Original Article

POMS

Production and Operations Management
2025, Vol. 34(10) 3010–3025

DOI: 10.1177/10591478251328630
journals.sagepub.com/home/pao


Lead Time Prediction for Inventory Optimization With Machine Learning

Robin Reiners[1], **Christiane B Haubitz**[1] and **Ulrich W Thonemann**[1]

Abstract

Modern decision-support applications build on planning parameters such as lead time, price, yield, etc., which are maintained as master data. The accuracy of master data significantly influences the viability of such applications. However, the maintenance of master data is considered a tedious and error-prone task. In this study, we explore the effectiveness of machine learning techniques to improve the accuracy of plan lead times. We apply both unsupervised and supervised learning methods for creating lead time prediction models. We test our approach using historical data of a global equipment manufacturer. In a numerical analysis the calculated plan lead times are over 30% more accurate than current plan lead times in terms of mean-squared-error (MSE). This increased accuracy of plan lead times reduces inventory investment by approximately 7%.

Keywords

Lead Times, Inventory, Machine Learning, Master Data, Decision Support

Date received 12 December 2023; accepted 18 December 2024 after two revisions

Handling Editor: Eric Johnson

1 Introduction

Supply chain management has become an integral factor in competitive strategy to enhance organizational productivity and profitability (Li et al., 2006). Considerable research and effort have been devoted to formulating optimization methods that require master data to achieve high quality results. Maintenance of master data is a time-consuming, tedious and error prone task since the data can be volatile, ambiguous, and influenced by factors outside the company's scope (Escudero et al., 1999).

Artificial intelligence (AI) technologies have gained prominence in their application to business-related problems and have been applied to pattern recognition, inference and learning from experience (Brynjolfsson and McAfee, 2017). In supply chain management, AI applications have been developed and deployed to, for example, inventory control and planning, transportation network planning, and purchasing and supply management (Min, 2010).

This study is motivated by a problem faced by a global equipment manufacturer which we refer to as "the company." The company has annual sales of several billion euros and operates world-wide. To provide maintenance and repair services for specialized equipment, the service department holds spare parts in stock. While some stock-keeping-units (SKU) are produced internally, many are purchased from external suppliers with different lead times. To coordinate inventory levels across multiple levels in their supply chain, the company employs optimization methods for determining inventory control parameters, that is re-order points, base-stock levels and order quantities. The calculation of these parameters relies on plan lead times that are retrieved from their enterprise resource planning (ERP) system. Due to inaccurate plan lead times of some SKUs, many inventory control parameters are suboptimal, resulting in lower service levels or higher inventory than optimal.

In this article, we address plan lead times and analyze how well their accuracy can be improved by machine learning. We propose a method to derive plan lead times with machine learning and compare the performance of three different machine learning regression algorithms: linear regression, random forests, and gradient boosting. We benchmark their performance with classical statistical approaches, that is

[1]Department of Supply Chain Management and Management Science, University of Cologne, Cologne, Germany

Corresponding author:
Ulrich W Thonemann, Department of Supply Chain Management and Management Science, University of Cologne, Cologne, Germany.
Email: ulrich.thonemann@uni-koeln.de

single-exponential-smoothing (SES) and the historical average, as well as the current plan lead times from the company's ERP system.

Our results show that plan lead times estimated with machine learning predict lead times more accurately than currently employed plan lead times. The best performing machine learning model improves lead time accuracy in terms of MSE by over 30% compared to current plan lead times. We also analyze the accuracy of our method for SKUs with respect to their procurement frequency and we find that our approach offers significant gains even for newly sourced SKUs. For very frequently procured SKUs, our approach still offers substantial gains in accuracy compared to classical statistical measures, but the difference is smaller than for infrequently procured SKUs.

To quantify the expected business benefits from improved plan lead times, we conducted an inventory simulation on historical data. The results from the simulation show that inventory holding can be reduced by approximately 7% while maintaining the same service level.

The remainder of this paper is structured as follows. In Section 2, we review the relevant literature on lead time prediction. In Section 3, we outline the problem and identify requirements for improving plan lead times with machine learning. In Section 4, we present our method of both supervised and unsupervised learning. In Section 5, we examine the economic implications of improved plan lead times and present numerical results. In Section 6, we derive practical implications, outline limitations and conclude.

2 Literature Review

Lead times are an important parameter in inventory optimization, extensively discussed in literature (Muthuraman et al., 2015; Silver et al., 2016; Wang, 2012). However, the literature has paid little attention to analyzing lead time accuracy and predicting lead times, especially in the context of inventory planning and optimization.

In practice, plan lead times are often derived from supply chain contracts, employee experience, or computed from historical data (Grout and Christy, 1993; Lawrenson, 1986; Lingitz et al., 2018; Urban, 2009). In some papers, lead times are predicted but not for stock replenishment. Berlec et al. (2008) examine how to accurately predict product delivery times, a critical factor in effective contract negotiations. Their study focuses on identifying commitment-worthy lead times for customers, ensuring that contracts reflect achievable delivery schedules. Similarly, Duffie et al. (2017) and Yang and Geunes (2007) explore aspects of customer lead time within contractual contexts, they do not specifically address its prediction for inventory replenishment.

Some recent studies have made promising strides in the field of lead time prediction through analytics. While these explorations are noteworthy, they present opportunities for further refinement and validation within the peer-reviewed scholarly community. Banerjee et al. (2015) combine matrix gamma distribution and step-wise linear regression to predict lead times. de Oliveira et al. (2021); Liu et al. (2018) employ regression models and compare various machine learning techniques, demonstrating high accuracy in predicting supplier lead times. However, these studies do not address the unique challenges in inventory management posed by spare parts, which are seldom re-ordered, leading to scarcity in historical data at the SKU level. These studies also neither address implementation of the approaches nor the financial implications of implementing such predictive models.

To address these gaps, our research leverages machine learning for lead time prediction. Insights from studies in related domains, such as transportation (Barbour et al., 2018; Choi et al., 2016; Hofleitner et al., 2012) and manufacturing (Bender and Ovtcharova, 2021; Burggräf et al., 2020; Lingitz et al., 2018), indicate the effectiveness of ensemble decision tree models. Specifically, the studies by Bender and Ovtcharova (2021); Gyulai et al. (2018); Lingitz et al. (2018); Oeztürk et al. (2006) have consistently demonstrated the superior performance of ensemble decision tree models such as random forest and gradient boosting in predicting various lead times.

While the existing literature offers insights into prediction methodologies, several open questions remain regarding how these predictions can be leveraged to derive accurate plan lead times for inventory control. Existing research focuses on predicting lead times at the individual purchase order level, whereas inventory control requires a lead time parameter to estimate the lead time demand distribution to compute base stock levels, safety stock or reorder points. This gap between order-level lead time predictions and the need for suitable parameters for inventory management is addressed in this article.

Moreover, while the literature addresses how lead time variability impacts base stock levels and inventory investments, it is typically assumed that the lead time distribution is already known or can be estimated from historical data. This assumption is particularly problematic for spare parts, where data is often sparse, making it difficult to estimate moments from historical observations.

Building on existing research in lead time prediction, we extend the literature by examining how varying SKU order frequencies affect bias in predictive models. The inherent bias introduced by SKU order frequency discrepancies presents a key challenge in procurement operations. Frequently ordered SKUs generate extensive historical data, allowing for higher prediction accuracy in machine learning models. Conversely, infrequently ordered SKUs have sparse historical data, leading to greater uncertainty in lead time predictions. This imbalance reflects the operational nature of procurement systems, where demand varies widely among SKUs. We address this challenge through a comprehensive approach combining feature engineering techniques and systematic evaluation of balancing methods for regression tasks.

Building on these findings, our analysis focuses on three machine learning models: linear regression, random forest, and gradient boosting (Breiman, 2001; Friedman, 2002; Weisberg, 2005). By applying these models to predict supplier lead times accurately, we aim to develop a framework that not only enhances inventory planning and optimization but also addresses the unique challenges posed by spare parts, thus contributing significantly to the existing body of knowledge in the field.

3 Problem Description

To offer repair and maintenance services to its customers, the company operates a warehouse network for its spare parts. To efficiently manage inventories, the service division of the company regularly optimizes its inventory control parameters (such as re-order points and base-stock levels) under service-level constraints. Due to inaccurate lead time parameters, inventory levels have been sub-optimal.

When analyzing the historical lead times, we observe a high dispersion. Figure 1 provides some information on lead times at the company. This data and other data of the company had to be sanitized by the company's request, but the general insights and relative performance data that we report are not affected by the sanitation. Figure 1(a) shows the lead time distribution of all SKUs. We can see that the distribution is highly skewed with a long right tail. Figure 1(b) plots order frequencies against procurement lead times. While procurements with longer lead times generally belong to SKUs with lower order frequencies, no strong relationship exists between these variables. This is evidenced by a single SKU ordered approximately 280 times spanning lead times from 20 to 170 days. Figure 1(c) shows the accuracy of current plan lead times. We can see that the deviations from the planned lead times occur with similar magnitude in both directions. Figure 1(d) shows the order frequency of all SKUs. We find a highly skewed number of purchase orders per SKU. While a few SKUs are ordered at a high frequency, the majority of SKUs are ordered infrequently.

Plan lead times are maintained as master data. These plan lead times are static and are rarely updated. The vendor management and the inventory management teams manage and maintain the plan lead times on the basis of the contractually agreed lead times. For most SKUs, the plan lead time is agreed upon with the individual supplier. For the remaining SKUs, the plan lead times are generated by calculating the historical average lead times of the SKU. Considering how infrequently most SKUs are procured, estimates based solely on historical lead time data (interpolation) are limited by the available observations. Together with experts from the company, we discussed which data from their ERP might contain causal information about lead times. After this qualitative approach, we queried operational attributes from the ERP system. We worked with two different data sources: SKU master data and transactional purchase order data.

