2006 to 2020.

The findings clearly indicate that single machine learning models exhibit improved prediction performance after optimisation via the KNNMTD method coupled with parameter tuning based on grid search and time series cross-validation. Additionally, this research extends existing literature by comprehensively investigating optimal combinations of diverse machine learning methods within a Stacking ensemble model. Previous studies primarily rely on single-model forecasts or simpler ensemble methods, such as Bagging or Boosting, which are often prone to overfitting due to high correlations among base learners (Mudannayake et al. 2022). In contrast, this research emphasises the importance of model heterogeneity to minimise correlation among base learners, thus enhancing predictive robustness and accuracy. More specifically, recognising the potential high correlation among tree-based models (XGBoost, LGBM, GBRT, DT, RF), this research strategically integrates these models with structurally distinct models (i.e. SVR, MLP).

Five different Stacking models are evaluated, each combining different base-model configurations. Ultimately, the Stacking model incorporating XGBoost, MLP, and SVR as base learners and linear regression as the meta-learner, demonstrated superior performance, achieving the lowest prediction error and highest R-squared value.

This finding aligns with previous research that highlights the superior performance of stacking ensembles due to their capability to integrate diverse model strengths effectively (Hoxha et al. 2023). This superior predictive capability enables remanufacturers to enhance the organisation of remanufacturing processes by providing accurate forecasts of available recycled materials, thus ensuring a steady flow of materials and optimising remanufacturing schedules.

5.2. Theoretical contribution and practical implication

Overall, previous research typically addresses forward supply chain issues (e.g. backorder risks) and reverse supply chain issues (e.g. recycled end-of-life product quantities) independently, often overlooking critical interdependencies between these processes (Shajalal, Hajek, and Abedin 2023; Zhang et al. 2022). The fundamental innovation of this research lies in its modelling of the inherent interconnectedness of forward and reverse supply chains processes. By integrating these predictive tasks into a cohesive two-step methodological framework, this research advances existing literature through a more holistic and realistic approach to manage supply chain uncertainties. Such integration enables manufacturers and remanufacturers to collaboratively optimise inventory levels, resource allocation, and production scheduling, effectively addressing the identified research gap of separate treatment highlighted in the literature review. Consequently, this research contributes significantly to the advancement of predictive analytics in managing customer-driven E2E supply chain uncertainties.

The integrated model helps advance the goals of the circular economy within the industrial sector, making it more adaptable and responsive to environmental and market changes. Specifically, the application of integrated machine learning-based model promotes circularity in manufacturing processes by improving the integration of recycled materials back into production cycles, thus supporting sustainable practices. Furthermore, the economic implications of such integration are substantial, including measurable reductions in operational costs, improved efficiency in supply chain planning, and enhanced responsiveness to market changes.

Additionally, by promoting material reuse and recycling, the integrated model economically incentivizes circular practices, minimising waste-related costs, and aligning industrial operations with evolving market and regulatory demands. Consequently, this research significantly advances predictive analytics in supply chain management, delivering tangible economic benefits and supporting sustainable, resilient, and human-centric industrial practices.

This research contributes to the management of customer-driven E2E supply chain uncertainties by demonstrating the importance of integrating forward and reverse supply chains. By providing actionable insights through advanced predictive models, this research offers clear guidance for supply chain managers on how to optimise inventory, enhance production efficiency, and support sustainability in supply chain operations. Moreover, this research can help manufacturers and remanufacturers make more informed and timely decisions, optimising production and remanufacturing processes, allocating resources more effectively, and ultimately aligning supply with demand. This approach not only reduces backorder occurrences and their associated costs but also promotes a more sustainable and circular value chain by efficiently handling product returns, thereby supporting the resilience and sustainability goals emphasised in Industry 5.0.

5.3. Limitation

This research contains some limitations. While the proposed Bayesian-optimised LGBM model is designed for efficiency, scalability challenges may arise as data volumes increase or when applied across highly intricate supply chains with multiple interdependencies. Furthermore, the model's effectiveness relies heavily on the quality and completeness of the input data, making it vulnerable to errors or biases introduced by missing or inaccurate information. Additionally, regarding the data handling strategy, this research employs a random under-sampling method yielding a 1:42 class ratio. Although this trade-off is justified to prevent information loss, the remaining imbalance may still bias the classifier slightly towards the majority class compared to advanced synthetic sampling techniques. For the proposed hybrid KNNMTD-optimised Stacking ensemble model, it is necessary to compare the performance with other ensemble models. The influence factors, such as customer preferences and economic policy, are not considered in this research. Finally, the interpretability of machine learning models, such as LGBM and stacking models, can pose challenges, particularly for stakeholders who require transparent insights into decision-making processes.

