Journal of Digital Economy 4 (2025) 144–155

Contents lists available at ScienceDirect

Journal of Digital Economy

journal homepage: www.keaipublishing.com/en/journals/journal-of-digital-economy

A machine learning approach to inventory stockout prediction

Yang Liu [a,*], Dimitra Kalaitzi [b], Michael Wang [c], Christos Papanagnou [d]

[a] Henley Business School, University of Reading, United Kingdom
[b] School of Management and Economics, Cyprus University of Technology, Cyprus
[c] Department of Management, Kingston University, Global Business School for Health, University College London, United Kingdom
[d] College of Engineering and Physical Sciences, Aston University, United Kingdom

ARTICLE INFO

Keywords:
Stockout
Backorder
Machine learning
Data analytics
Data imbalance
Inventory management

ABSTRACT

The retail industry continues to experience frequent stockouts, driven by the rise of e-commerce and disruptive events such as the COVID-19 pandemic, which have significantly impacted both profitability and supply chain stability. As a result, developing effective models for stockout prediction has become increasingly critical for enhancing the efficiency and resilience of retail operations. The growing availability of data, challenges posed by data imbalance, and high demand uncertainty underscore the need to transition from traditional forecasting models to more intelligent, data-driven approaches that integrate multiple relevant features alongside sales data. In this study, we utilise a large dataset from a retailer comprising over 1.6 million stock keeping units (SKUs) to develop an analytical model based on classical machine learning algorithms aimed at improving stockout prediction accuracy. Our results demonstrate that the proposed approach performs well in handling large-scale, imbalanced data and significantly enhances predictive performance. Feature importance analysis reveals that current inventory levels, short-term demand forecasts (three months), and recent sales data are the most influential factors in predicting stockouts. Furthermore, the findings suggest that recent demand forecasts and sales data have greater predictive power than longer-term projections (six and nine months), highlighting the importance of near-term indicators in inventory stockout prediction accuracy. To the best of our knowledge, these insights provide valuable contributions to understanding stockout dynamics and improving inventory management strategies within the retail sector.

1. Introduction

Trade wars, Coronavirus disease 2019 (COVID-19) pandemic and Brexit highlighted the poor inventory management practices that are followed by the retail industry (e.g., fashion retailer H&M faced with excess inventory, and UK shoppers experienced out-of-stock products in supermarkets). Retailers found themselves grappling with the challenges posed by a surge in consumer demand for popular products, further complicating an already intricate landscape. According to Andaur et al. (2021), around four percent of the total annual revenue worldwide in the retail sector is lost due to out-of-stock which is close to $984 billion loss.

Most retail supply chains are structured based on anticipated behavioural trends. However, because of the COVID-19 pandemic, retailers experienced a sudden shift from traditional forms of purchasing (e.g., physical stores) to online retail-oriented consumption

* Corresponding author.
E-mail addresses: kurt.liu@henley.ac.uk (Y. Liu), dimitra.kalaitzi@cut.ac.cy (D. Kalaitzi), m.wang@kingston.ac.uk (M. Wang), c.papanagnou@aston.ac.uk (C. Papanagnou).

Peer review under responsibility of Editorial Board of Journal of Digital Economy.

https://doi.org/10.1016/j.jdec.2025.06.002
Received 22 May 2025; Received in revised form 12 June 2025; Accepted 15 June 2025

(e.g., e-stores), coupled with consumer stockpiling, resulting in starkly empty shelves for particular products. Studies highlighted that customers favour more precise time slots and increase speed for home deliveries of e-grocery products (Amorim et al., 2023). While this upheaval represented a disruption in demand pattern, the next significant change could just as easily manifest as a disruption of supply, giving rise to stockouts of essential products in the market. Such scarcity has the potential to escalate product prices, fuel conflicts, and incite political or civil unrest (Jones et al., 2023). The invasion of Ukraine in 2022 reinforced these disruptions, impacting the global supply of critical resources such as wheat, plastics and metal, resulting in diminished availability and heightened prices. As the enduring repercussions of such disruptions are likely to remain uncertain in the future, retailers should implement alternative strategies to address immediate needs and implement short-term decision-making tools to anticipate new retail landscapes.

The negative consequences of stockout occurrences are often associated with lost sales and dissatisfied customers. Stockouts can cause disruptions within the supply chain, resulting in delays in fulfilling orders and meeting customer demand. These disruptions have the potential to induce inefficiencies in production and distribution processes, consequently affecting the overall flow of goods and stocks. Retailers rely significantly on customers and their purchasing power. Stockouts can lead to low customer satisfaction levels, eroding trust and damaging relationships. Customers may seek alternative brands or competitors when faced with stockouts, affecting brand loyalty and overall customer satisfaction levels within the supply chain. Stockouts stress the critical significance of proficient inventory management and the need for precise forecasting, streamlined replenishment protocols, and enhanced visibility into inventory levels as pivotal strategies aimed at pre-empting disruptions within the supply chain (Amorim et al., 2021).

The shift towards digitalisation has intensified the complexity for retailers of managing inventory by expanding the range of products offered to cater to diverse customer preferences and behaviours (Fildes et al., 2022). Thus, retailers seek analytics methods to predict customer needs, optimise replenishment processes, prevent stockouts, speed up order fulfilment, increase customer satisfaction, and reduce costs (Ulrich et al., 2021; Loureiro et al., 2018). Digitalizing every aspect of life has also made data more readily available for analytics (Hübner et al., 2021a). Essentially, retailers can leverage nowadays a plethora of different data and factors (e.g., historical sales data, inventory levels, supplier performance, market trends, and seasonal patterns) to develop predictive models (Fathi et al., 2021; Hübner et al., 2021b). For example, Walmart has invested in an artificial intelligence (AI) and machine learning (ML) powered inventory management system to improve the stock level and satisfy the customer demand. In the same vein, the Otto Group used predictive ML applications and minimised the stockout rate by 80 % (Oosthuizen et al., 2021). These techniques are able to support the correlations between various causal elements having nonlinear relational demand patterns, thus improving the retail chain performance (Mitra et al., 2022).

The utilisation of data analytics and ML methods holds significant potential in forecasting stockouts within the retail industry. Accurate demand forecasts enable retailers to optimise inventory levels and minimize the likelihood of stockouts (Kourentzes et al., 2020). However, given the contemporary constraints, large amounts of data and the number of different scenarios that retailers need to consider (e.g. pricing, promotions, and competitor actions) (Taparia et al., 2023), there is a need for intelligent inventory systems to produce more efficient and flexible inventory classification models (Kartal et al., 2016; Stranieri et al., 2024a). Nowadays, with the aid of advanced technologies (e.g., Internet of Things) retailers have started to invest in real-time analytics enabling them to continuously monitor inventory levels, sales trends, and market dynamics. By detecting fluctuations and anomalies in live data streams, retailers can promptly adjust inventory replenishment strategies to minimize the risk of stock shortage.

However, there is a notable lack of comprehensive studies that specifically explore the application of data analytics and ML in inventory management within retail operations. Furthermore, more in-depth research is needed to examine the challenges and limitations associated with integrating these technologies into existing systems. To maximize the effectiveness of predictive models, retailers must also continuously refine and update them based on new data and real-time feedback. While the existing literature offers valuable insights into the use of ML algorithms in areas such as optimal route prediction (Neves-Moreira et al., 2022) and broader supply chain management, research on their application in inventory control—particularly in addressing stockout scenarios—remains limited (Theodorou et al., 2023).