4 Methodology for Accurate Plan Lead Time Prediction

Our methodology is structured into three fundamental stages. The first stage involves a rigorous data preparation process. We briefly touch our cleaning and merging process of the raw data sets, ensuring the integrity of the data. Subsequently, we conduct an insightful initial analysis to identify features with potential significance for our predictive models. The second stage encompasses feature engineering. This step is dedicated to the creation and transformation of features to improve the performance of the applied machine learning techniques. In the final stage we build upon the prepared data set and apply machine learning algorithms to refine our predictive models. We define robust evaluation techniques, serving as benchmarks for comparing and selecting the most effective models. Through regularization and hyper-parameter optimization, we tailor the models to predict plan lead times accurately.

4.1 Data Preparation and Exploratory Analysis

We work with two different data sources: SKU master data and transactional purchase order data. SKU master data contain data such as the supplier from which the SKU is sourced, the country it is sourced from, and the current plan lead time. Transactional purchase data contain information on purchased SKUs over four years (2018–2021), including the request and delivery date and the plan lead time at the time when the SKU was ordered. Therefore, this data set holds information on how master data, such as plan lead time, has changed over time.

Most pre-processing of our data consisted of storing information as correct data types and identifying SKUs with missing entries. Since we could not fix missing values, these SKUs representing less then 2% of the total data set were dropped.

An analysis of categorical variables reveals that a large share of purchase orders is sourced from the country in which the global warehouse is located (Figure EC.3). We also observe countries of origin with only few purchase orders and therefore we clustered countries from the same continent with less than 1% of the order volume. The remaining categorical variables seem plausible and do not require further processing.

While assessing the numerical features of our data sets, we observe some extreme outliers for the variable's valuation price and purchase order quantity. We decide to winsorize them based on Tukey's rule of thumb (Tukey, 1992). This method is widely used in statistical analysis (Carling, 2000; Wiley and Wiley, 2019) and applied in machine learning to ensure models account for all data points while mitigating the impact of extremes, where models benefit from capturing the full range of data without being skewed by extremes. It identifies outlier values based on the inter-quartile range (IQR). Corresponding outliers are truncated and set to a constant value equal to the IQR times a factor, which, following Tukey's rule, we set to 1.5. For linear regression, we normalize numerical data by re-scaling features to the range of zero to one. We do not normalize the numerical data for random forest and

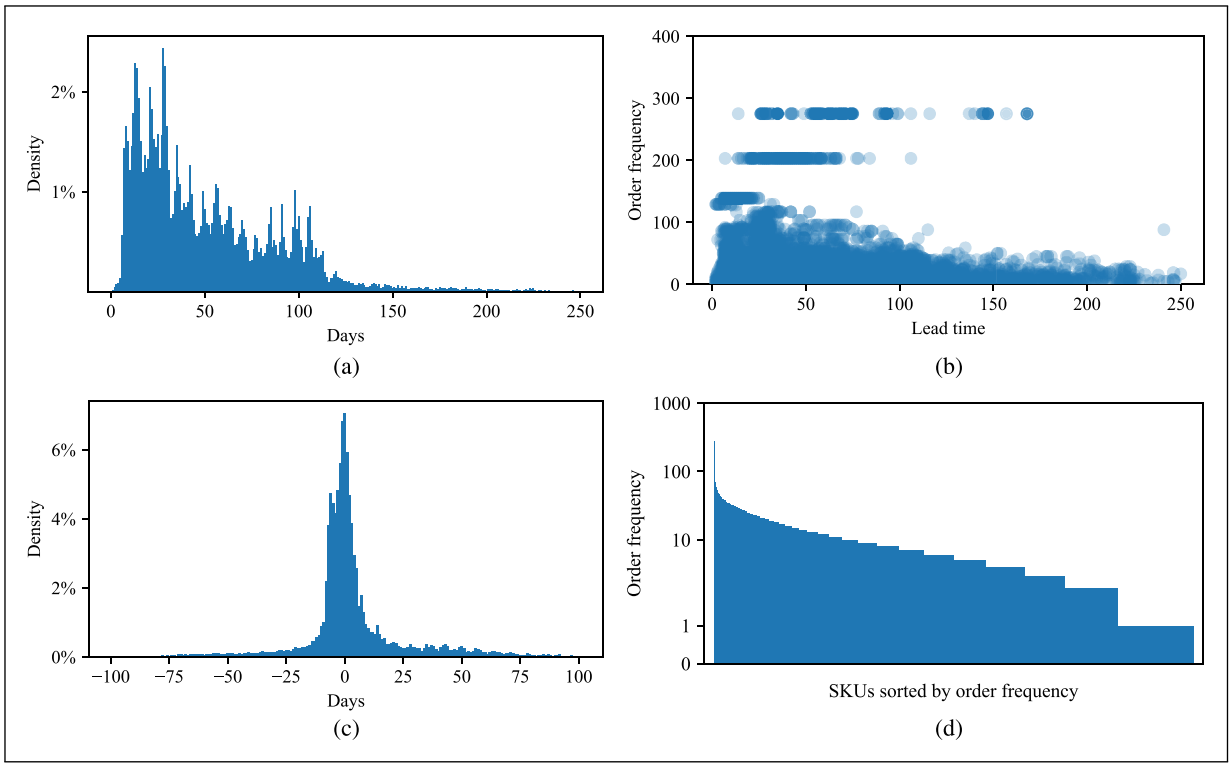

Figure 1. Plan and actual lead time characteristics of all purchase orders: (a) lead time dispersion; (b) order frequency against lead time; (c) plan lead time accuracy; (d) SKU order frequency.

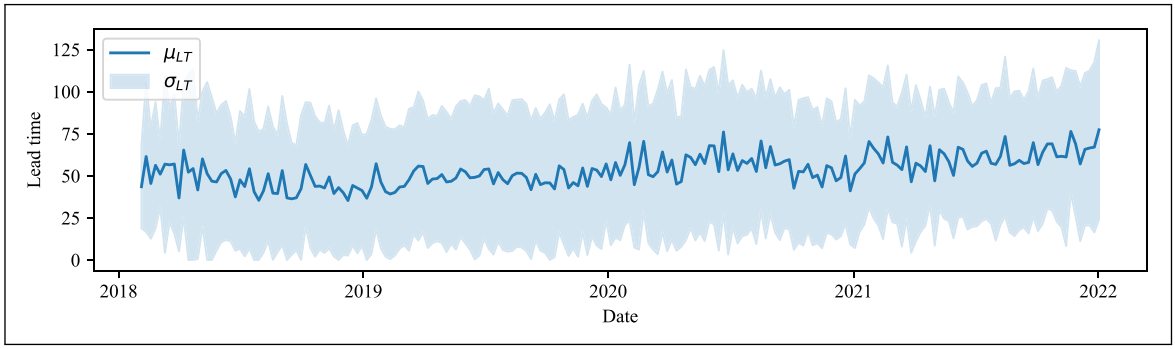

Figure 2. Lead time volatility over time.

gradient boosting, as tree based models do not require feature scaling (Brownlee, 2020).

Despite initial concerns about the potential impact of COVID-19 on lead-time volatility, our analysis did not uncover any significant fluctuations or patterns in the data, as illustrated in Figure 2. The lead times remain relatively stable over the observed period, with no extreme variances that would necessitate adjustments to the data set.

Our final data set includes 160,048 purchase orders across 17,728 SKUs, which are used for model training and testing (cf. Table 1). For performance evaluation, we implement a fixed origin evaluation procedure, splitting the data into a training set and a test set (Tashman, 2000). The first 42 months of data comprise the training set, while the final 6 months, selected to coincide with the company's semi-annual review period, form the test set.

To validate that the data of the training set do not deviate heavily from the test set, we compare both sets in Table 1. Key statistics of all features for each set indicate that both share similar characteristics.

Although the overall lead time volatility remains relatively consistent over time, we observe an increase of 6.67 days in average lead times compared to the training set. This shift

Table 1. Pre-processed features.

Column	Data type	Training set statistics		Test set statistics	
Purchase order	Category	Unique: 144,059		Unique: 15,989	
SKU description	String	Unique: 16,090		Unique: 7,667	
SKU ID	Category	Unique: 16,090		Unique: 7,667	
Supplier ID	Category	Unique: 331		Unique: 246	
Business unit	Category	Unique: 12		Unique: 12	
Business line	Category	Unique: 13		Unique: 13	
Business type	Category	Unique: 2		Unique: 2	
Business area	Category	Unique: 5		Unique: 5	
Criticality	Category	Unique: 3		Unique: 3	
ABC / XYZ	Category	Unique: 9		Unique: 9	
Country code	Category	Unique: 11		Unique: 11	
Requested in time	Boolean	Y (27 %)	N (73 %)	Y (36 %)	N (66 %)
Repairable part	Boolean	Y (4 %)	N (96 %)	Y (4 %)	N (96 %)
Manual correction	Boolean	Y (2 %)	N (98 %)	Y (1 %)	N (99 %)
Long tail lead time	Boolean	Y (6 %)	N (94 %)	Y (8 %)	N (92 %)
Valuation price	Float	μ: 331.17	σ: 571.21	μ: 335.41	σ: 565.82
Order quantity	Integer	μ: 9.31	σ: 13.73	μ: 11.38	σ: 15.35
Current plan lead time	Integer	μ: 48.32	σ: 39.16	μ: 54.14	σ: 41.21
End of support (days)	Integer	μ: 3,527.19	σ: 571.21	μ: 3,546.50	σ: 535.16
Lead time	Integer	μ: 49.37	σ: 39.16	μ: 56.04	σ: 44.60

does not indicate a broader pattern but rather reflects normal operational variations. We explored various splitting points as suggested by Hyndman and Athanasopoulos (2018), but ultimately, we chose the above mentioned split to maintain sufficient observations. We exercise the train-test split at this point to ensure that no information from the test set is used for feature engineering.