6. Conclusion

The transition from Industry 4.0 to Industry 5.0, with its emphasis on sustainability, human-centricity, and resilience, highlights the urgent need for research into customer-driven E2E supply chain uncertainty to ensure adaptability and success in an evolving manufacturing environment. Addressing this uncertainty, which includes managing backorder risks and handling recycled end-of-life product quantities, is crucial for promoting circularity in manufacturing processes, thus supporting sustainable practices and enhancing supply chain resilience against disruptions. An integrated machine learning-based model is proposed to mitigate customer-driven E2E supply chain uncertainty. This model combines a Bayesian-optimised LGBM to predict backorder risks and a KNNMTD-optimised Stacking ensemble model to forecast recycled end-of-life product quantities. Case studies reveal the proposed integrated model significantly improves prediction performance and generalisation capability for managing customer-driven supply chain uncertainties, which potentially enhances supply chain resilience and aligns supply with demand.

The findings of this research contribute to more resilient and efficient supply chains by reducing uncertainty and promoting sustainable practices. The models support the goals of the circular economy by integrating recycled materials back into production processes and enhancing the adaptability of supply chains to disruptions. Future work could focus on explicitly modelling the interdependencies between the forward and reverse processes to make the coupling more operational, further enhancing the predictive power of the system. Furthermore, future research should explore advanced data balancing techniques, such as synthetic minority over-sampling technique (SMOTE) or adaptive synthetic (ADASYN) sampling, as well as algorithmic approaches like cost-sensitive learning or focal loss functions, to further address class imbalance without the information loss associated with under-sampling. Additionally, future research could focus on extending the models to different industries and supply chain contexts to enhance their generalizability. Incorporating real-time data streams and adaptive learning algorithms could improve the models' ability to respond dynamically to disruptions and market changes. Another area of focus is enhancing the explainability and interpretability of the models to foster trust and facilitate adoption by industry stakeholders.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors



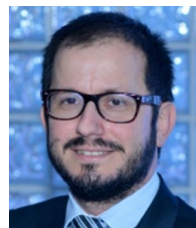
Hanbing Xia is currently a Postdoctoral Researcher at South China University of Technology, China. Prior to this role, Dr. Xia completed a PhD in Manufacturing at the Sustainable Manufacturing Systems Centre, Faculty of Engineering and Applied Science, Cranfield University, United Kingdom, in 2025, under the supervision of Dr. Jelena Milisavljevic-Syed and Prof. Konstantinos Salonitis. Dr. Xia received a Master's degree in Logistics Management from Jilin University, China, in 2019, and a Bachelor's degree in Logistics Management from the same institution in 2017. Xia's research focuses on the optimization of reverse supply chains for end-of-life products, developing computational frameworks that integrate machine learning, blockchain technology, and digital product passports to enable sustainable circular supply chains. This work explores how supply chain systems, particularly in manufacturing environment, can be transformed towards Industry 5.0 by addressing uncertainties in demand forecasting, logistics network design, and resource recovery. Xia has published seven journal papers and five conference papers in venues such as Transportation Research Part E, Resources, Conservation and Recycling, Journal of Cleaner Production, International Journal of Production Research, and ASME IDETC-CIE. Dr. Xia is a member of ASME and serves as a reviewer for the Journal of Manufacturing Systems, Journal of Cleaner Production, Cleaner Logistics and Supply Chain, and Procedia CIRP.



Shehu Sani received his PhD in 2025 from Cranfield University, United Kingdom, where his research focused on enhancing resilience in supply chains under conditions of disruption and uncertainty. His work focus on the development of decision-support models that generate satisficing solution to facilitate timely decision-making during unexpected disruptions, particularly through the application of the Compromise Decision Support (cDSP) construct and Decision Support in the Design of Engineered Systems (DSIDES) framework. His research has been published in several peer-reviewed scholarly journals addressing supply chain resilience, decision analytics and engineered systems design. He previously obtained a Masters' Degree in Mechanical Engineering (Energy Studies) from the Federal University of Agriculture, Makurdi, Nigeria. Dr. Sani is a registered member of the Council for the Regulation of Engineering in Nigeria (COREN) and a member of both the Nigerian Society of Engineers (NSE) and the American Society of Mechanical Engineers (ASME).