Accurately predicting stockouts remains a significant challenge, particularly for large retailers managing high-dimensional datasets comprising millions of SKUs. These predictions inform critical merchandising decisions such as product placement, replenishment policies, and promotional strategies. However, several factors complicate this process. SKUs often have diverse and complex attributes, while inventory records can be prone to errors (Rekik et al., 2019). Moreover, sales datasets frequently exhibit an imbalanced distribution across different types of SKUs. Compounding this issue is the typically low average stockout rate across industries (Jenkins, 2022), which leads to an inherent imbalance between stockout and non-stockout instances in the data. This imbalance can significantly degrade the performance and accuracy of stockout prediction models (Shajalal et al., 2023; Andaur et al., 2021). Given these challenges, the analysis of large-scale SKU data in the retail sector demands more robust and advanced analytical tools than traditional forecasting methods—such as Croston's method or more recent approaches like TBATS (Dharmawardane et al., 2021; Hübner et al., 2013).

In this study, we address the aforementioned challenges by proposing an analytical model based on classical ML algorithms to enhance stockout prediction performance. This research contributes to the emerging body of literature focused on forecasting stockouts in the retail industry and distinguishes itself by utilizing a large-scale dataset containing over 1.6 million SKUs—an uncommon scale in existing studies. Our findings offer valuable insights into the dynamics of stockouts at scale, illuminating both previously debated and underexplored aspects of inventory management.

This study enriches the inventory management literature by demonstrating the efficacy of classical ML models, particularly Random Forest (RF), in accurately predicting stockouts across large-scale, imbalanced SKU datasets. The integration of near-term demand forecasts and current inventory levels as the most significant predictors reinforces existing findings on the primacy of recent, time-sensitive data for demand-driven decision-making (Goltsos et al., 2022; Ntakolia et al., 2021). Moreover, the application of SMOTE with Tomek Links to correct data imbalance addresses a persistent methodological issue in retail analytics and enhances model generalisability across diverse datasets (Chawla et al., 2002; Yan et al., 2019). These contributions extend the theoretical understanding of intelligent inventory systems by offering an interpretable and replicable framework for stockout prediction, paving the way for future research on reinforcement learning and deep learning applications in dynamic inventory environments (Stranieri et al., 2024b; Liu et al., 2024).

Regarding managerial implications, this study is highly relevant for retailers seeking to optimise inventory management and minimize stockouts. Implementing predictive models based on short-term forecasts and real-time inventory data enables firms to enhance forecasting accuracy, reduce lost sales, and improve customer satisfaction (Andaur et al., 2021; Shajalal et al., 2023). This is particularly critical in sectors prone to demand surges and supply shocks, such as grocery retail, where disruptions can undermine service continuity and consumer trust. By minimising such disruptions, firms can strengthen supply chain resilience and responsiveness.

Moreover, the model's adaptability and scalability make it suitable for critical sectors such as pharmaceutical supply chains, where preventing stockouts of essential medicines is vital. Retailers can integrate the ML-based analytical model proposed in this study into real-time inventory tracking and dynamic safety stock adjustments, ensuring precise replenishment decisions. Accurate stockout prediction in this context could support more efficient batch production planning, regulatory compliance, and cost-effective inventory management aiming at reducing disruptions and optimizing supply chain resilience (Galli et al., 2020; Hajek and Abedin, 2020).

The rest of the paper is organized as follows: Section 2 provides a state-of-the-art literature review on inventory management, including techniques used to address it with a particular focus on the retail industry. Section 3 introduces the methodology including our proposed analytical model, while Section 4 presents and discusses the results and the study contributions. Finally, conclusion and directions for further research are presented in Section 5.

2. Literature review

2.1. Inventory management

Historically, inventory management has centred on balancing replenishment lead times, carrying costs, and demand forecasting. Williams and Tokar (2008) provided a comprehensive review of inventory management research in major logistics journals, identifying two primary themes: the integration of inventory decisions with logistics functions and the emergence of collaborative inventory models. These collaborative approaches emphasize joint planning and information sharing among supply chain partners to enhance inventory efficiency. Munyaka and Yadavalli (2022) further explored inventory management concepts, highlighting its role as a cornerstone of supply chain and logistics systems. Their systematic review underscored the importance of aligning inventory strategies with organizational objectives and the need for adaptive approaches in response to dynamic market demands.

According to Deng and Liu (2021), there are two inventory main management approaches: traditional (i.e., methods that rely on static mathematical statistical analysis) and intelligent (i.e., methods deploying machine learning approach), which also can be linked to the three main approaches indicated by Stranieri et al. (2024a) that have been used in the literature to manage the inventory decisions namely: multi-stage stochastic programming, reordering policies and reinforcement learning. In the literature, most of the studies focus on inventory management methods that rely on traditional mathematical statistical analysis (e.g., McGillivray and Silver, 1978; Parlar, 1985). For instance, Bertsimas et al. (2016) developed a stochastic optimization model to obtain the optimal inventory quantity for a retail network.

Over the past decade, a variety of factors, including the abundance of products, limited warehouse spaces, evolving consumer preferences, and fierce market competition, have engendered heightened complexities for retailers. These intricacies have led to a range of operational inefficiencies, particularly instances of stockouts, which in turn manifest in meagre profit margins (Tran et al., 2024; Koren et al., 2022). In addressing these operational challenges, retailers have undertaken various strategies that embrace digital technology and smart algorithms that make formal and accurate decisions about replenishment policies and optimal inventory targets.

Although the progression of digital technology has revolutionised the operations of retailers and Fast-Moving Consumer Goods (FMCG) markets, there remains a pressing need to address the challenges posed by the accuracy of demand forecasts and rising product prices. Often, the factors impacting demand, especially promotional information, introduce significant complexity, potentially leading forecasters to encounter the dimensionality problem characterised by an abundance of variables and insufficient data (Fildes et al., 2022). In response, supply chain managers must harness the power of innovative solutions such as data analytics and machine learning tools (Liu, 2022).

With the advent of advanced technologies and data analytics, supply chain procurement and supply chain managers have increasingly turned to supply chain analytics (SCA) to improve inventory management processes. According to Kalaitzi and Tsolakis (2022, p. 2) SCA can be defined as “A set of capabilities and qualitative/quantitative techniques utilised to analyse traditional data and big data to inform decision-making in operations to achieve improved supply chain performance and competitive advantage”. SCA enables companies to forecast future demand more accurately by leveraging data such as historical sales data, and external factors (Seyedan and Mafakheri, 2020). In order to optimise inventory levels across supply chains, these analytics techniques take into consideration factors such as lead times and carrying costs.

These analytical techniques and tools can support overcoming challenges such as overstocking, stockouts, demand volatility, and supply chain disruptions by providing actionable insights into inventory optimization and risk mitigation strategies, thus optimizing the inventory ordering decisions (Wang et al., 2016a). For example, Sharma and Garg (2016) and Arya et al. (2017) discussed the relationship between SCA and inventory control and visibility. Several studies such as Giannoccaro and Pontrandolfo (2002) and Jiang and Sheng (2009) have utilised the reinforcement learning method. In the retail industry, Kara and Dogan (2018) focused on perishable products and used the reinforcement learning approach namely the genetic algorithm, Q-learning and Sarsa algorithms. Shi et al. (2020) also utilised machine learning techniques in the predictive analytics phase in a Chinese fashion retailer. Demizu et al. (2023) also attempted to optimise the control of inventory for new products, which is challenging as there is not enough data for learning by combining, by developing a model-based on deep reinforcement learning. Liu et al. (2024) developed the Maskable LSTMProximal policy optimization, a deep reinforcement learning algorithm that combines current observations and future predictions to tackle inventory management problems of omni-channel retailers.