Table 2. Example count vectorization.

		Word counts					
SKU	Description	8	16	GB	DDR4	RAM	Memory
00001	8 GB DDR4 RAM	1	0	1	1	1	0
00002	16 GB DDR4 Memory	0	1	1	1	0	1

4.2 Feature Engineering

We next present feature engineering techniques to extract meaningful information from the existing data set and to transform this information so that machine learning models can process it.

In our data set, we have short descriptions of the SKUs. We anticipate that our models can learn from other SKUs with similar lead times. For example, electronic parts with a similar description of "8 GB DDR4 RAM" and "16 GB DDR4 Memory" can have similar lead times.

Strings cannot be interpreted by machine learning algorithms and we must convert them into ratio scale numeric representations that machine learning algorithms can process. We use KNN-regression to identify similar SKUs from the same supplier based on their description (unsupervised learning). For each SKU, we calculate a measure of lead time central tendency from its similar SKUs and use it as a feature for the machine learning algorithm.

Before starting with feature engineering, we have to preprocess some data. Because the SKU descriptions are not very extensive, most data cleaning, like removing stop words or expanding contractions, is redundant (Webster and Kit, 1992). We merely remove extra white spaces, lowercase all characters, and split the descriptions into single words (tokenization).

Because the SKU descriptions are concise, we use the bag of words count vectorizer, a vector space representation model for unstructured text (Zhang et al., 2010). From a total of N SKU descriptions, each SKU description n can be represented as a set of unique words (tokens) S_n. The corpus is the union from these sets $\cup_{n=1}^{N} S_n$. It includes every unique word from each SKU description and a single part description can be represented as a $(1 \times \cup_{n=1}^{N} S_n)$ vector, where each column corresponds to a specific word from the corpus. The value of each column is the frequency of how often that word appears in the description. Table 2 illustrates the concept.

We cluster SKUs which are sourced from the same supplier based on their cosine similarity. To evaluate if these clusters are meaningful to predict lead times, we analyze if SKUs from the same group have similar lead times. The part descriptions of the SKUs shown in Table 2, for example, have a cosine similarity of 0.83. For each SKU, we calculate its average lead time and compare it to the average of all historical lead times from the SKUs in the same cluster (cf. Figure 3(b)). We see that the average lead time of an SKU correlates with the average

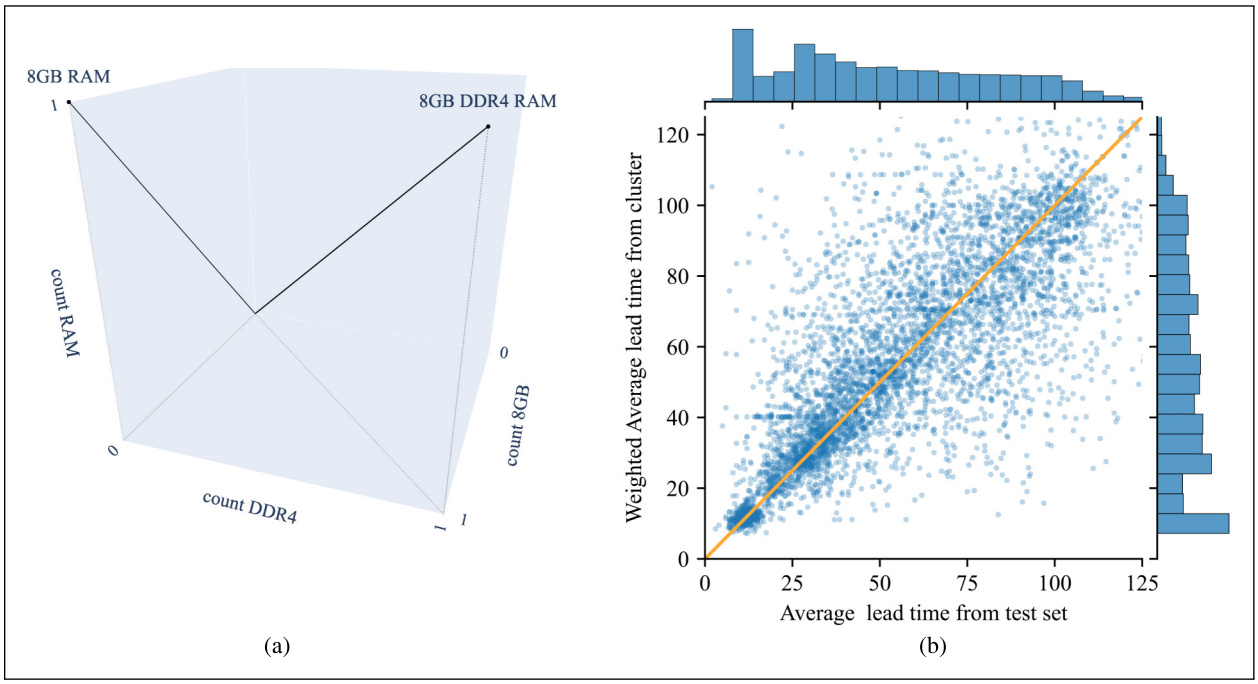

Figure 3. Creation of similarity feature: (a) 3D vector space cosine similarity; (b) average lead time correlation between SKU and cluster.

lead time of the SKUs from the same group. Therefore, we add the average lead time from the clustered SKUs as an additional feature. For SKUs which are not in any cluster, we take their average historical lead time. For those SKUs that have not been procured before we take the average lead times from the corresponding supplier.

As shown in Figure 1(a), the distribution of the target variable (lead time) is skewed, indicating an imbalance with a long tail extending towards longer lead times. To address this, we explore various balancing techniques such as random over- and under-sampling, as well as synthetic minority oversampling for regression, which have been proposed for handling imbalanced regression tasks (Branco et al., 2017; Torgo et al., 2013). In our analysis, these sampling methods do not improve results, probably because such techniques are more suited for contexts where the under-represented numeric value interval is most critical, and a wrongful prediction can be costly, often at the expense of overall performance (Torgo et al., 2013).

Conducting a point-biserial correlation analysis, we find a weak but significant negative correlation $r = -0.109, p \leq 0.001$ between lead times and punctuality (defined as a deviation of ± 3 days from the contractually agreed lead time). To improve the machine learning algorithm's ability to detect this correlation, we include a Boolean indicator for long lead times in the training set. We define long lead times as those in the 95 percentile, corresponding to a value of 119 days. This threshold represents the minority class and aligns with the typical range for rare events reported in the literature (Torgo et al., 2013).

As an additional feature, we compute a measure of supplier performance according to the business rules applied at the company. At the company, the supplier performance is measured as the share of orders delivered in time compared to all orders placed at the supplier. We first compute this measure for each supplier in the training set. Afterwards, we append this measure to the corresponding supplier in the test set. Some suppliers included in the test set do not appear in the training set. For those, we take the average supplier performance across all suppliers in the training set to complete the records.

As in most machine learning applications, the target variable is influenced by categorical (nominal scale) features. One-hot encoding, the most widely used coding scheme (Rodríguez et al., 2018), but in our case leads to undesirable sparsity in the data, especially with numerous categories such as the supplier ID (Gupta and Asha, 2020; Prokhorenkova et al., 2017). To address this, we implement multiple Bayesian encoding techniques: target encoder, polynomial encoder, Helmert encoder, James-Stein encoder, m-estimate encoder, weight of evidence encoder and catboost encoder—for a full reference, please refer to Pedregosa et al. (2011); Prokhorenkova et al. (2017). We choose the one which performs best in cross-validation, an approach that we cover in the next section.

4.3 Model Optimization and Evaluation

Informed by our literature review, we train three established machine learning algorithms to predict lead times: linear regression, random forest, and gradient boosting. Linear regression serves as our baseline, leveraging its capacity for

modeling relationships between scalar responses and explanatory variables through a least squares approach (Weisberg, 2005). The random forest algorithm, aggregates predictions from a multitude of decision trees trained in parallel on random data subsets to enhance predictive accuracy and mitigate over-fitting (Breiman, 2001). Gradient boosting constructs an additive model in a forward stage-wise fashion, and is particularly adept at addressing the residuals of preceding trees, sharpening accuracy on more complex or noisy data sets (Friedman, 2002).

Our primary performance metric is the MSE as the performance of an inventory control system heavily depends on achieving a low MSE in lead time demand forecasts (Syntetos et al., 2009). The MSE penalizes large deviations from observations disproportionately compared to minor deviations.

However, we recognize that the MSE can be challenging to interpret due to its dependence on the scale of lead time, especially when assessments are conducted across various SKUs. To address this, we also report the mean-absolute-percentage-error (MAPE), a scale-free metric, which prevents high or low performance in terms of MSE from being disproportionately influenced by a few SKUs. In addition to MSE and MAPE, we also evaluate the bias (mean forecast error). Evaluating bias is crucial because even with high accuracy, a model that consistently over- or underestimates can lead to poor inventory performance.