Jelena Milisavljevic-Syed assumed the role of Associate Professor in Manufacturing Systems at Cranfield University within the Sustainable Manufacturing Systems Centre in September 2022. Prior to her current role, Dr Milisavljevic-Syed held an Assistant Professor position in Industrial Design at the University of Liverpool and a leadership position as a Director of the Systems Realization Laboratory (SRL) since 2022. Dr Milisavljevic-Syed received her PhD in Mechanical Engineering in 2018 from the University of Oklahoma, Norman, USA. Her research focuses on the design of sustainable evolving service systems, developing theories and computational frameworks for cyber-physical-social systems (eCPSS) that integrate human, artificial, and organisational intelligence. Her work explores how complex service systems – particularly in healthcare and manufacturing – can evolve safely and sustainably by addressing asymmetries in information, intelligence, and resources. She contributes to emerging design paradigms for complex socio-technical systems, including the development of Multi-Intelligence Agents (MIAs) for coordinating decision-making across distributed actors and enabling resilient, equitable, and adaptive service systems. Milisavljevic-Syed has co-authored a monograph and published over forty journal and forty-one technical publications, as well as several magazine articles, addressing the dynamic management of manufacturing systems and the design and analysis of complex systems. She is an Honorary Research Fellow at the School of Electrical Engineering and Electronics, University of Liverpool, and an active member of several professional organisations, including IEEE, MIET, ASME, ASEE, the Design Society, and the EPSRC Peer Review College.



Konstantinos Salonitis is a Professor of Manufacturing Systems and the Head of Sustainable Manufacturing Systems Centre at Cranfield University (UK). He has a PhD in manufacturing processes modelling (2006) from the University of Patras (Greece). He has a research track record of more than 25 years, having secured research funding exceeding £10M. He has published more than 350 research papers in major international journals and at internationally refereed conferences and has authored two books. He has supervised more than 25 PhD students to completion. His current research focuses on the sustainability of manufacturing systems and processes, the design and optimization of manufacturing systems, and the use of lean and green approaches to manage manufacturing. He is a Chartered Engineer and a Fellow of the IMechE, the IET and the HEA.

Data availability statement

The data sources are listed in this paper.