In short, retailers can no longer rely on traditional inventory management methods to survive in an ever-changing environment but must follow a more digitalised approach and innovate by adopting new technologies (Oosthuizen et al., 2021). Big data analytics techniques are widely used in key areas of operational management such as inventory management. These novel methods are based on well-known machine learning models (e.g., deep learning and reinforcement learning methods). However, this research is still in its early stages and opportunities exist to advance our knowledge and the accuracy of demand prediction for inventory management (Deng and Liu, 2021).

2.2. Factors to predict future stockout events

Prior studies have explored several key features (or product attributes) that may lead to stockouts when there is insufficient inventory and stockouts. For example, Pritchard et al. (2023) suggested that high product sales rates increase the risk of stockouts because of the quick depletion of inventory. This can occur if inventory levels are not properly managed, causing the rapid sales to exhaust the available stock faster than expected, resulting in stockouts. Past sales data is a critical element in predicting future demand and preventing stockouts and to reduce the risk of stockouts caused by high sales velocity, retailers require sophisticated demand forecasting and inventory management systems. However, relying solely on historical sales data without considering other factors can result in inaccuracies in forecasting stockouts (Ovezmyradov, 2022).

Pan and Hsiao (2005) suggested that lead time is a known factor that impacts stockouts in the retail industry. Longer lead times result in a prolonged duration for retailers to replenish inventory once it is exhausted. If lead times are not accurately accounted for in forecasting models, there is a risk of misjudging future demand during the lead period. This can result in stockouts as retailers fail to replenish inventory in time to meet customer demand. In addition, underestimating lead times can result in insufficient inventory levels, leading to stockouts and lost sales opportunities. Wolters and Huchzermeier (2021) pointed out that lead times can vary seasonally, especially in the retail industry where demand is influenced by factors such as holidays, promotions, and changing consumer preferences. Failure to account for seasonal variations in lead times can lead to inaccurate forecasts and stockouts during peak demand periods or overstock situations during slower seasons (Li, 2020).

Previous research has also suggested that the choice of ordering policy (e.g., reorder point or reorder quantity) and inventory policies (e.g., continuous replenishment) can influence the impact of stockout forecasting (Roldán et al., 2016). For example, if the reorder point is set too low, there is a higher likelihood of stockouts occurring before new inventory arrives. Similarly, if the reorder quantity is insufficient to cover demand, stockouts may occur even if the reorder point is reached. Typically, inventory policies use three key parameters: when to replenish, how much to replenish, and how often the inventory level is reviewed, decisions that may lead to stockout occurrences (Roldán et al., 2016).

Table 1 summarizes the various input features or attributes used by ML models to predict whether a stockout will occur for a given SKU. These features contribute to accurate stockout prediction, enabling retailers to implement proactive inventory strategies such as adaptive replenishment and demand-driven stock optimization. As can be seen, they range from sales features, demands, lead times to inventory features, to name a few. Worth mentioning, Papakiriakopoulos et al. (2009) is one of the first studies that adopted machine learning techniques in the grocery retail sector to predict products that are not on the shelf based on sales, inventory and other data. Ntakolia et al. (2021) employed machine learning models such as RF, XGB, LGBM, and BB to predict the backorder rate in an inventory

Table 1
Features to predict stockout management from recent literature.

Author(s)	Key Predictive Features for Stockout Management
Pritchard et al. (2023); Ovezmyradov (2022)	High sales velocity/turnover, sophisticated demand forecasting, dynamic replenishment, real-time inventory -monitoring system.
Pan and Hsiao (2005); Wolters and Huchzermeier (2021); Li (2020)	Lead times, seasonal demand variations.
Papakiriakopoulos et al. (2009)	Sales velocity, inventory level, promotional product, shelf space, stock centralization, market share, seasonality, store size, employees, store managers decisions
Andaur et al. (2021)	Sales features (e.g., the number of days that a product has not sold any single unit in a period/total number of days in a period (percent), average number of items sold in a period); Inventory features (e.g., average inventory level using only the days that a product has positive stock in a period); Ordering features (e.g., number of items received on a period to the D.C. for a specific store).
Ntakolia et al. (2021)	Low stock, quantity in-transit, high short-term and mid-term future demand
Shajalal et al. (2023)	Supplier performance, lead times, inventory tracking, product sales.
Siddiqui et al. (2023)	Forecast error and demand variability
Stranieri et al. (2024b)	Number of product types, number of local warehouses, length of lead Times

system and provided interpretations of the major contributing factors that lead to product backorder. Most specifically, they found that precise predictions on future demands, product quality and transit time, quantity in transit are key elements to handle and avoid backorders and optimise the inventory system. Stranieri et al. (2024b) also utilised a similar set of parameters when evaluating DRL algorithms and static inventory policies. Andaur et al. (2021) to predict out of stock events utilised data from retail packaged foods manufacturing company in Latin America and developed two machine learning based systems and found that the RF classifier presented the best performance in a real-world setting. In this work, the issue of data imbalance was highlighted, and suggested and applied data balancing strategies namely oversampling or undersampling. For instance, Shajalal et al. (2023) utilised deep neural network and also accomplish to balance the dataset by applying different techniques such as randomised oversampling and SMOTE oversampling.

2.3. Intrinsic challenges from SKUs data

In addition to the aforementioned factors, managing and analysing data for a multitude of SKUs can become very challenging, particularly if the data is scattered across various systems or sources (Hirche et al., 2021). This complexity can hinder the ability to detect patterns, trends, and anomalies that may suggest potential stockout risks, making it challenging to ensure adequate inventory levels and avoid stockouts. Having access to extensive data on sales, inventory, customer behaviour, market trends, and supply chain operations can be daunting to manage and analyse efficiently for companies. This situation can compromise forecasting, especially when inventory managers face challenges in deriving actionable insights promptly from the data, thereby impeding their capacity to forecast and avert stockouts.

The veracity of SKU datasets constitutes an important factor in inventory management. Retailers often face the dilemma of analysing certain SKUs that are overrepresented in the dataset or when they are underrepresented or omitted entirely. Kalla et al. (2020) argued that data imbalance may cause forecasting models to inaccurately estimate demand for the under/overrepresented SKUs. This can lead to stockouts for those SKUs as inventory levels are not adequately replenished to meet actual demand. Additionally in an unbalanced dataset, resources such as inventory, marketing efforts, and shelf space may be disproportionately allocated to high-demand SKUs, leaving low-demand SKUs with insufficient resources. As a result, low-demand SKUs may experience stockouts due to inadequate inventory levels or visibility (Li et al., 2022). One of the primary reasons for stockout prediction failure is inaccurate demand forecasting, which may be caused because companies may lack the tools, data, or expertise to forecast demand effectively, leading to underestimation or overestimation of future demand for products. By accurately predicting demand and maintaining appropriate inventory levels, retailers can ensure they have sufficient stock to meet customer demand without experiencing stockouts (Abolghasemi et al., 2020).

The analysis of the sheer volume of SKUs combined with imbalanced datasets that this article demonstrates poses retailers with a contemporary challenge at present and may hinder effective inventory management practices, such as safety stock allocation and reorder point optimization. Without accurate demand forecasts for all SKUs, retailers may struggle to maintain optimal inventory levels and mitigate the risk of stockouts effectively. In our study, we attempt to address these challenges by analysing a large imbalanced SKUs dataset and proposing an analytical model to more accurately predict stockout occurrences, thereby offering useful insights into effective inventory management, especially for stockout management.