As discussed in Section 3, the procurement frequency of SKUs in our dataset is highly skewed, potentially leading to bias for infrequently ordered items. To address this, we explore balancing techniques such as minority over-sampling (OS) and majority under-sampling (US), as well as sample weighting based on the inverse of order frequency. These methods aim for a more balanced representation of SKUs and try to improve the model's ability to generalize across both frequently and infrequently ordered items. However, our results indicate that these techniques do not consistently improve model performance or significantly reduce bias. Therefore, we do not use these methods in the main analysis. Detailed results from the experiments are provided in Table EC.2 of the e-companion for reference and transparency.

To determine the effectiveness of these techniques, as well as to optimize hyper-parameters and perform feature elimination, we calibrate our models using a cross-validation procedure based on a rolling forecasting origin, applied to the training data (Hyndman, 2014). This technique takes into account the temporal nature of our data. The process involves dividing the training data set into several subsets, or 'folds,' considering the sequential order of the data. During each iteration, one fold is reserved for validation, while the remaining folds are used for training the model. This sequential approach ensures that the model is trained on past data and tested on future data. For our analysis, we implement the commonly used $k = 5$ folds.

In the initial configuration, we evaluate a base machine learning model, utilizing all available features and default hyper-parameters. This configuration serves as a benchmark for assessing model performance. To improve the model, we then perform hyper-parameter tuning through a grid search in a cross-validation framework, optimizing the model's performance with the full feature set (Kuhn and Johnson, 2019).

For regularization, we use the recursive feature elimination method (Athey and Imbens, 2019). In an iterative process, the models are trained on the full feature set. In each iteration, the one feature with the lowest importance score is eliminated. This step is repeated until all but one feature has been removed. In order to find appropriate hyper-parameter values we use a grid search approach after every feature elimination round. The grid design includes 100 parameter combinations (Santner et al., 2003). For our final model we choose the feature sub-set and hyper-parameter configuration that minimizes the MSE averaged over the five folds (cf. Figure 4). A summary for each algorithm of the included features, as well as hyper-parameters and their optimal values can be found in Table EC.1 of the e-companion.

To estimate plan lead times, we need to derive a representative purchase order from historical purchase orders for each SKU. We pass the representative purchase order's feature set into our trained machine learning model and use the response value as new plan lead times. The approach is rather trivial: To calculate a representative purchase based on information from the train-set, we take averages for the numerical values. For categorical values we take the most frequent value.

5 Numerical Results

5.1 Prediction Accuracy

We analyze the accuracy of our models in predicting purchase order lead times for our test set. 490 SKUs were newly introduced, for which traditional forecasting techniques like SES can not be applied due to the absence of historical data. Figure 5 illustrates the prediction errors for different lead time forecasting methods. While the top graphs show the MSE, the graphs in the middle show the MAPE, and finally the bottom graphs show the bias. The metrics for the three graphs on the left were calculated for all 7,667 SKUs and the metrics for the graphs on the right were calculated for the subset of 7,177 SKUs for which historical data exists.

To analyze our numerical results, we first focus on MSE for SKUs with demand in the training set, as shown in Figure 5(b). The data indicates that algorithms, in general, provide more accurate lead time predictions than those from the ERP system. In bench-marking classical methods against regression methods, the latter demonstrate improved MSE outcomes compared to methods such as SES and historical averages. SES and historical averages perform similarly. The inclusion of additional features in regression models appears to contribute to their enhanced predictive performance. Within the category of regression models, it is the machine learning models, gradient boosting and random forest, that further enhance the predictive accuracy. Within this subset, random

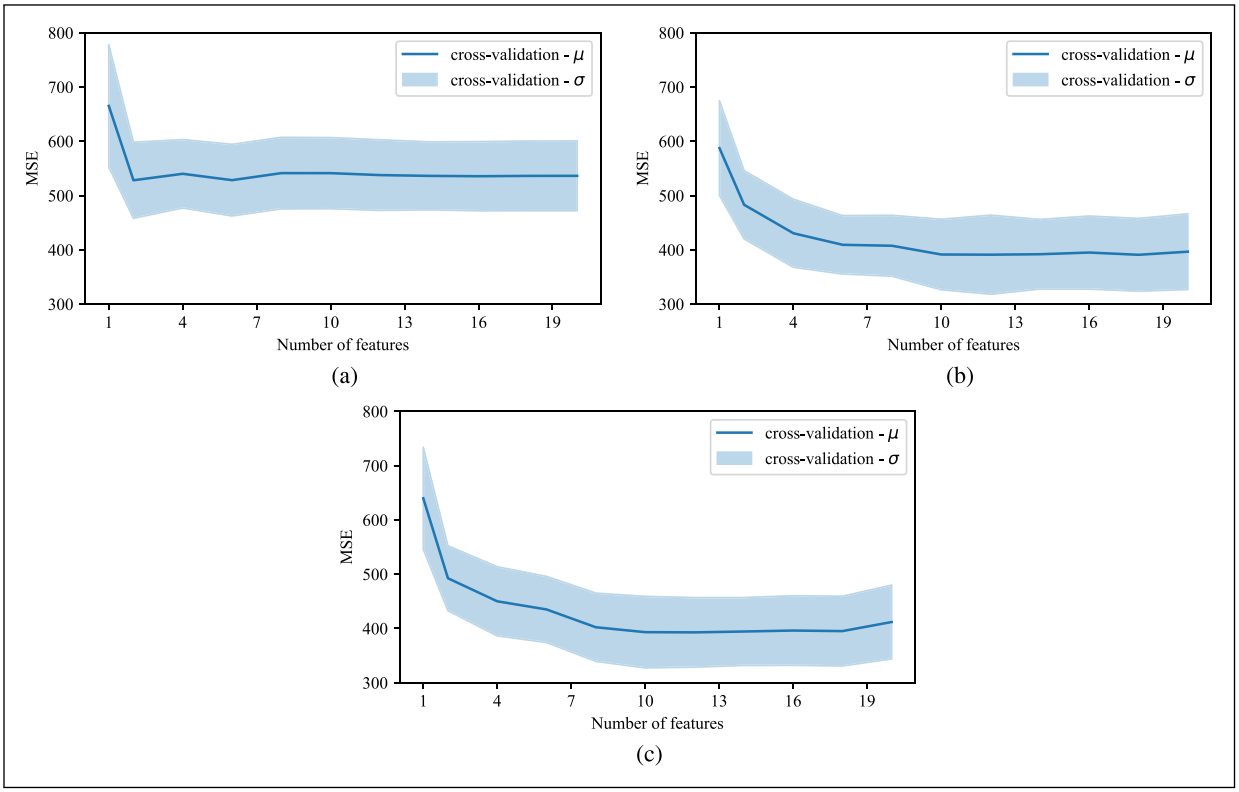

Figure 4. Cross-validation results of reverse-feature elimination with hyper-parameter tuning process: (a) linear regression; (b) gradient boosting; (c) random forest.

forest stands out, improving MSE by 34.84% over current ERP plan lead times, marking a significant improvement in forecast accuracy.

After confirming the effectiveness of the random forest model in terms of MSE for SKUs with demand in the training set, we examine the other graphs for consistent patterns. Upon extending the analysis to all SKUs as shown in Figure 5(a), consistency in algorithmic performance is observed even when evaluating SKUs without demand in the training set, with the random forest maintaining its lead in predictive accuracy. A similar improvement is observed in Figure 5(c) and (d). The random forest model not only lowers MSE but also reduces MAPE by 2.62 and 3.11 percentage points, respectively, which corresponds to improvements of 8.36% and 9.94%. These reductions underscore the model's increased predictive accuracy, confirming that the improvements are not disproportionately affected by outliers.

Continuing the analysis, we observe that both historical averages and SES exhibit a notable negative bias, as shown in Figure5(f), aligning with expectations given the approximate 6-day shift in average lead times between the training and test sets (cf. Section 4.1). These methods inherently lack the capability to adapt to future developments. Consequently, their performance during inference is suboptimal. In contrast, current plan lead times (ERP) incorporate some level of foresight by reflecting temporal adjustments made by planners or adjustments derived from contractual obligations. This foresight allows for the anticipation of potential shifts. However, these plan lead times systematically underestimate actual lead times, as shown in Figure 5(e) and (f), indicating a persistent negative bias.

Regression models, particularly machine learning approaches like gradient boosting and random forest, are designed to identify and adjust for underlying data patterns. This capability enables them to effectively "debias" their predictions, as evidenced by the minimal bias observed even amidst slight shifts between training and test data. The high accuracy and low bias exhibited by these models demonstrate their robustness and their ability to generalize effectively to new data.