ORCID

Hanbing Xia <https://orcid.org/0000-0003-1815-1045>

Jelena Milisavljevic-Syed <http://orcid.org/0000-0003-2957-8933>

References

- AAM China Association of Automobile Manufacturers. Accessed May 5, 2024. www.caam.org.cn.
- Abbasi, M., M. Abduli, B. Omidvar, and A. Baghvand. 2014. "Results Uncertainty of Support Vector Machine and Hybrid of Wavelet Transform-support Vector Machine Models for Solid Waste Generation Forecasting." *Environmental Progress & Sustainable Energy* 33 (1): 220–228. <https://doi.org/10.1002/ep.11747>.
- Abbasi, M., and A. El Hanandeh. 2016. "Forecasting Municipal Solid Waste Generation Using Artificial Intelligence Modelling Approaches." *Waste Management* 56:13–22. <https://doi.org/10.1016/j.wasman.2016.05.018>.
- Adamović, V. M., D. Z. Antanasijević, MĐ Ristić, A. A. Perić-Grujić, and V. V. Pocajt. 2017. "Prediction of Municipal Solid Waste Generation using Artificial Neural Network Approach Enhanced by Structural Break Analysis." *Environmental Science and Pollution Research* 24 (1): 299–311. <https://doi.org/10.1007/s11356-016-7767-x>.
- Agrawal, S., and R. K. Singh. 2019. "Forecasting Product Returns and Reverse Logistics Performance: Structural Equation Modelling." *Management of Environmental Quality: An International Journal* 31 (5): 1223–1237. <https://doi.org/10.1108/meq-05-2019-0109>.
- Akanbi, L. A., A. O. Oyedele, L. O. Oyedele, and R. O. Salami. 2020. "Deep Learning Model for Demolition Waste Prediction in a Circular Economy." *Journal of Cleaner Production* 274:122843. <https://doi.org/10.1016/j.jclepro.2020.122843>.
- Al Daoud, E. 2019. "Comparison between XGBoost, LightGBM and CatBoost Using a Home Credit Dataset." *International Journal of Computer Information Engineering* 13 (1): 6–10. <https://doi.org/10.52344/hfr.2021.4.0.33>.
- Al-Omran, K., and E. Khan. 2024. "Predicting Medical Waste Generation and Associated Factors Using Machine Learning in the Kingdom of Bahrain." *Environmental Science and Pollution Research* 31 (26): 38343–38357. <https://doi.org/10.1007/s11356-024-33773-1>.
- Altin, F. G., İ Budak, and F. Özcan. 2023. "Predicting the Amount of Medical Waste Using Kernel-Based SVM and Deep Learning Methods for a Private Hospital in Turkey." *Sustainable Chemistry and Pharmacy* 33:101060. <https://doi.org/10.1016/j.scp.2023.101060>.
- Baryannis, G., S. Validi, S. Dani, and G. Antoniou. 2019. "Supply Chain Risk Management and Artificial Intelligence: State of the art and Future Research Directions." *International Journal of Production Research* 57 (7): 2179–2202. <https://doi.org/10.1080/00207543.2018.1530476>.
- Bhatnagar, R., and A. S. Sohal. 2005. "Supply Chain Competitiveness: Measuring the Impact of Location Factors, Uncertainty and Manufacturing Practices." *Technovation* 25 (5): 443–456. <https://doi.org/10.1016/j.technovation.2003.09.012>.
- Björk, K.-M. 2009. "An Analytical Solution to a Fuzzy Economic Order Quantity Problem." *International Journal of Approximate Reasoning* 50 (3): 485–493. <https://doi.org/10.1016/j.ijar.2008.10.001>.
- Cha, G.-W., Y.-C. Kim, H. J. Moon, and W.-H. Hong. 2017. "New Approach for Forecasting Demolition Waste Generation Using chi-squared Automatic Interaction Detection (CHAID) Method." *Journal of Cleaner Production* 168:375–385. <https://doi.org/10.1016/j.jclepro.2017.09.025>.