2.4. Research gap

Several data-driven models introduced that integrate deep reinforcement learning to optimise dynamic pricing and inventory strategies for omni-channel retailing, under conditions of demand uncertainty and variable customer behaviour or combined deep reinforcement learning with multi-stage stochastic programming, however, most of these studies focused on pricing optimization and assumes balanced datasets (Hajek and Abedin, 2024; Liu et al., 2024; Stranieri et al., 2024a), thereby overlooking the challenges associated with imbalanced and high-dimensional stock-keeping unit (SKU) data, especially for stockout prediction. Collectively, these works illustrate a growing scholarly interest in intelligent and adaptive inventory control frameworks, yet they also reveal a pressing need for scalable and interpretable models that sustain performance in highly imbalanced data environments. These insights have directly driven our decision to employ classical ML models (e.g. a RF model) enhanced with SMOTE and Tomek Links, ensuring not only robust performance on imbalanced data but also improved interpretability for real-world inventory challenges. Thus, while existing research often emphasizes aggregate demand forecasting or assumes balanced data, this study highlights the need for scalable, interpretable models capable of operating under real-world constraints. In response, the paper contributes by developing and validating a machine learning-based analytical model for accurate stockout prediction across a large-scale, imbalanced retail dataset. By addressing these gaps, the study underscores the tangible value of integrating machine learning into inventory management and delivers actionable insights in the dynamic retail environment.

Note that this research is grounded in a hybrid theoretical framework that integrates classical inventory management theory with modern statistical learning principles. Drawing from established operations research models (such as EOQ and dynamic inventory control theories) and data-driven decision-making frameworks, our study leverages advances in ML to enhance predictive accuracy. This integrative approach not only bridges the gap between traditional forecasting methods and contemporary big data analytics but also provides a robust foundation for developing scalable and interpretable stockout prediction models. By aligning empirical insights from prior studies (e.g., Ntakolia et al., 2021; Shajalal et al., 2023) with these theoretical constructs, our work addresses both the practical challenges of inventory management and the methodological limitations observed in existing literature.

3. Methodology

3.1. Data preparation and preprocessing

The data we used in our stockout prediction are from an anonymous retailer with over 1.6 million SKUs. The company records several key features for each SKU, including the current inventory level, the lead time for stock replenishment, in-transit quantity, the historical sales (past one, three, six, and nine months), future demand forecasts (next three, six, and nine months), recommended minimum safety stock level, overdue amounts from suppliers, average historical supplier performance (for past six and 12 months), risks factors, and stockout status. We performed data cleaning and transformation to prepare the data for model training. First, incomplete data entries were removed, errors were corrected, and missing values were replaced where possible using mean values. Second, the categorical data were transformed using dummy coding. Third, we used centring and scaling to standardize the dataset to handle skewness and outliers issues.

Our target variable is the stockout status, which was denoted by "Yes" or "No" in the dataset. We transferred the target variable using "0" to indicate non-stockout, and "1" to represent stockout. In practice, the chances of experiencing stockout should only represent a small percentage in any types of successful business operations. It is the same for our case firm. In this dataset, 99.33 % SKUs are non-stockout, whereas only 0.67 % are stockout. For our model training and prediction purposes, it is important to maintain the same class imbalance in both train and test sets. Thus, we applied a stratified shuffle split technique to create splits by preserving the same percentage for each target class as in the complete set. The stratified shuffle split method can reduce sampling bias where a random slit may fail (Ma et al., 2018; Szeghalmy and Fazekas, 2023).

3.2. Handling dataset imbalance

The imbalance in the train set can cause biases in data classification, leading to poor accuracy and generalization performance (Sain and Purnami, 2015; Zheng et al., 2015; Zou et al., 2016). For binary classification problems, the performance of any classifier is usually evaluated by *accuracy*, *precision* and *recall* (Liu, 2022; Zhang et al., 2022). These standard metrics are calculated by using the confusion matrix (Shajalal et al., 2023) as illustrated in Table 2.

- Precision $= \frac{TP}{TP+FP}$ (1)
- Recall $= \frac{TP}{TP+FN}$ (2)
- Accuracy $= \frac{TP+TN}{TP+FN+FP+FN}$ (3)

However, for imbalanced classification, *accuracy* is not adequate to reflect the classifier performance (Evgeniou et al., 2022; Shajalal et al., 2023). To examine the issue, we first trained a logistic regression model (LR) with the train set and make predictions, which returned a 99.33 % accuracy on the test set. Then, we used confusion matrix to evaluate the performance of the classifier as shown in Fig. 1. It can be seen that even though the classifier has a high accuracy rate, the correctly predicted number of TPs (i.e., true stockout) to be zero, indicating a rather poor performance of the model in handling the imbalanced dataset.

To deal with the imbalance in our dataset, we thus applied a combined over- and under-sampling technique SMOTE + Tomek Links (Batista et al., 2004; Sain and Purnami, 2015). The synthetic minority oversampling technique (SMOTE) has long been proven to be very effective in handling class-imbalanced problem (Chawla et al., 2002), but it can also create overlapping problems between classes (Wang and Japkowicz, 2004). Tomek Links can be used as an under-sampling method or as a data cleaning method (Batista et al., 2004), which can remove the overlapping samples that are introduced by SMOTE (Yan et al., 2019). The resampled train set with balanced stockout vs. non-stockout target variables is illustrated in Fig. 2.

3.3. The stockout prediction analytical model

We adopted five classical supervised learning models: LR, SVM, RF, AdaBoost, and GB. These models are representative classifiers that have been used in various machine learning tasks (Zhang et al., 2022). In addition, we used eXtreme GB (XGBoost) that augments GB by enhancing regularisation terms to prevent overfitting (Chen and Guestrin, 2016). We first trained these models using our balanced train set and then used our test set to make predictions as shown in Fig. 3. Grid search was also applied to find the best combination of hyperparameters for a given model. The prediction performances of the trained models were then evaluated and results from the best model were presented.

Table 2
Confusion matrix.

	PREDICTED Negative	*PREDICTED Positive*
ACTUAL Negative	TRUE Negative (TN)	FALSE Positive (FP)
ACTUAL Positive	FALSE Negative (FN)	TRUE Positive (TP)

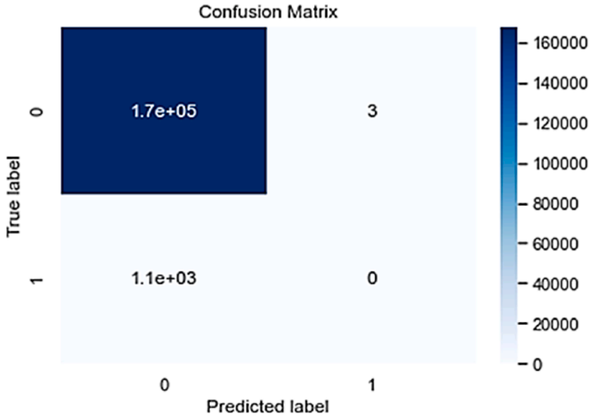

Fig. 1. Confusion matrix for the LR model.