By evaluating model performance in relation to the order frequency in the training set, we gain insights into the reliability of our predictive methods for SKUs with varying levels of historical data. Figure 6 presents the performance of different forecasting methods against the procurement frequency of SKUs from the training set. The frequency categories range from "new," indicating no historical procurement data in the train-set, to "monthly," indicating frequent procurement. The graph plots the performance of our best performing

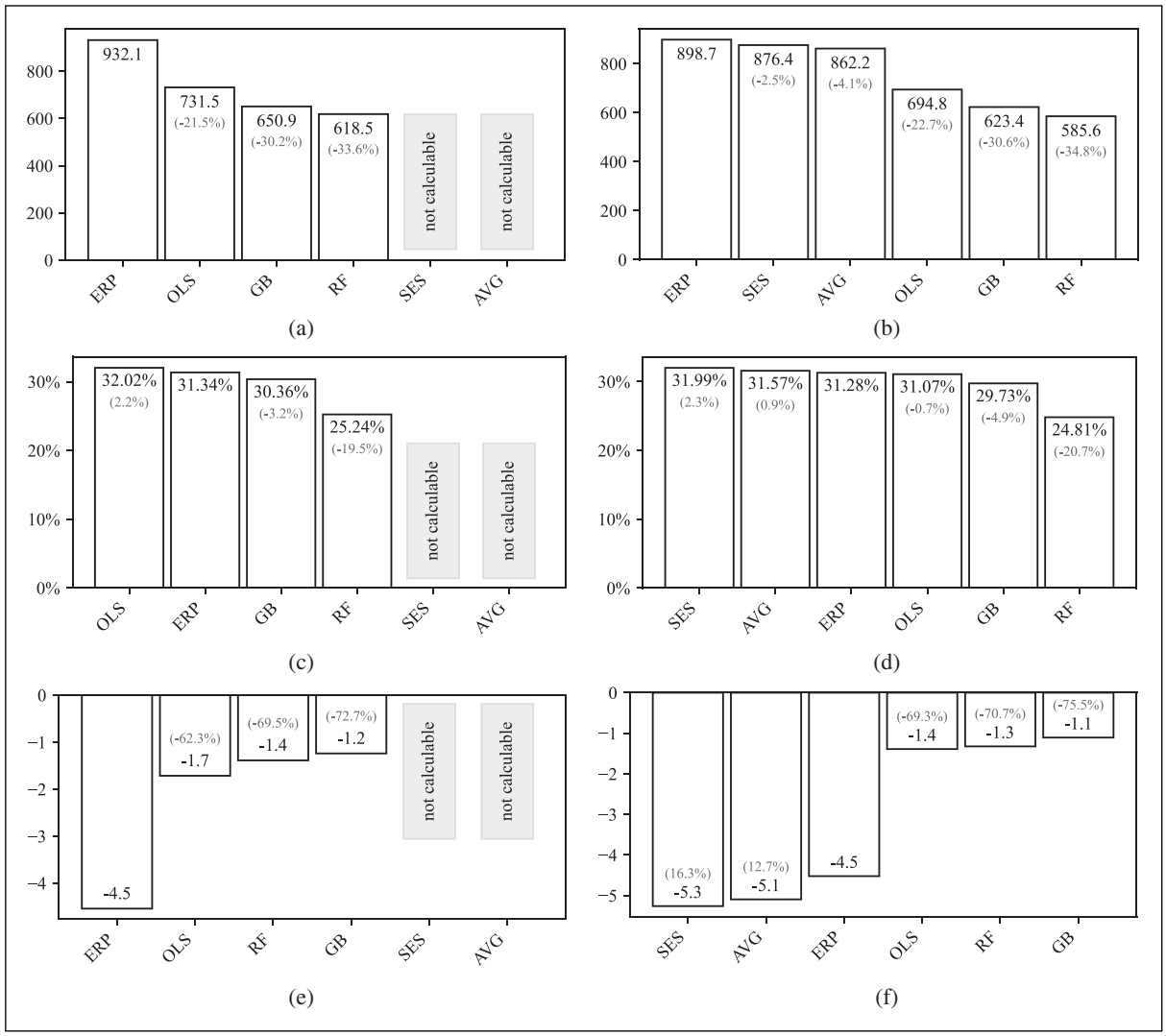

Figure 5. Accuracy of different potential plan lead times: (a) all SKUs ($n = 7,667$) - MSE; (b) SKUs with demand in training set ($n = 7,177$) - MSE; (c) all SKUs ($n = 7,667$) - MAPE; (d) SKUs with demand in training set ($n = 7,177$) - MAPE; (e) all SKUs ($n = 7,667$) - bias; (f) SKUs with demand in training set ($n = 7,177$) - bias.

machine learning model (random forest), current plan lead times (ERP), as well as conventional statistical methods (SES and historical averages). It illustrates a trend towards improved performance as the procurement frequency increases, signifying better accuracy with more data.

Notably, the random forest model consistently outperforms both the current plan lead times and classical statistical measures across all performance measures and almost all frequency categories. For instance, for new SKUs our method delivers an MSE roughly 25% lower than that of the current plan lead times. This significant reduction in error is indicative of the robustness of our method, particularly for SKUs without prior order history.

To generate some insights into how the features contribute to the prediction, we show the feature importance for random forest and gradient boosting in Figure 7. It becomes evident that the current plan lead time is a dominant feature, suggesting that the models heavily utilize this information to refine their predictions.

However, the significance of the "Long tail lead time," "Cluster feature" and the advanced encoding of "Part ID" and "Supplier ID" emphasize the critical role of feature engineering in enhancing model performance. It indicates that the models are not solely reliant on current lead time parameters and historical lead times but also on the nuanced interplay of SKU characteristics. These features capture underlying patterns and similarities across SKUs that are otherwise not directly observable, thus providing powerful predictors that complement and enhance the historical data. Their substantial role in both models underscores the effectiveness of our

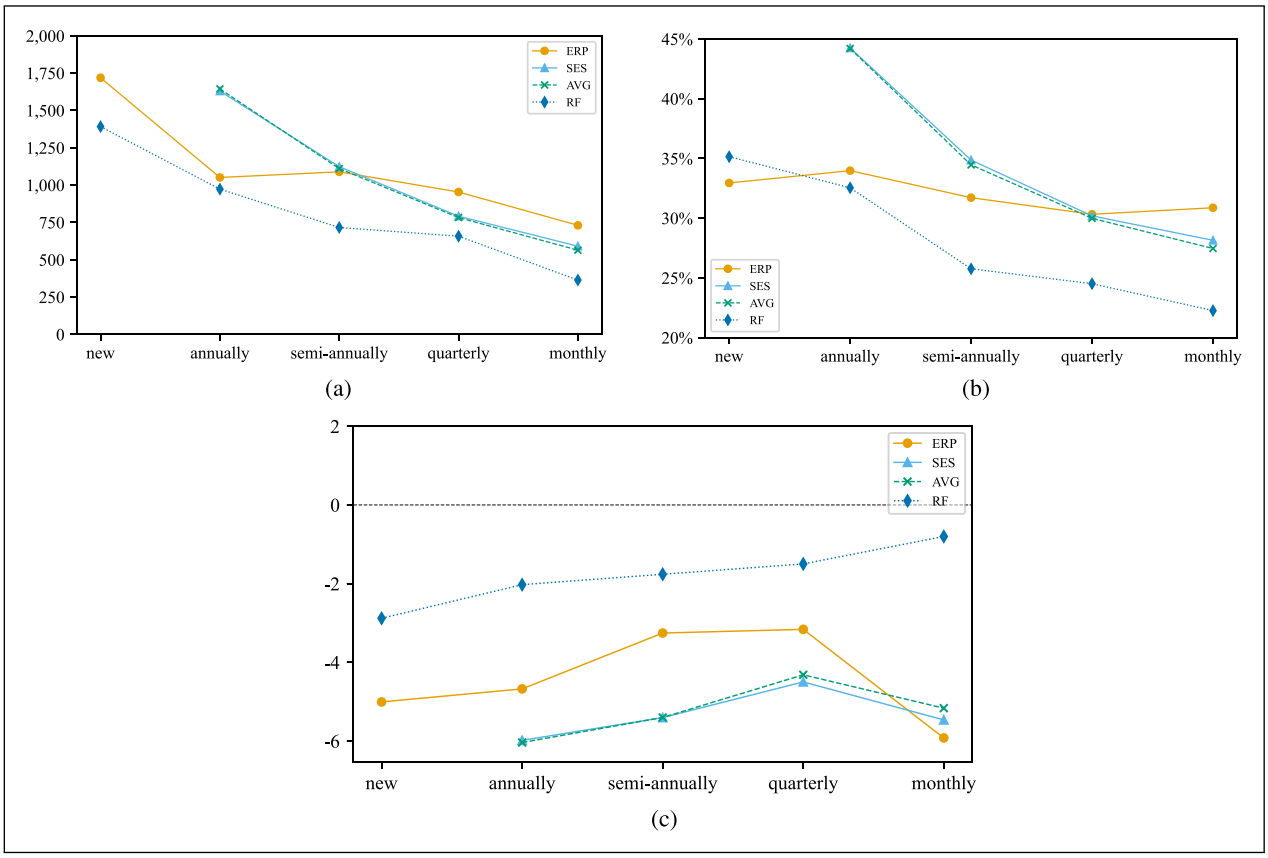

Figure 6. Accuracy of different potential plan lead times relative to order frequency: (a) MSE; (b) MAPE; (c) bias.

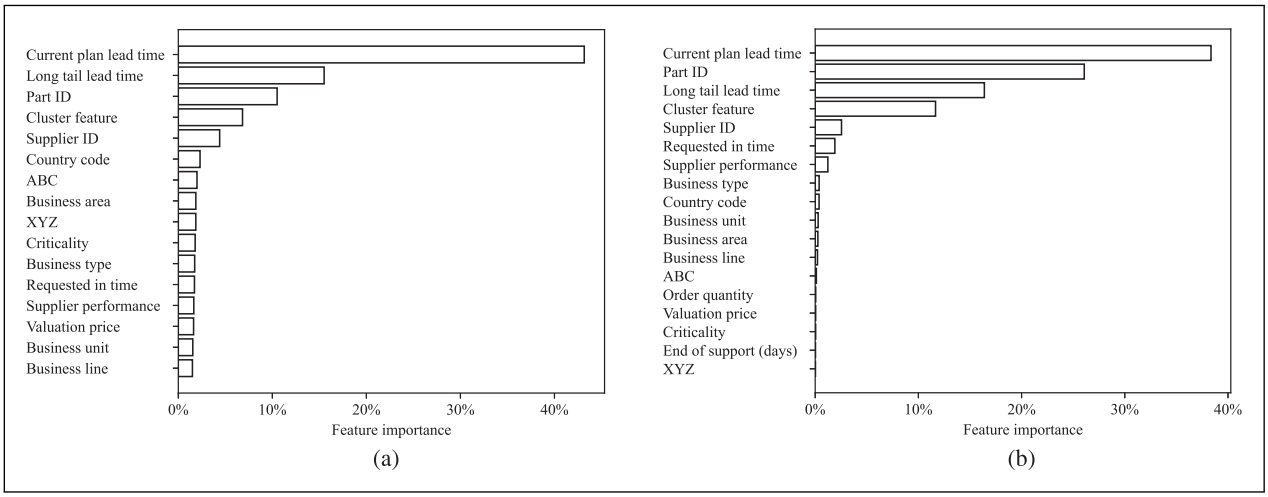

Figure 7. Feature importance: (a) random forest; (b) gradient boosting.

feature engineering process, greatly enhancing the model's ability to forecast lead times accurately.