- Cha, G.-W., H. J. Moon, and Y.-C. Kim. 2022. "A Hybrid Machine-Learning Model for Predicting the Waste Generation Rate of Building Demolition Projects." *Journal of Cleaner Production* 375:134096. <https://doi.org/10.1016/j.jclepro.2022.134096>.
- Chen, C., Q. Zhang, Q. Ma, and B. Yu. 2019. "LightGBM-PPI: Predicting Protein-Protein Interactions through LightGBM with Multi-information Fusion." *Chemometrics and Intelligent Laboratory Systems* 191:54–64. <https://doi.org/10.1016/j.chemolab.2019.06.003>.
- Chilmon, B., and N. S. Tipi. 2020. "Modelling and Simulation Considerations for an end-to-end Supply Chain System." *Computers & Industrial Engineering* 150:106870. <https://doi.org/10.1016/j.cie.2020.106870>.
- Christopher, M. 2022. *Logistics and Supply Chain Management*. Harlow: Pearson UK.
- Coskuner, G., M. S. Jassim, M. Zontul, and S. Karateke. 2020. "Application of Artificial Intelligence Neural Network Modeling to Predict the Generation of Domestic, Commercial and Construction Wastes." *Waste Management and Research* 39 (3): 499–507. <https://doi.org/10.1177/073424220935181>.
- Crandall, R. E., W. R. Crandall, and C. C. Chen. 2009. *Principles of Supply Chain Management*. Boca Raton: CRC Press.
- de Lima, F. A., S. Seuring, and P. C. Sauer. 2022. "A Systematic Literature Review Exploring Uncertainty Management and Sustainability Outcomes in Circular Supply Chains." *International Journal of Production Research* 60 (19): 6013–6046. <https://doi.org/10.1080/00207543.2021.1976859>.
- Disney, S. M., and D. R. Towill. 2003. "On the Bullwhip and Inventory Variance Produced by an Ordering Policy." *Omega* 31 (3): 157–167. [https://doi.org/10.1016/s0305-0483\(03\)00028-8](https://doi.org/10.1016/s0305-0483(03)00028-8).
- Eriksson, D., M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek. 2019. "Scalable Global Optimization via Local Bayesian Optimization." *Advances in Neural Information Processing Systems* 32:9849–9860.
- Feng, G., L. Chen-Yu, X. Feng-Lei, and L. Wei-Ling. 2012. "Demand Prediction of LRU Parts with Backorder for SRU." In *2012 Fifth International Symposium on Computational Intelligence and Design*, edited by L. Sun and Y. Tang, 530–532. Hangzhou, China: IEEE.
- Flynn, B. B., X. Koufteros, and G. Lu. 2016. "On Theory in Supply Chain Uncertainty and Its Implications for Supply Chain Integration." *Journal of Supply Chain Management* 52 (3): 3–27. <https://doi.org/10.1111/jscm.12106>.
- Frazier, P. I. 2018. "Bayesian Optimization." In *Recent Advances in Optimization and Modeling of Contemporary Problems*, edited by E. Gel and L. Ntamo, 255–278. Catonsville, MD: Informs.
- Galang, M. G. K., and F. Ballesteros Jr. 2018. "Estimation of Waste Mobile Phones in the Philippines Using Neural Networks." *Global NEST Journal* 20 (4): 767–772. <https://doi.org/10.30955/gnj.002534>.
- Hajek, P., and M. Z. Abedin. 2020. "A Profit Function-Maximizing Inventory Backorder Prediction System Using big Data Analytics." *IEEE Access* 8:58982–58994. <https://doi.org/10.1109/access.2020.2983118>.
- Handfield, R. 2002. *Supply Chain Redesign: Transforming Supply Chains into Integrated Value Systems*. Upper Saddle River, NJ: FT Press.