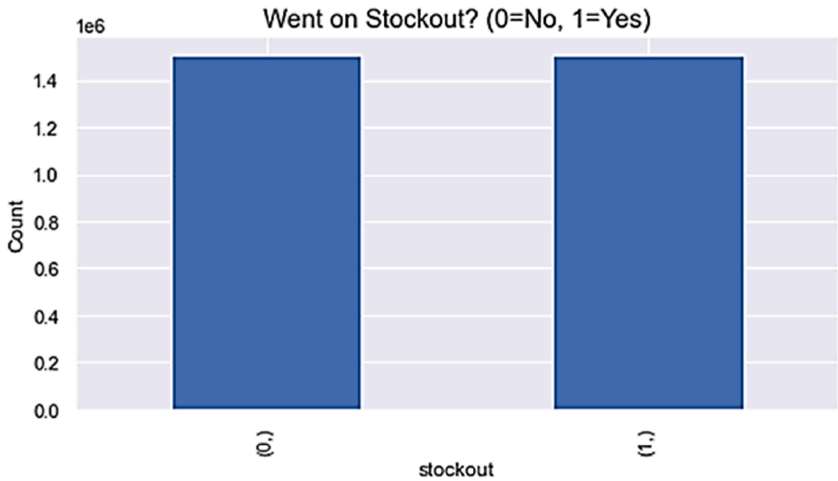

Fig. 2. Balanced train set after transformation.

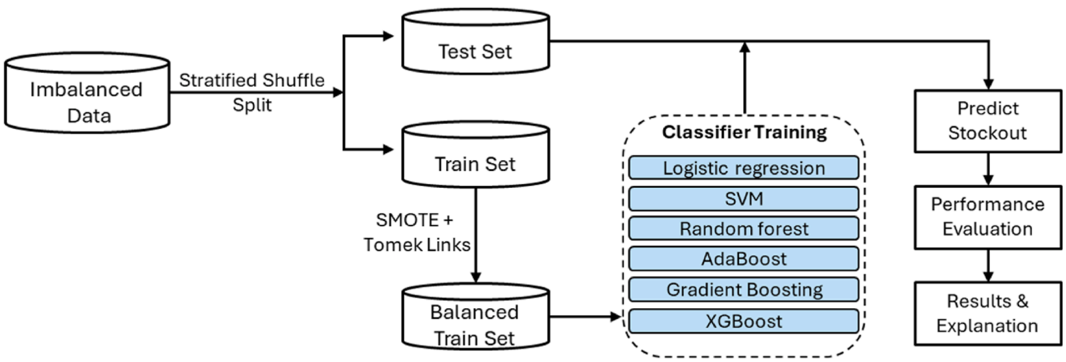

Fig. 3. The methodological stockout prediction Flowchart.

3.4. Results and models performance evaluation

To evaluate the prediction performance of our trained models, we calculated the area under the curve (AUC) of the receiver operating characteristic (ROC) curve (Fawcett, 2006). AUC measures the discriminative power of the classifier and how well the false positive and true positive rates perform in comparison to the random choice model, which cannot be measured by the standard *accuracy* (Evgeniou et al., 2022). This is especially important when there is imbalance in the dataset (Zhang et al., 2022). An AUC score equal to or close to 1 indicates a perfect classifier (Liu, 2022). Table 3 summarizes the performance scores.

It can be seen that the RF model outperforms all the other five models, with an AUC score of 0.978 indicating a well performed classifier.

3.5. Feature importance

RF are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001). By computing the mean and standard deviation of accumulation of the impurity decrease within each tree, we are able to derive the feature importances. Based on our RF model, we plotted the feature importance with the features included in the dataset. It can be observed that the current inventory level, 3-month forecast, 6-month forecast and 9-month forecast are highly important in predicting stockout as well as the past 1-month sales (see Fig. 4).

4. Discussion

As noted in previous studies, inventory stockout or supply shortage can pose significant challenges to business success of firms (Chang et al., 2019), especially for the retail sector (Jing and Lewis, 2011). It can produce adverse effects on the company reputation and consequently lead to loss of customers and sales because of customer dissatisfaction. Therefore, the high accuracy of stockout prediction offers a greater possibility of developing more efficient inventory management systems to effectively tackle back-orders problems faced by many retailers (Shajalal et al., 2023). However, inventory stockout usually takes place due to many varying factors, making accurate predictions difficult to achieve, especially when there is a large number of SKUs and imbalance in the dataset.

In our study, to deal with the imbalanced dataset, we first applied the stratified shuffle split technique to create splits by preserving the same percentage for each target class as in the complete set for our model training and testing purposes. Then, we used SMOTE + Tomek Links sampling technique to balance our train set. These techniques turned out to be effective in handling the imbalance in our dataset (Batista et al., 2004; Sain and Purnami, 2015). We also trained various classical machine learning algorithms and evaluated their performances. Even though we did not apply more complicated algorithms in our study such as deep neural network (Shajalal et al., 2023), the overall prediction performance is dramatically improved. Most importantly, the results suggest that the proposed analytical model for stockout prediction as depicted in Fig. 3 can help improve stockout prediction accuracy. In real-world applications, it is vital for inventory managers to fully understand their inventory data and the average stockout rates as per the number of SKUs. If there is an obvious imbalance in the dataset, the managers could follow the proposed stockout prediction process in this study to first deal with the imbalance before feeding the data into any forecasting models in order to improve the prediction performance and accuracy.

Fig. 4 reveals that the current inventory level and demand forecasts for the next three, six and nine months play more important roles in stockout prediction, followed by past one-month and three-month sales. These results partially support the previous findings, which confirm the relevance of specific predictors and the importance of addressing data imbalance in inventory forecasting. Consistent with Ntakolia et al. (2021), we found that low inventory levels and short-term demand forecasts are critical indicators of potential stockouts. Their study emphasized features such as low stock, quantity in transit, and short-to mid-term demand forecasts as the most influential variables in predicting backorders, aligning closely with our feature importance results. Similarly, Shajalal et al. (2023) highlighted the value of real-time, recent data—particularly sales and supplier-related information—for improving prediction accuracy, a conclusion echoed in our finding that near-term sales and forecasts outperformed longer-term indicators. Furthermore, both studies addressed the challenge of imbalanced datasets using SMOTE-based techniques; however, our study contributes further by applying a combined SMOTE and Tomek Links approach to a uniquely large and highly imbalanced retail dataset comprising over 1.6 million SKUs; thus, our research not only aligns with but also extends the current literature by offering scalable, interpretable solutions for stockout prediction in complex, real-world inventory environments.

To interpret the results from a practical perspective, if the current inventory level of a particular SKU is low, there will be a high risk of having stockout for that item, which is quite straightforward to understand. The findings regarding demand forecasts are consistent with prior research that accurate forecasting is important for effective inventory control (Goltsos et al., 2022; Rego and Mesquita, 2015; Tiacci and Saetta, 2009). Poor demand forecasting can often lead to inaccurate replenishment decisions, resulting in stockout. Therefore, inventory managers should strive to make accurate demand forecasts and accordingly make effective inventory control policies to adjust inventory levels and replenishment decisions.

Our findings further indicate that while demand forecasting is a key factor in stockout prediction, short-term forecasts—particularly those for the next three months—are more influential than longer-term forecasts (six- and nine-month horizons). This suggests that recent forecasts more accurately capture current demand trends, thereby exerting a more immediate and significant impact on inventory levels and the likelihood of stockouts. Similarly, recent sales data have a stronger effect on stockout prediction

Table 3
Model performance metrics.