In conclusion, the findings suggest that machine learning regression models, particularly the random forest, offer substantial improvements over traditional statistical methods and current ERP planning estimates in predicting lead times.

5.2 *Inventory Performance*

To quantify the expected business benefits of increased plan lead time accuracy we conduct a simulation study. We evaluate how the more accurate plan lead times also improve the company's inventory performance. In our simulation, we compare how inventory levels evolve, using (a) the current plan

lead times, and (b) the improved plan lead times. We use actual historical lead times for our benchmark. Our purchase orders hold information on when replenishment orders were submitted and when they arrived. We can simulate a realistic inventory development on an SKU level by modeling a periodic review base-stock policy with base-stock level S (Silver et al., 2016). This inventory control policy is applied in many real-world spare parts inventory systems (Boylan and Syntetos, 2010; Cavalieri et al., 2008; Syntetos et al., 2012; Wang, 2012) and also used by the company.

The intraperiod timing assumptions for the base-stock policy are as follows: each SKU is reviewed daily, and if the inventory position is below the base-stock level S, a corresponding replenishment order is submitted to the supplier. After submitting the replenishment order, previous replenishment orders may arrive with lead times according to historical data. Subsequently, potential back-orders are fulfilled. Then demands d_t arrive and are either fulfilled or back-ordered. We assume that lead times and demands are independent from each other.

To determine the appropriate base-stock level S that meets the target β service level, the model must account for uncertainties in both demand and lead time. Traditionally, this uncertainty is addressed by modeling the shortfall, which includes variability in order arrival sequences. In the presence of stochastic lead times, order crossover can occur, where orders are received in a different sequence than they were placed. This may result in higher inventory levels when earlier orders arrive. As reported by Robinson et al. (2001) neglecting order crossover under order-up-to inventory policies can lead to significantly higher inventory costs. However, we do not consider the effect of order crossover due to its complexity, its limited relevance to our primary objectives, and its low occurrence rate in our data set (0.18% of all purchase orders). Most SKUs are sourced from a single supplier with long intervals between orders relative to the variability in lead time.

Therefore, we use the lead time demand distribution to calculate the base-stock level S^*. Lead time demand describes the cumulative demand in the risk period with the duration of the lead time plus the review interval $D_{LT+1} = \sum_{t=1}^{LT+1} D_t$. Employing a base-stock policy, our objective is to ensure that a certain percentage of lead time demand is satisfied from stock (β service level). To determine the appropriate base-stock level S a probability function of lead time demand exceeding S is required.

For fast-moving items, the assumption of normally distributed lead time demand is typically reasonable. However, our SKUs often exhibit intermittent demand patterns that the normal distribution may not represent. This is almost invariably the case for spare parts (Turrini and Meissner, 2019). For slow-moving items the poisson distribution typically offers a good fit for the arrival of spare part demand (Silver et al., 1998). For demands that are not unit-sized but of constant size the resulting distribution of demand per period can be modeled by a "package Poisson" distribution (Ritchie and Kingsman, 1985). If demand size is not constant, it is reasonable to assume that demand arrivals are poisson distributed, and the order size follows a logarithmic distribution. In such cases, total demand is negative binomial distributed over lead time (Quenouille, 1949). For our simulation, we use negative binomial distributions to model our spare part demand, noting that the Poisson distribution is a special case of this distribution.

To compute the moments of the distributions, it is necessary to forecast both lead time and demand. We utilize the predicted lead times to estimate the average lead time, denoted as μ_{LT}. To quantify lead time uncertainty, we employ the root-mean-squared-error (RMSE) as an estimator for the standard deviation of lead time, denoted as σ_{LT}, following the methodology suggested by Barrow (2016). Expected demand, μ_{D}, and demand variance, σ^2_{D}, are derived from the purchase order data within the training dataset.

The moments of the lead time demand distribution are calculated using established formulas for the mean and variance of the sum of a random number of random variables $\mu_{\mathrm{LTD}} = \mu_D \cdot \mu_{\mathrm{LT}}$ and $\sigma^2_{\mathrm{LTD}} = \mu_{\mathrm{LT}}\sigma^2_D + \mu^2_D\sigma^2_{\mathrm{LT}}$ (Zipkin, 2000: 285).

To determine the appropriate base-stock level, we aim to ensure that the expected back-orders during a replenishment cycle do not exceed a specified fraction of demand within the same cycle. This objective is formalized through the following optimization problem (Sieke et al., 2012):

$$\min S$$

$$\text{s.t.} \left[\sum_{y=S+1}^{\infty} (y - S) \cdot f_{\mathrm{LT+R}}(y) - \sum_{y=S+1}^{\infty} (y - S) \cdot f_{\mathrm{LT}}(y) \right] \leq (1 - \beta)\mu_{\mathrm{R}}$$

Here, $f_{\mathrm{LT+R}}$ and f_{LT} are the demand distributions over lead time plus review period, and lead time, respectively.

We use S^*_{ML} to refer to the base-stock level based on the plan lead time derived from our proposed machine learning method, and S^*_{ERP} to refer to the base-stock level based on the current plan lead time.

For our simulation, the initial inventory position starts at the respective base-stock level S^*_{ML} or S^*_{ERP}. Assessing inventory performance right away might be unrepresentative, as the system has not had time to settle into its steady-state. Therefore, we do not measure inventory performance for the duration of the training set but only evaluate on the test set.

The amount of capital lockup has been scaled linearly in order to not reveal information about the company. Figure 8 summarizes our simulation results. The trade-off curve shows the capital lockup for a given realized service level. The grey curve shows inventory performance based on base-stock levels S^*_{ERP} (β), while the black curve describes inventory performance based on base-stock levels S^*_{ML} (β).

The black curve is constantly below the grey curve, implying that compared to the current plan lean times, our approach leads to stock levels that achieve higher service levels while having the same capital lockup. To derive some financial implications of our proposed method, we investigate inventory

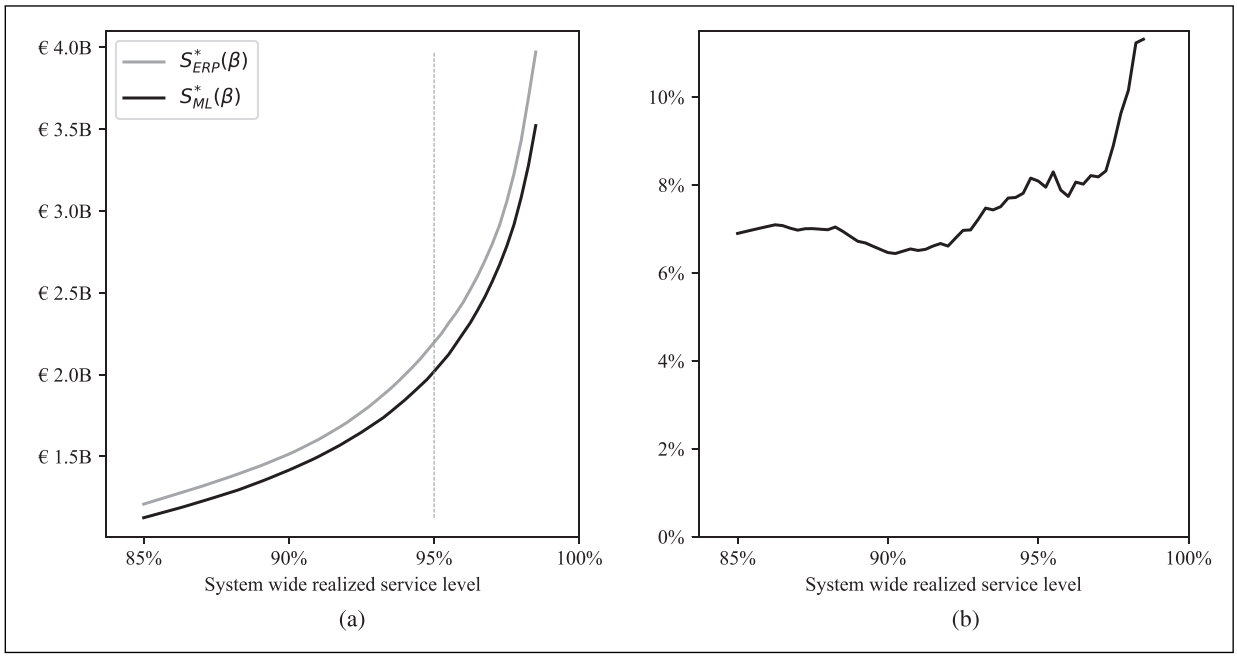

Figure 8. Simulation results: (a) capital lockup for a given realized service level; (b) capital savings (%) for a given realized service level.

performance for an achieved 95% service level. We find that to attain the targeted service level, base-stock levels S^*_{ML} require roughly 7% less capital than base-stock levels S^*_{ERP}.