- He, H., and E. A. Garcia. 2009. "Learning from Imbalanced Data." *IEEE Transactions on Knowledge and Data Engineering* 21 (9): 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>.
- Hichri, A. B. 2016. *Examining the Moderating Effect of Demand Uncertainty on Supply Chain and Technology Strategies and Performance*. Prescott, AZ: Northcentral University.
- Hoxha, J., M. Y. Çodur, E. Mustafaraj, H. Kanj, and A. El Masri. 2023. "Prediction of Transportation Energy Demand in Türkiye using Stacking Ensemble Models: Methodology and Comparative Analysis." *Applied Energy* 350:121765. <https://doi.org/10.1016/j.apenergy.2023.121765>.
- Hoy, Z. X., K. S. Woon, W. C. Chin, H. Hashim, and Y. Van Fan. 2022. "Forecasting Heterogeneous Municipal Solid Waste Generation via Bayesian-optimised Neural Network with Ensemble Learning for Improved Generalisation." *Computers & Chemical Engineering* 166:107946. <https://doi.org/10.1016/j.compchemeng.2022.107946>.
- Huang, Y., and L. Li. 2011. "Naive Bayes Classification Algorithm Based on Small Sample set." 2011 IEEE International Conference on Cloud Computing and Intelligence Systems. IEEE, 34–39.
- Islam, S., and S. H. Amin. 2020. "Prediction of Probable Backorder Scenarios in the Supply Chain Using Distributed Random Forest and Gradient Boosting Machine Learning Techniques." *Journal of Big Data* 7:1–22. <https://doi.org/10.1186/s40537-020-00345-2>.
- Ivanov, D. 2020. "Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-based Analysis on the Coronavirus Outbreak (COVID-19/SARS-CoV-2) Case." *Transportation Research Part E: Logistics and Transportation Review* 136:101922. <https://doi.org/10.1016/j.tre.2020.101922>.
- Jafari, M., and E. Mousavi. 2024. "Machine Learning-Based Prediction of Construction and Demolition Waste Generation in Developing Countries: A Case Study." *Environmental Science and Pollution Research* 32 (32): 19562–19573. <https://doi.org/10.1007/s11356-024-34527-9>.
- Jassim, M. S., G. Coskuner, N. Sultana, and S. Z. Hosain. 2023. "Forecasting Domestic Waste Generation during Successive COVID-19 Lockdowns by Bidirectional LSTM Super Learner Neural Network." *Applied Soft Computing* 133:109908. <https://doi.org/10.1016/j.asoc.2022.109908>.
- Kazemi, N., E. Ehsani, and M. Y. Jaber. 2010. "An Inventory Model with Backorders with Fuzzy Parameters and Decision Variables." *International Journal of Approximate Reasoning* 51 (8): 964–972. <https://doi.org/10.1016/j.ijar.2010.07.001>.
- Kontokosta, C. E., B. Hong, N. E. Johnson, and D. J. Starobin. 2018. "Using Machine Learning and Small Area Estimation to Predict Building-Level Municipal Solid Waste Generation in Cities." *Computers, Environment and Urban Systems* 70:151–162. <https://doi.org/10.1016/j.compenvurbysys.2018.03.004>.
- Lee, H. L. 2004. "The Triple-A Supply Chain." *Harvard Business Review* 82 (10): 102–113.
- Li, D. C., W. T. Huang, C. C. Chen, and C. J. Chang. 2013. "Employing Virtual Samples to Build Early High-Dimensional Manufacturing Models." *International Journal of Production Research* 51 (11): 3206–3224. <https://doi.org/10.1080/00207543.2012.746795>.
- Li, K., and H. Ma. 2019. "Prediction of Municipal Solid Waste Generation with Elman Neural Network—Case Study: Shanghai City in China." 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 1174–1177. IEEE.
- Lin, H.-T., K. Nakajima, E. Yamasue, and K. Ishihara. 2018. "Recycling of End-of-Life Vehicles in Small Islands: The Case of Kinmen, Taiwan." *Sustainability* 10 (12): 4377–4390. <https://doi.org/10.3390/su10124377>.
- Liu, H.-M., H.-H. Sun, R. Guo, D. Wang, H. Yu, D. Do Rosario Alves, and W.-M. Hong. 2022. "Prediction of China's Industrial Solid Waste Generation based on the PCA-NARBP Model." *Sustainability* 14 (7): 4294. <https://doi.org/10.3390/su14074294>.
- Lu, W., W. Huo, H. Gulina, and C. Pan. 2022. "Development of Machine Learning Multi-city Model for Municipal Solid Waste Generation Prediction." *Frontiers of Environmental Science & Engineering* 16 (9): 119. <https://doi.org/10.1007/s11783-022-1551-6>.
- Maitra, S., and S. Kundu. 2023. "Backorder Prediction in Inventory Management: Classification Techniques and Cost Considerations." 2023 27th International Computer Science and Engineering Conference (ICSEC), 1–8. Samui Island, Thailand. IEEE.
- Meza, J. K. S., D. O. Yepes, J. Rodrigo-Illarri, and E. Cassiraga. 2019. "Predictive Analysis of Urban Waste Generation for the City of Bogotá, Colombia, through the Implementation of Decision Trees-Based Machine Learning, Support Vector Machines and Artificial Neural Networks." *Heliyon* 5 (11): e02810. <https://doi.org/10.1016/j.heliyon.2019.e02810>.
- Milislavljevic-Syed, J., J. Li, and H. Xia. 2023. "Realisation of Responsive and Sustainable Reconfigurable Manufacturing Systems." *International Journal of Production Research* 62 (8): 2725–2746. <https://doi.org/10.1080/00207543.2023.2233625>.
- Monczka, R. M., R. B. Handfield, L. C. Giunipero, and J. L. Patterson. 2020. *Purchasing and Supply Chain Management*. Boston, MA: Cengage Learning.
- Mudannayake, O., D. Rathnayake, J. D. Herath, D. K. Fernando, and M. Fernando. 2022. "Exploring Machine Learning and Deep Learning Approaches for Multi-step Forecasting in Municipal Solid Waste Generation." *IEEE Access* 10:122570–122585. <https://doi.org/10.1109/ACCESS.2022.3221941>.
- Nahmias, S., and T. L. Olsen. 2015. *Production and Operations Analysis*. Long Grove, IL: Waveland Press.
- NBS China National Bureau of Statistics. Accessed December 5, 2025. <https://www.stats.gov.cn/>.
- Noori, R., A. Karbassi, and M. S. Sabahi. 2010. "Evaluation of PCA and Gamma Test Techniques on ANN Operation for Weekly Solid Waste Prediction." *Journal of Environmental Management* 91 (3): 767–771. <https://doi.org/10.1016/j.jenvman.2009.10.007>.
- NRRA China National Resources Recycling Association. Accessed December 5, 2025. <https://www.crra.org.cn/>.
- Ntakolia, C., C. Kokkoti, P. Karlsson, and S. Moustakidis. 2021. "An Explainable Machine Learning Model for Material Backorder Prediction in Inventory Management." *Sensors* 21 (23): 7926. <https://doi.org/10.3390/s21237926>.
- Oguz-Ekim, P. 2021. "Machine Learning Approaches for Municipal Solid Waste Generation Forecasting." *Environmental Engineering Science* 38 (6): 489–499. <https://doi.org/10.1089/ees.2020.0232>.