	LR	SVM	RF	AdaBoost	GB	XGBoost
Precision	0.022	0.015	0.428	0.145	0.148	0.181
AUC	0.759	0.688	0.978	0.938	0.944	0.950

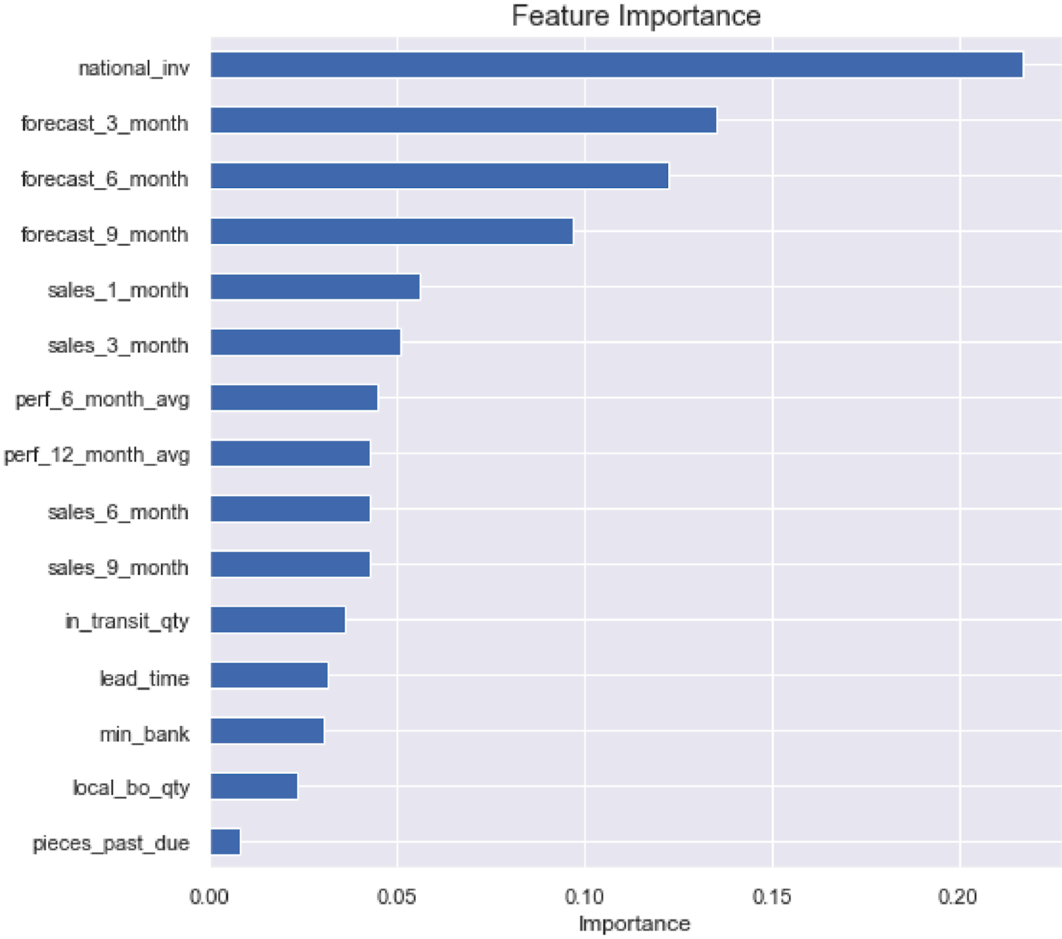

Fig. 4. RF feature importance.

than older sales figures, although the overall influence of sales is less pronounced compared to demand forecasts and current inventory levels. This reinforces the idea that recent sales are more reflective of current demand conditions and, therefore, more relevant for accurate forecasting.

These insights highlight the underlying relationship between time-series demand forecasting and historical sales, supporting previous findings that emphasize the importance of recent data in predictive accuracy (Huber et al., 2017). By analysing extensive historical datasets, we find that recent sales data play a critical role in improving the accuracy of demand forecasts. This underscores the value of integrating high-volume, real-time sales data into predictive models to enhance their effectiveness in anticipating stockouts.

In practice, inventory managers should closely monitor the current inventory levels of their SKUs and incorporate the most recent demand forecasts and sales data into their inventory control and replenishment strategies to minimize the risk of stockouts. Specifically, we recommend the following actions: i) **Continuous monitoring of current inventory levels** to enable timely replenishment decisions and prevent stockouts; ii) **Accurate forecasting of future SKU demand**, particularly in the near term, to allow for proactive adjustments in safety stock levels and reorder quantities; iii) **Ongoing analysis of recent sales figures** to better anticipate short-term demand fluctuations and adjust inventory policies accordingly.

Moreover, adopting advanced inventory management technologies that support real-time tracking and dynamic adjustment of replenishment timing and quantities based on up-to-date sales and demand forecasts can significantly enhance operational efficiency. Integrating an early-warning stockout prediction system into the inventory management process can further reduce the risk and cost of stockouts, while contributing to overall inventory optimization (Shajalal et al., 2023).

Such a system should be capable of pre-emptively identifying potential stockouts, thereby empowering retailers to take corrective action before disruptions occur. Future research should explore the design, implementation, and optimization of these predictive systems to improve inventory control strategies and strengthen supply chain resilience.

Last but not least, our research highlights the significant potential of integrating advanced ML techniques to improve the accuracy of stockout predictions. By leveraging ML algorithms, retailers can uncover complex patterns and detect anomalies within sales data, thereby enhancing the precision of their forecasting models. The incorporation of these sophisticated methodologies into predictive analytics frameworks offers substantial promise for strengthening the resilience of retail operations—particularly when dealing with imbalanced SKU datasets. This, in turn, can lead to more efficient inventory management and improved customer satisfaction.

However, the successful implementation of such technologies requires organizations to address a range of technological, organizational, and environmental challenges. It also necessitates fostering a culture of data-driven decision-making throughout the entire supply chain ecosystem (Kalaitzi and Tsolakis, 2022). Companies that cultivate strong data cultures are up to 23 times more likely to outperform competitors in customer acquisition and 19 times more likely to be profitable. Data-driven organizations are better equipped to respond to disruption, optimise processes, and scale predictive technologies effectively (Chatterjee et al., 2024; Deloitte, 2022; Ghafoori et al., 2024). In addition, a robust culture of data-driven decision-making is integral to successfully translating our predictive model into operational excellence especially during emergencies (Papanagnou et al., 2022). In the context of retail inventory management, this cultural foundation enables cross-functional teams to align around shared metrics, implement advanced analytics tools with confidence, and continuously improve forecasting and replenishment strategies based on real-time feedback. Developing such a culture involves training staff, establishing clear data governance policies, and incentivizing evidence-based decision-making across the supply chain ecosystem. In practice, this means that beyond deploying technical solutions, companies must invest in training, establish clear data governance policies, and foster an environment that values evidence-based approaches to decision-making.

5. Conclusion

During recent years, retail industry has dynamically shifted to omnichannel retailing due to various factors (e.g. the Covid-19 pandemic), but are confronting unprecedented challenges and disruptions, which makes inventory management especially dealing with stockout a significant and critical topic to address confronting many supply chain managers. In this study, we attempted to address the stockout prediction problems by utilizing a large dataset containing over 1.6 million SKUs and developed an effective analytic model consisting of classical machine learning algorithms. The results indicated that our proposed analytic model can improve stockout prediction performance when dealing with large and imbalanced dataset. The results also showed that RF model outperforms all the other ML algorithms, which enabled us to examine the important features for stockout management. A limitation of this study is that we only used the classical machine learning algorithms for the stockout prediction problem, whereas more sophisticated ML algorithms such as deep learning models might offer improved prediction performance. A sensitivity analysis can also be performed for each model to offer useful insights into different ML models' behaviour, help with feature selection and improve our understanding of the relationships between input variables and output. Future research can attempt to address this limitation by developing and adopting more advanced models. Last but not least, in the era of digital economy, data privacy and ethical consideration should be carefully considered for any future research that include the use of operational data.