6 Discussion

6.1 Implications

The implications that we derive from our results provide recommendations for implementing lead time prediction approaches with machine learning and how they can be used in practice to exploit potential benefits.

Effective feature engineering can provide significant benefits. In our spare parts setting, data on SKU level is often limited. We successfully develop a text classification algorithm to incorporate information from similar SKUs based on their descriptions. Additionally, we enrich our data set with general information on supplier performance. State-of-the-art algorithms allow us to train our models on sparse categorical features. While conducting our study, we have found that there is a relationship between the extent of information extraction from the available data and the accuracy of our models. The feature importance analysis in Section 4.2 has clearly indicated that the additional features derived from feature engineering significantly contribute to accuracy improvements for all machine learning models we investigate.

Machine learning enables a more accurate prediction of lead time solely based on observed purchase orders. There are very few studies attempting to predict lead times solely based on the limited information available at a company. Companies often depend on (a) lead times specified in supply contracts, (b) employees' experience, or (c) average values based on historical data. We have seen that (a) lead times specified in supply contracts can significantly deviate from actual lead times. (b) Relying on expert knowledge can result in predictions being biased by stock-out aversion. Additionally, by relying on human experience, the company is exposed to the risk of unique human knowledge leaving the company in employee churn. (c) Average values based on historical data have been shown to be limited by the frequency of observations.

Our results align well with current evidence from an emerging stream of literature addressing the viability of machine learning applications for lead time prediction. We contribute to this literature stream by confirming that machine learning applications can reliably predict lead times with high accuracy. This is an opportunity for many companies to predict lead times operationally and take short-term measures. Indications as to whether an order arrives earlier or later than expected can be used to manage critical inventories tightly.

Plan lead times derived from machine learning are particularly valuable for procurement structures with large number of different SKUs with only a few purchase orders each. We have shown that compared to other alternatives, our method performs best for SKUs that are more frequently procured than once every half a year. This demonstrates a major advantage of our machine learning approach to determine plan lead times compared to methods that rely solely on past data: the data for available SKUs can be exploited to predict lead times for SKUs with an insufficiently large set of historical data. With increasing procurement frequency the accuracy of other methods converge, due to many observations being available from which estimates can be derived.

Plan lead time accuracy improves inventory performance. When lead time is used in the planning process to set inventory control parameters, increased lead time accuracy is expected to improve business outcomes. The business benefits of increased plan lead time accuracy are challenging to quantify since they are based on adjusted processes to reflect the improved accuracy. Therefore, we simulated inventory developments based on current and improved plan lead times. We estimate that for achieving a targeted service level of 95%, our approach is expected to reduce capital lockup by 7%.

While the findings in our study are promising, it is important to discuss the generalizability of these results. Our primary goal is not to identify a universally superior model but to demonstrate the benefits of applying machine learning techniques to lead time prediction in contexts characterized by sparse data and lead time uncertainty, such as spare parts management. A distinct and important contribution of our paper, compared to the existing literature, is our focus on predicting the lead time distribution from the supplier's perspective. Suppliers, who order from various manufacturers, often have limited information about the production state of products and manage a wide variety of product types. Our methodology addresses this complexity by using machine learning to effectively handle diverse and incomplete data, enabling more accurate predictions of lead times.

The higher accuracy of the random forest model in our study can be attributed to its robustness against skewed feature distributions and its ability to handle complex, non-linear relationships. However, we recognize that these findings may not directly apply to all contexts. Supply chain environments vary significantly, with differences in data availability, supplier behavior, and SKU characteristics. Consequently, no single model is likely to consistently outperform others across all scenarios. Therefore, while random forest demonstrates high effectiveness in our setting, we do not claim it will always be the best-performing model.

Our findings emphasize the critical importance of robust feature engineering and the incorporation of domain-specific knowledge. These elements are likely to be more decisive in the success of any machine learning model for lead time prediction than the specific choice of algorithm. The general methodology we use is likely to yield good results in similar contexts where suppliers face uncertainty due to diverse products and limited visibility into production processes. This adaptability underscores the broader applicability of our approach beyond our specific study.

Our study offers a scalable and solid methodology that can be adapted by researchers and practitioners alike to address the challenge of lead time prediction in various contexts, particularly those involving spare parts with limited historical data. As more companies develop in-house data science capabilities, implementing these advanced techniques becomes increasingly feasible and less resource-intensive. Although implementing data collection processes and operationalizing machine learning pipelines may initially seem complex and costly compared to simpler methods like SES, the significant accuracy improvements demonstrated in this study make a strong case for the investment.

We have successfully implemented this machine learning tool within a company, demonstrating its practical applicability and tangible benefits. The long-term cost savings from more accurate lead time predictions, as evidenced by our simulation, are likely to outweigh the initial implementation costs. The strategic advantage gained through improved inventory management and procurement processes justifies the complexity of the machine learning methods used. This approach provides a robust framework for companies to enhance operational efficiency and make informed decisions, thereby transforming how lead time predictions are handled in practice.

6.2 Limitations

In this study, we derived accurate plan lead times for SKUs. These plan lead times are point estimates of lead time which are used by optimization tools for inventory planning to set inventory control parameters. While many firms use inventory management software ignoring lead time uncertainty (Dolgui et al., 2013), lead time fluctuations strongly degrade the tools' performance and cause high inventory costs. Babai et al. (2022); Chopra et al. (2004); Song et al. (2010), as well as others, have demonstrated the economic implications of lead time uncertainty. They have shown the importance of accounting for its effects regardless of the procedures used to compensate for demand uncertainty.

In addition to our proposed machine learning approach, it is important to acknowledge that alternative or complementary strategies, such as continuous improvement methodologies, could be employed to improve plan lead time accuracy. Continuous improvement could involve identifying and focusing on the most critical or problematic parts and suppliers to either better predict lead times or create robust processes that can respond to the uncertainty. By integrating such continuous improvement efforts with machine learning models, companies can potentially enhance their overall planning accuracy and operational resilience.

In this study, we also illustrate the impact that ill-estimated lead times can have on inventory performance. Yet, we only cover the forecasting of point estimates of lead time. Modern optimization tools account for lead time uncertainty and depend on estimates on the lead time uncertainty, that is, the variance of the response value (Cachon and Terwiesch, 2008; Silver et al., 2016; Song et al., 2010).

In the future it would be great if the methods we describe in our article, were used for non-parametric conditional density estimation (Bertsimas and Kallus, 2020; Dalmasso et al., 2020; Pospisil and Lee, 2018). The conditional density estimation could be used to estimate lead time demand directly fitting into recent research of Babai et al. (2022); Boylan and Babai (2022). Their research is dedicated to estimating lead

time demand distributions directly, albeit considering only univariate lead time samples.

6.3 Conclusion

In this study, we addressed the challenge of maintaining inaccurate master data on which many inventory optimization models rely. We investigated with special regard to plan lead times how well the maintenance of this planning parameter can be automated by means of machine learning.

We applied unsupervised and supervised machine learning in our lead time prediction approach. Based on natural language processing techniques, we identified similar SKUs and used this information to generate features for our model. To derive plan lead times from our machine learning models, we used the features of representative purchase orders to generate a response value from our model. This value can be used as plan lead time.

We tested our approach on historical data of a global equipment manufacturer. We bench-marked the results against current plan lead times from the company as well as simple statistical measures of central tendency. We found that our approach significantly increases the accuracy of plan lead times by over 30% compared to current plan lead times. Simulations have shown that the increased accuracy of plan lead times reduces capital lockup by approximately 10% while increasing the likelihood of achieving the targeted service level for a given capital lockup. This translates into significant cost reduction for inventory-keeping as well as into reputational benefits due to fewer back-orders. Moreover, we developed a method to predict lead times for new SKUs, which can be used as decision support for contract negotiations with the suppliers. With specific regard to plan lead times, machine learning has proven to be a viable technique for improving the quality of master data. Thus our method creates tangible and intangible added value by automating a notoriously tedious and error-prone task.

Acknowledgments

The authors are grateful to the editors and the anonymous review team for the insightful guidance provided throughout the revision process. The authors would also like to thank all research participants and organizations involved in this research study for their support.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Robin Reiners https://orcid.org/0009-0008-8873-9403
Christiane B Haubitz https://orcid.org/0000-0002-7126-0606
Ulrich W Thonemann https://orcid.org/0000-0002-3507-9498

Supplemental Material

Supplemental material for this article is available online (doi: 10.1177/10591478251328630).

References

Athey S and Imbens GW (2019) Machine learning methods that economists should know about. *Annual Review of Economics* 11(1): 685–725.

Babai MZ, Dai Y, Li Q, Syntetos A and Wang X (2022) Forecasting of lead-time demand variance: implications for safety stock calculations. *European Journal of Operational Research* 296(3): 846–861.

Banerjee AG, Yund W, Yang D, Koudal P, Carbone J and Salvo J (2015) A hybrid statistical method for accurate prediction of supplier delivery times of aircraft engine parts. In: *International design engineering technical conferences and computers and information in engineering conference*, Boston, MA: American Society of Mechanical Engineers, V01BT02A037.

Barbour W, Samal C, Kuppa S, Dubey A and Work DB (2018) On the data-driven prediction of arrival times for freight trains on U.S. railroads. *21st International conference on intelligent transportation systems (ITSC)*, pp.2289–2296.

Bender J and Ovtcharova J (2021) Prototyping machine-learning-supported lead time prediction using autoML. *Procedia Computer Science* 180: 649–655.