- Ojha, R., A. Ghadge, M. K. Tiwari, and U. S. Bititci. 2018. "Bayesian Network Modelling for Supply Chain Risk Propagation." *International Journal of Production Research* 56 (17): 5795–5819. <https://doi.org/10.1080/00207543.2018.1467059>.
- Oyewole, A. T., C. C. Okoye, O. C. Ofodile, and E. Ejairu. 2024. "Reviewing Predictive Analytics in Supply Chain Management: Applications and Benefits." *World Journal of Advanced Research and Reviews* 21 (3): 568–574. <https://doi.org/10.30574/wjarr.2024.21.3.0673>.
- Predictive Backorder Competition. 2025. Accessed March 1, 2025. https://github.com/rodrigasantis1/backorder_prediction.
- Puntarić, E., L. Pezo, Zgorelec, J. Gunjača, D. Kučić Grgić, and N. Voća. 2022. "Prediction of the Production of Separated Municipal Solid Waste by Artificial Neural Networks in Croatia and the European Union." *Sustainability* 14 (16): 10133. <https://doi.org/10.3390/su141610133>.
- Rafrastara, F. A., C. Supriyanto, C. Paramita, Y. P. Astuti, and F. Ahmed. 2023. "Performance Improvement of Random Forest Algorithm for Malware Detection on Imbalanced Dataset using Random Under-sampling Method." *Jurnal Informatika: Jurnal Pengembangan IT* 8 (2): 113–118. <https://doi.org/10.30591/jpit.v8i2.5207>.
- Ramyachitra, D., and P. Manikandan. 2014. "Imbalanced Dataset Classification and Solutions: A Review." *International Journal of Computing Business Research* 5 (4): 1–29.
- Rashid, H. M. M., S. R. M. Yatim, S. Abdullah, M. Y. A. F. Pejabat, I. A. Nazim, and Z. Ahmad. 2022. "Forecasting Medical Waste Generation in Hospital Taiping, Perak Using Nonlinear Autoregressive (NAR) Neural Network Method." 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), 243–247. IEEE.
- Rolf, B., I. Jackson, M. Müller, S. Lang, T. Reggelin, and D. Ivanov. 2023. "A Review on Reinforcement Learning Algorithms and Applications in Supply Chain Management." *International Journal of Production Research* 61 (20): 7151–7179. <https://doi.org/10.1080/00207543.2022.2140221>.
- Sani, S., H. Xia, J. Milisavljevic-Syed, and K. Salonitis. 2023. "Supply Chain 4.0: A Machine Learning-Based Bayesian-Optimized LightGBM Model for Predicting Supply Chain Risk." *Machines* 11 (9): 888. <https://doi.org/10.3390/machines11090888>.
- Satapathy, S. C., K. S. Raju, J. K. Mandal, and V. Bhateja. 2015. *Proceedings of the Second International Conference on Computer and Communication Technologies: IC3 T 2015, Volume 1*. New Delhi: Springer.
- Shajalal, M., P. Hajek, and M. Z. Abedin. 2023. "Product Backorder Prediction Using Deep Neural Network on Imbalanced Data." *International Journal of Production Research* 61 (1): 302–319. <https://doi.org/10.1080/00207543.2021.1901153>.
- Shamshiry, E., M. Mokhtar, A.-M. Abdulai, I. Komoo, and N. Yahaya. 2014. "Combining Artificial Neural Network-Genetic Algorithm and Response Surface Method to Predict Waste Generation and Optimize Cost of Solid Waste Collection and Transportation Process in Langkawi Island, Malaysia." *Malaysian Journal of science* 33 (2): 118–140. <https://doi.org/10.22452/mjs.vol33no2.1>.
- Simangunsong, E., L. C. Hendry, and M. Stevenson. 2012. "Supply-chain Uncertainty: A Review and Theoretical Foundation for Future Research." *International Journal of Production Research* 50 (16): 4493–4523. <https://doi.org/10.1080/00207543.2011.613864>.
- Sivakumar, J., K. Ramamurthy, M. Radhakrishnan, and D. Won. 2022. "Synthetic Sampling from Small Datasets: A Modified Mega-trend Diffusion Approach Using k-Nearest Neighbors." *Knowledge-Based Systems* 236:107687. <https://doi.org/10.1016/j.knsys.2021.107687>.
- Su, D., J. Zheng, Y.-k. Shao, J.-y. Liu, X.-x. Liu, K. Yu, B.-h. Feng, H. Mei, and S. Qin. 2025. "Developing and Validating a Machine Learning-Based Model for Predicting in-hospital Mortality among ICU-Admitted Heart Failure Patients: A Study Utilizing the MIMIC-III Database." *Digital Health* 11:20552076251335705. <https://doi.org/10.1177/20552076251335705>.
- Supply Chain Brain. 2022. Accessed December 5, 2024. <https://www.supplychainbrain.com/supplychainbrain-2022-event-coverage>.
- Tan, M., S. Yuan, S. Li, Y. Su, H. Li, and F. He. 2019. "Ultra-short-term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning." *IEEE Transactions on Power Systems* 35 (4): 2937–2948. <https://doi.org/10.1109/TPWRS.2019.2963109>.
- Tanha, J., Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour. 2020. "Boosting Methods for Multi-class Imbalanced Data Classification: An Experimental Review." *Journal of Big Data* 7 (1): 1–47. <https://doi.org/10.1186/s40537-020-00349-y>.
- Tian, R., Z. X. Hoy, P. Y. Liew, M. M. Hanafiah, G. R. Mong, C. T. Chong, M. U. Hossain, and K. S. Woon. 2023. "Socio-economic Correlation Analysis and Hybrid Artificial Neural Network Model Development for Provincial Waste Electrical and Electronic Equipment Generation Forecasting in China." *Journal of Cleaner Production* 418:138076. <https://doi.org/10.1016/j.jclepro.2023.138076>.
- Tzanos, G., C. Kachris, and D. Soudris. 2019. "Hardware Acceleration on Gaussian Naive Bayes Machine Learning Algorithm." 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAS), 1–5, IEEE.
- Wang, D., Y.-a. Yuan, Y. Ben, H. Luo, and H. Guo. 2022. "Long Short-term Memory Neural Network and Improved Particle Swarm Optimization-based Modeling and Scenario Analysis for Municipal Solid Waste Generation in Shanghai, China." *Environmental Science and Pollution Research* 29 (46): 69472–69490. <https://doi.org/10.1007/s11356-022-20438-0>.
- Wang, Z., M. Zoghi, F. Hutter, D. Matheson, and N. De Freitas. 2013. Bayesian Optimization in High Dimensions via Random Embeddings. *IJCAI*. Beijing, China, 3–9 August. 1778–1784.
- Weiss, G. M., and F. Provost. 2003. "Learning When Training Data Are Costly: The Effect of Class Distribution on Tree Induction." *Journal of Artificial Intelligence Research* 19:315–354. <https://doi.org/10.1613/jair.1199>.
- Wyrembek, M., G. Baryannis, and A. Brintrup. 2025. "Causal Machine Learning for Supply Chain Risk Prediction and Intervention Planning." *International Journal of Production Research* 63 (15): 5629–5648. <https://doi.org/10.1080/00207543.2025.2458121>.