CRediT authorship contribution statement

Yang Liu: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Dimitra Kalaitzi:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Michael Wang:** Writing – original draft, Formal analysis, Conceptualization. **Christos Papanagnou:** Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The author Yang Liu is an Associate Editor for ***Journal of Digital Economy*** and was not involved in the editorial review or the decision to publish this article.

References

Abolghasemi, M., Hurley, J., Eshragh, A., Fahimnia, B., 2020. Demand forecasting in the presence of systematic events: cases in capturing sales promotions. Int. J. Prod. Econ. 230, 107892.

Amorim, P., Eng-Larsson, F., Pinto, C., 2021. The cost of a broken promise: understanding and mitigating the impact of failure in online retail. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3855425. (Accessed 12 April 2025).

Amorim, P., DeHoratius, N., Eng-Larsson, F., Martins, S., 2023. Customer preferences for delivery service attributes in attended home delivery. Chicago Booth Research Paper No. 20-07. Available at: https://ssrn.com/abstract=3592597. (Accessed 12 April 2025).

Andaur, J.M.R., Ruz, G.A., Goycoolea, M., 2021. Predicting out-of-stock using machine learning: an application in a retail packaged foods manufacturing company. Electronics 10, 2787.

Arya, V., Sharma, P., Singh, A., De Silva, P.T.M., 2017. An exploratory study on supply chain analytics applied to spare parts supply chain. Benchmark Int. J. 24 (6), 1571–1580.

Batista, G.E.A.P.A., Prati, R.C., Monard, M.C., 2004. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorat. Newsletter 6, 20–29.

Bertsimas, D., Kallus, N., Hussain, A., 2016. Inventory management in the era of big data. Prod. Oper. Manag. 25 (12), 2006–2009.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Chang, J.A., Lu, H., Shi, J., 2019. Stockout risk of production-inventory systems with compound Poisson demands. Omega 83, 181–198.

Chatterjee, S., Chaudhuri, R., Vrontis, D., 2024. Does data-driven culture impact innovation and performance of a firm? An empirical examination. Ann. Oper. Res. 333 (2), 601–626.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
Deloitte, 2022. The state of AI and data-driven decision making in supply chains. https://www2.deloitte.com.
Demizu, T., Fukazawa, Y., Morita, H., 2023. Inventory management of new products in retailers using model-based deep reinforcement learning. Expert Syst. Appl. 229 (Part A), 120256.
Deng, C., Liu, Y., 2021. A deep learning-based inventory management and demand prediction optimization method for anomaly detection. Wireless Commun. Mobile Comput. 2021, 14. https://doi.org/10.1155/2021/9969357, 9969357.
Dharmawardane, C., Sillanpää, V., Holmström, J., 2021. High-frequency forecasting for grocery point-of-sales: intervention in practice and theoretical implications for operational design. Oper. Manag. Res. 14 (1–2), 38–60.
Evgeniou, T., Fekom, M., Ovchinnikov, A., Porcher, R., Pouchol, C., Vayatis, N., 2022. Pandemic lockdown, isolation, and exit policies based on machine learning predictions. Prod. Oper. Manag. 32, 1307–1322.
Fathi, M., Khakifirooz, M., Diabat, A., Chen, H., 2021. An integrated queuing-stochastic optimization hybrid Genetic Algorithm for a location-inventory supply chain network. Int. J. Prod. Econ.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874.
Fildes, R., Ma, S., Kolassa, S., 2022. Retail forecasting: research and practice. Int. J. Forecast. 38 (4), 1283–1318.
Galli, L., Levato, T., Schoen, F., Tigli, L., 2020. Prescriptive analytics for inventory management in health care. J. Oper. Res. Soc. 1–14.
Ghafoori, A., Gupta, M., Merhi, M.I., Gupta, S., Shore, A.P., 2024. Toward the role of organizational culture in data-driven digital transformation. Int. J. Prod. Econ. 271, 109205.
Giannoccaro, I., Pontrandolfo, P., 2002. Inventory management in supply chains: a reinforcement learning approach. Int. J. Prod. Econ. 78 (2), 153–161.
Goltsos, T.E., Syntetos, A.A., Glock, C.H., Ioannou, G., 2022. Inventory – forecasting: mind the gap. Eur. J. Oper. Res. 299, 397–419.
Hajek, P., Abedin, M.Z., 2020. A profit function-maximizing inventory backorder prediction system using big data analytics. IEEE Access 8, 58982–58994.
Hirche, M., Farris, P., Greenacre, L., Quan, Y., Wei, S., 2021. Predicting under- and overperforming SKUs within the distribution–market share relationship. J. Retailing.
Huber, J., Gossmann, A., Stuckenschmidt, H., 2017. Cluster-based hierarchical demand forecasting for perishable goods. Expert Syst. Appl. 76, 140–151.
Hübner, A.H., Kuhn, H., Sternbeck, M.G., 2013. Demand and supply chain planning in grocery retail: an operations planning framework. Int. J. Retail Distrib. Manag. 41 (7), 512–530.
Hübner, A., Amorim, P., Fransoo, J., Honhon, D., Kuhn, H., de Albeniz, V.M., 2021a. Digitalization and omnichannel retailing: innovative OR approaches for retail operations. Eur. J. Oper. Res. 294 (3), 817–819.
Hübner, A., Amorim, P., Fransoo, J.C., Honhon, D., Kuhn, H., Martinez-de-Albeniz, V., Robb, D., 2021b. Digitalization and omnichannel retailing: innovative or approaches for retail operations. Eur. J. Oper. Res. 294 (3), 817–819.
Jenkins, A., 2022. 'Stockouts defined', oracle NetSuite. https://www.netsuite.com/portal/resource/articles/inventory-management/stockout.shtml. (Accessed 12 April 2025).
Jiang, C., Sheng, Z., 2009. Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. Expert Syst. Appl. 36 (3, Part 2), 6520–6526.
Jing, X., Lewis, M., 2011. Stockouts in online retailing. J. Mark. Res. 48, 342–354.
Jones, A., Bridle, S., Denby, K., Bhunnoo, R., Morton, D., Stanbrough, L., Coupe, B., Pilley, V., Benton, T., Falloon, P., Matthews, T.K., 2023. Scoping potential routes to UK civil unrest via the food system: results of a structured expert elicitation. Sustainability 15 (20), 14783.
Kalaitzi, D., Tsolakis, N., 2022. Supply chain analytics adoption: determinants and impacts on organisational performance and competitive advantage. Int. J. Prod. Econ. 248 (C).
Kalla, R., Murikinjeri, S., Abbaiah, R., 2020. An improved demand forecasting with limited historical sales data. In: 2020 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–5.
Kara, A., Dogan, I., 2018. Reinforcement learning approaches for specifying ordering policies of perishable inventory systems. Expert Syst. Appl. 91, 150–158.
Kartal, H., Oztekin, A., Gunasekaran, A., Cebi, F., 2016. An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. Comput. Ind. Eng. 101, 599–613.
Koren, M., Perlman, Y., Shnaiderman, M., 2022. Inventory management model for stockout based substitutable products. IFAC-PapersOnLine 55 (10), 613–618.
Kourentzes, N., Trapero, J., Barrow, D., 2020. Optimising forecasting models for inventory planning. Int. J. Prod. Econ. 225.
Li, X., 2020. Valuing lead-time and its variance in batch-ordering inventory policies. Int. J. Prod. Econ. 228, 107731.
Li, S., Lu, L., Lu, S., Huang, S., 2022. Estimating the stockout-based demand spillover effect in a fashion retail setting. Manuf. Serv. Oper. Manag. 25, 468–488.
Liu, K.Y., 2022. Supply Chain Analytics: Concepts, Techniques and Applications, first ed. Palgrave Macmillan.
Liu, S., Wang, J., Wang, R., Zhang, Y., Song, Y., Xing, L., 2024. Data-driven dynamic pricing and inventory management of an omni-channel retailer in an uncertain demand environment. Expert Syst. Appl. 244.
Loureiro, A., Miguéis, V., Silva, L., 2018. Exploring the use of deep neural networks for sales forecasting in fashion retail. Decis. Support Syst. 114, 81–93.
Ma, J., Ovalle, A., Woodbridge, D.M., 2018. Medhere: a smartwatch-based medication adherence monitoring system using machine learning and distributed computing. In: The 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Honolulu, HI, pp. 4945–4948.
McGillivray, A.R., Silver, E.A., 1978. Some concepts for inventory control under substitutable demand. INFOR 16 (1), 47–63.
Mitra, A., Jain, A., Kishore, A., Kumar, P.A., 2022. A comparative study of demand forecasting models for a multi-channel retail company: a novel hybrid machine learning approach. Operat. Res. Forum 3 (4), 58.
Munyaka, J.B., Yadavalli, V.S.S., 2022. Inventory management concepts and implementations: a systematic review. S. Afr. J. Ind. Eng. 33 (2), 15–36.
Neves-Moreira, F., Almada-Lobo, B., Guimarães, L., Amorim, P., 2022. The multi-product inventory-routing problem with pickups and deliveries: mitigating fluctuating demand via rolling horizon heuristics. Transport. Res. E Logist. Transport. Rev. 164, 102791.
Ntakolia, C., Kokkotis, C., Karlsson, P., Moustakidis, S., 2021. An explainable machine learning model for material backorder prediction in inventory management. Sensors 21, 7926.
Oosthuizen, K., Botha, E., Robertson, J., Montecchi, M., 2021. Artificial intelligence in retail: the AI-enabled value chain. Australas. Market J. 29 (3), 264–273.
Ovezmyradov, B., 2022. Product availability and stockpiling in times of pandemic: causes of supply chain disruptions and preventive measures in retailing. Ann. Oper. Res. 1–33.
Pan, J., Hsiao, Y., 2005. Integrated inventory models with controllable lead time and backorder discount considerations. Int. J. Prod. Econ. 93, 387–397.
Papakiriakopoulos, D., Pramatari, K., Doukidis, G., 2009. A decision support system for detecting products missing from the shelf based on heuristic rules. Decis. Support Syst. 46 (3), 685–694.
Papanagnou, C., Seiler, A., Spanaki, K., Papadopoulos, T., Bourlakis, M., 2022. Data-driven digital transformation for emergency situations: the case of the UK retail sector. Int. J. Prod. Econ. 250, 108628.
Parlar, M., 1985. Optimal ordering policies for a perishable and substitutable product: a Markov decision model. INFOR 23, 182–195.
Pritchard, A., Sweeney, K., Celebi, H., Evers, P., 2023. The impact of stockout-based switching on fill rates. J. Bus. Logist. 44 (4), 741–763.
Rego, J.R., Mesquita, M.A., 2015. Demand forecasting and inventory control: a simulation study on automotive spare parts. Int. J. Prod. Econ. 161, 1–16.
Rekik, Y., Syntetos, A., Glock, C., 2019. Measuring the Sales Impact of Improving Inventory Records: How Improving the Accuracy of Inventory Records Can Grow Sales by 4-8%. Efficient Consumer Response (ECR). Available at: https://www.ecrloss.com/research/grow-sales-by-imp. (Accessed 12 April 2025).
Roldán, R.F., Basagoiti, R., Coelho, L., 2016. Robustness of inventory replenishment and customer selection policies for the dynamic and stochastic inventory-routing problem. Comput. Oper. Res. 74, 14–20.
Sain, H., Purnami, S.W., 2015. Combine sampling support vector machine for imbalanced data classification. Procedia Comput. Sci. 72, 59–66.