Berlec T, Govekar E, Grum J, Potočnik P and Starbek M (2008) Predicting order lead times. *Strojniški vestnik—Journal of Mechanical Engineering* 54(5): 308–321.

Bertsimas D and Kallus N (2020) From predictive to prescriptive analytics. *Management Science* 66(3): 1025–1044.

Boylan JE and Babai MZ (2022) Estimating the cumulative distribution function of lead-time demand using bootstrapping with and without replacement. *International Journal of Production Economics* 252: 108586.

Boylan JE and Syntetos A (2010) Spare parts management: a review of forecasting research and extensions. *IMA Journal of Management Mathematics* 21(3): 227–237.

Branco P, Torgo L and Ribeiro RP (2017) SMOGN: a pre-processing approach for imbalanced regression. In: *First international workshop on learning with imbalanced domains: theory and applications*, vol. 74, pp.36–50. PMLR.

Breiman L (2001) Random forests. *Machine Learning* 45(1): 5–32.

Brownlee J (2020) *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. San Francisco, CA: Machine Learning Mastery.

Brynjolfsson E and McAfee A (2017) Artificial intelligence, for real. *Harvard Business Review* 2017: 1–30.

Burggräf P, Wagner J, Koke B and Steinberg F (2020) Approaches for the prediction of lead times in an engineer to order environment: a systematic review. *IEEE Access* 8: 142434.

Cachon G and Terwiesch C (2008) *Matching Supply with Demand: An Introduction to Operations Management*. New York, NY: McGraw-Hill.

Carling K (2000) Resistant outlier rules and the non-Gaussian case. *Computational Statistics & Data Analysis* 33(3): 249–258.

Cavalieri S, Garetti M, Macchi M and Pinto R (2008) A decision-making framework for managing maintenance spare parts. *Production Planning and Control* 19(4): 379–396.

Choi S, Kim YJ, Briceno S and Mavris D (2016) Prediction of weather-induced airline delays based on machine learning algorithms. In: *IEEE/AIAA 35th digital avionics systems conference (DASC)*, Sacramento, CA, USA. IEEE.

Chopra S, Reinhardt G and Dada M (2004) The effect of lead time uncertainty on safety stocks. *Decision Sciences* 35(1): 1–24.

Dalmasso N, Pospisil T, Lee AB, Izbicki R, Freeman PE and Malz AI (2020) Conditional density estimation tools in python and R with applications to photometric redshifts and likelihood-free cosmological inference. *Astronomy and Computing* 30(1): 100362.

de Oliveira MB, Zucchi G, Lippi M, Cordeiro D, Da Silver NR and Iori M (2021) Lead time forecasting with machine learning techniques for a pharmaceutical supply chain. In: *Proceedings of the 23rd international conference on enterprise information systems*, vol. 1, pp.634–641. SCITEPRESS - Science and Technology Publications.

Dolgui A, Ammar OB, Hnaien F and Louly M (2013) A state of the art on supply planning and inventory control under lead time uncertainty. *Studies in Informatics and Control* 22(3): 255–268.

Duffie N, Bendul J and Knollmann M (2017) An analytical approach to improving due-date and lead-time dynamics in production systems. *Journal of Manufacturing Systems* 45: 273–285.

Escudero LF, Galindo E, Garcıia G, Gomez E and Sabau V (1999) Schumann, a modeling framework for supply chain management under uncertainty. *European Journal of Operational Research* 119(1): 14–34.

Friedman JH (2002) Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38(4): 367–378.

Grout JR and Christy DP (1993) An inventory model of incentives for on-time delivery in just-in-time purchasing contracts. *Naval Research Logistics* 40(6): 863–877.

Gupta H and Asha V (2020) Impact of encoding of high cardinality categorical data to solve prediction problems. *Journal of Computational and Theoretical Nanoscience* 17(9-10): 4197–4201.

Gyulai D, Pfeiffer A, Nick G, Gallina V, Sihn W and Monostori L (2018) Lead time prediction in a flow-shop environment with analytical and machine learning approaches. *International Federation of Automatic Control* 51(11): 1029–1034.

Hofleitner A, Herring R and Bayen A (2012) Arterial travel time forecast with streaming data: a hybrid approach of flow modeling and machine learning. *Transportation Research Part B: Methodological* 46: 1097–1122.

Hyndman RJ (2014) *Business Forecasting: Practical Problems and Solutions*. John Wiley & Sons.

Hyndman RJ and Athanasopoulos G (2018) *Forecasting: Principles and Practice*. OTexts.

Kuhn M and Johnson K (2019) *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Boca Raton, FL: Chapman and Hall/CRC Press.

Lawrenson J (1986) Effective spares management. *International Journal of Physical Distribution & Materials Management* 16(5): 3–111.

Li S, Ragu-Nathan B, Ragu-Nathan TS and Subba R S (2006) The impact of supply chain management practices on competitive advantage and organizational performance. *Omega* 34(2): 107–124.

Lingitz K, Gallina V, Ansari F, Gyulai D, Pfeiffer A, Sihn W and Monostori L (2018) Lead time prediction using machine learning algorithms: a case study by a semiconductor manufacturer. *Procedia CIRP* 72: 1051–1056.

Liu J, Hwang S, Yund W, Boyle LN and Banerjee AG (2018) Predicting purchase orders delivery times using regression models with dimension reduction. In: *38th computers and information in engineering conference*, vol. 1B, Quebec City, QC, Canada: American Society of Mechanical Engineers, V01BT02A034.

Min H (2010) Artificial intelligence in supply chain management: theory and applications. *International Journal of Logistics Research and Applications* 13(1): 13–39.

Muthuraman K, Seshadri S and Wu Q (2015) Inventory management with stochastic lead times. *Mathematics of Operations Research* 40(2): 302–327.

Oeztürk A, Kayalıgil S and Özdemirel N (2006) Manufacturing lead time estimation using data mining. *European Journal of Operational Research* 173(2): 683–700.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E (2011) Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12(85): 2825–2830.

Pospisil T and Lee AB (2018) RFCDE: random forests for conditional density estimation. *arXiv preprint*.

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV and Gulin A (2017) CatBoost: unbiased boosting with categorical features. *arXiv preprint*.

Quenouille MH (1949) A relation between the logarithmic, Poisson, and negative binomial series. *Biometrics* 5(2): 162–164.

Ritchie E and Kingsman BG (1985) Setting stock levels for wholesaling: performance measures and conflict of objectives between supplier and stockist. *European Journal of Operational Research* 20(1): 17–24.

Robinson LW, Bradley JR and Thomas LJ (2001) Consequences of order crossover under order-up-to inventory policies. *Manufacturing and Service Operations Management* 3: 175–188.

Rodríguez P, Bautista MA, Gonzalez J and Escalera S (2018) Beyond one-hot encoding: lower dimensional target embedding. *Image and Vision Computing* 75: 21–31.

Santner TJ, Williams BJ, Notz WI and Williams BJ (2003) *The Design and Analysis of Computer Experiments*. vol. 1, New York, NY: Springer.

Sieke MA, Seifert RW and Thonemann UW (2012) Designing service level contracts for supply chain coordination. *Production and Operations Management* 21(4): 698–714.

Silver WA, Pyke DF and Peterson R (1998) *Inventory Management and Production Planning and Scheduling*. 3rd ed. New York, NY: John Wiley & Sons.

Silver WA, Pyke DF and Thomas D (2016) *Inventory and Production Management in Supply Chains*. 4th ed. Boca Raton, FL: CRC Press.

Song JS, Zhang H, Hou Y and Wang M (2010) The effect of lead time and demand uncertainties in (r, q) inventory systems. *Operations Research* 58(1): 68–80.

Syntetos AA, Babai MZ and Altay N (2012) On the demand distributions of spare parts. *International Journal of Production Research* 50(8): 2101–2117.

Syntetos AA, Boylan JE and Disney SM (2009) Forecasting for inventory planning: a 50-year review. *The Journal of the Operational Research Society* 60: 149–160.

Tashman LJ (2000) Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16(4): 437–450.

Torgo L, Ribeiro R, Pfahringer B and Branco P (2013) Smote for regression. *Portuguese Conference on Artificial Intelligence* 8154: 378–389.

Tukey JW (1992) *Breakthroughs in Statistics: Methodology and Distribution*. New York, NY: Springer.

Turrini L and Meissner J (2019) Spare parts inventory management: new evidence from distribution fitting. *European Journal of Operational Research* 273(1): 118–130.

Urban T (2009) Establishing delivery guarantee policies. *European Journal of Operational Research* 196(3): 959–967.

Wang W (2012) A stochastic model for joint spare parts inventory and planned maintenance optimisation. *European Journal of Operational Research* 216(1): 127–139.

Webster JJ and Kit C (1992) Tokenization as the initial phase in NLP. In: *Proceedings of the 14th conference on computational linguistics*, vol. 4, pp.1106–1110. Nantes, France: Association for Computational Linguistics.

Weisberg S (2005) *Applied Linear Regression*. vol. 528, John Wiley & Sons.

Wiley M and Wiley JF (2019) *Advanced R Statistical Programming and Data Models: Analysis, Machine Learning, and Visualization*. Berkeley, CA: Apress Media LLC.

Yang B and Geunes J (2007) Inventory and lead time planning with lead-time-sensitive demand. *IIE Transactions* 39(5): 439–452.

Zhang Y, Jin R and Zhou Z (2010) Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics* 1(1): 43–52.

Zipkin PH (2000) *Foundations of Inventory Management*. McGraw-Hill.

How to cite this article

Reiners R, Haubitz CB and Thonemann UW (2025) Lead Time Prediction for Inventory Optimization With Machine Learning. *Production and Operations Management* 34(10): 3010–3025.