- Xia, H., Z. Chen, J. Milisavljevic-Syed, and K. Salonitis. 2024. "Uncertain Programming Model for Designing Multi-objective Reverse Logistics Networks." *Cleaner Logistics and Supply Chain* 11:100155. <https://doi.org/10.1016/j.clscn.2024.100155>.
- Xia, H., J. Han, and J. Milisavljevic-Syed. 2022. "Forecasting the Number of End-of-Life Vehicles: State of the Art Report." 16th International Design Conference (Online). Croatia. Cambridge University Press.
- Xia, H., J. Han, and J. Milisavljevic-Syed. 2023a. "Predicting the Quantity of Recycled End-of-Life Products Using a Hybrid SVR-Based Model." ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Boston, Massachusetts, USA.
- Xia, H., J. Han, and J. Milisavljevic-Syed. 2023b. "Predictive Modeling for the Quantity of Recycled end-of-life Products Using Optimized Ensemble Learners." *Resources, Conservation and Recycling* 197:107073. <https://doi.org/10.1016/j.resconrec.2023.107073>.
- Xia, H., J. Li, Q. J. Li, J. Milisavljevic-Syed, and K. Salonitis. 2025. "Integrating Blockchain with Digital Product Passports for Managing Reverse Supply Chain." *Transportation Research Part E: Logistics and Transportation Review* 202:104336. <https://doi.org/10.1016/j.tre.2025.104336>.
- Xu, G., J. Yano, and S.-i. Sakai. 2018. "Recycling Potentials of Precious Metals from end-of-life Vehicle Parts by Selective Dismantling." *Environmental Science and Technology* 53 (2): 733–742. <https://pubs.acs.org/doi/abs/10.1021acs.est.8b04273>.
- Xu, S., B. Lu, M. Baldea, T. F. Edgar, W. Wojsznis, T. Blevins, and M. Nixon. 2015. "Data Cleaning in the Process Industries." *Reviews in Chemical Engineering* 31 (5): 453–490. <https://doi.org/10.1515/revce-2015-0022>.
- Zhang, C., H. Dong, Y. Geng, H. Liang, and X. Liu. 2022. "Machine Learning Based Prediction for China's Municipal Solid Waste under the Shared Socioeconomic Pathways." *Journal of Environmental Management* 312:114918. <https://doi.org/10.1016/j.jenvman.2022.114918>.
- Zheng, G., D. Ivanov, and A. Brintrup. 2025. "An Adaptive Federated Learning System for Information Sharing in Supply Chains." *International Journal of Production Research* 63 (11): 3938–3960. <https://doi.org/10.1080/00207543.2024.2432469>.
- Zhou, F., and P. Ma. 2020. "End-of-life Vehicle Amount Forecasting based on an Improved GM (1, 1) Model." *Engineering Letters* 28 (3): 777–782.



DSPACE

<https://dspace.org/>

Integrated machine learning model for managing customer-driven end-to-end supply chain uncertainty

Xia, Hanbing; Sani, Shehu; Milisavljevic-Syed, Jelena; Salonitis, Konstantinos

2026-12-31

Taylor & Francis

<https://doi.org/10.1080/00207543.2026.2648764>

Xia H, Sani S, Milisavljevic-Syed J, Salonitis K. (2026) Integrated machine learning model for managing customer-driven end-to-end supply chain uncertainty. International Journal of Production Research, Available online 26 March 2026