Seyedan, M., Mafakheri, F., 2020. Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. J. Big Data 7, 53.
Shajalal, M., Hajek, P., Abedin, M.Z., 2023. Product backorder prediction using deep neural network on imbalanced data. Int. J. Prod. Res. 61 (1), 302–319.
Sharma, M., Garg, N., 2016. Inventory control and big data. In: Mittal, M., Shah, N.H. (Eds.), Optimal Inventory Control and Management Techniques. IGI Global.
Siddiqui, K.I., Lee, M.M.Y., Koch, T., Dugundji, E., 2023. Case fill rate prediction. In: Terzi, S., Madani, K., Gusikhin, O., Panetto, H. (Eds.), Innovative Intelligent Industrial Production and Logistics. IN4PL 2023, Communications in Computer and Information Science, 1886. Springer, Cham.
Stranieri, S., Fadda, E., Stella, F., 2024a. Combining deep reinforcement learning and multi-stage stochastic programming to address the supply chain inventory management problem. Int. J. Prod. Econ. 268, 109099.
Stranieri, F., Stella, F., Kouki, C., 2024b. Performance of deep reinforcement learning algorithms in two-echelon inventory control systems. Int. J. Prod. Res. 1–16.
Szeghalmy, S., Fazekas, A., 2023. A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. Sensors 23 (4), 2333.
Taparia, V., Mishra, P., Gupta, N., Kumar, D., 2023. Improved demand forecasting of a retail store using a hybrid machine learning model. J. Graphic Era Univer. 12 (1), 15–36.
Theodorou, E., Spiliotis, E., Assimakopoulos, V., 2023. Optimizing inventory control through a data-driven and model-independent framework. EURO J. Transport. Log. 12, 100103.
Tiacci, L., Saetta, S., 2009. An approach to evaluate the impact of interaction between demand forecasting method and stock control policy on the inventory system performances. Int. J. Prod. Econ. 118, 63–71.
Ulrich, M., Jahnke, H., Langrock, R., Pesch, R., Senge, R., 2021. Classification-based model selection in retail demand forecasting. Int. J. Forecast.
Wang, B.X., Japkowicz, N., 2004. Imbalanced data set learning with synthetic samples. In: Proceedings of IRIS Machine Learning Workshop, p. 435.
Wang, G., Gunasekaran, A., Ngai, E.W., Papadopoulos, T., 2016. Big data analytics in logistics and supply chain management: certain investigations for research and applications. Int. J. Prod. Econ. 176, 98–110.
Williams, B.D., Tokar, T., 2008. A review of inventory management research in major logistics journals. Int. J. Logist. Manag. 19 (2), 212–232.
Wolters, J., Huchzermeier, A., 2021. Joint in-season and out-of-season promotion demand forecasting in a retail environment. J. Retailing.
Yan, Y., Liu, R., Ding, Z., Du, X., Chen, J., Zhang, Y., 2019. A parameter-free cleaning method for SMOTE in imbalanced classification. IEEE Access 7, 23537–23548.
Zhang, X., Du, Q., Zhang, Z., 2022. A theory-driven machine learning system for financial disinformation detection. Prod. Oper. Manag. 31, 1–20.
Zheng, Z., Cai, Y., Li, Y., 2015. Oversampling method for imbalanced classification. Comput. Inf. 34, 1017–1037.
Zou, Q., Xie, S., Lin, Z., Wu, M., Ju, Y., 2016. Finding the best classification threshold in imbalanced classification. Big Data Res. 5, 2–8.