

Chapter 1

INTRODUCTION

1.1 Background

In today's competitive manufacturing environment, accurate demand forecasting is essential for achieving operational efficiency and ensuring customer satisfaction. This is especially true for manufacturers of heavy-duty filter products, where thousands of Stock Keeping Units (SKUs) exhibit diverse demand patterns and customer-specific order behaviors. The forecast challenge is intensified by irregular purchasing cycles, long product life spans, and the sector's exposure to external industrial volatility.

The root causes of demand forecasting failures in such environments stem from four main dimensions, as identified through a Fishbone (Ishikawa) analysis: (1) data limitations, such as missing or misaligned sales records; (2) method-related issues, including reliance on simplistic or outdated forecasting models; (3) process deficiencies, such as lack of historical tracking or integration with planning systems; and (4) environmental volatility, such as economic shocks or seasonal procurement behavior. These issues create significant disruptions in planning activities, leading to inventory inaccuracy, reactive ordering, and missed sales opportunities.

The ability to predict customer orders becomes even more critical when demand is influenced by recurring seasonal patterns, such as religious holidays or fiscal year-end purchasing cycles. Recognizing these patterns is fundamental to building robust forecasting models that support proactive planning and timely stock replenishment.

Traditional statistical forecasting methods, such as moving averages and exponential smoothing, often fail to capture the complex and nonlinear nature of real-world sales data. These methods depend on assumptions of stability and linearity, making them poorly suited to handle dynamic demand shifts, noise, or irregular behavior. In contrast, machine learning (ML)-based forecasting offers adaptive capabilities to learn hidden relationships and generate more accurate predictions.

Recent studies confirm the value of ML in supply chain forecasting. For example, Vikram Pasupuleti (2024) demonstrated that ML-based forecasts improved accuracy by 15%, reduced stock shortages by 10%, and improved on-time delivery rates. Similarly, Feizabadi

(2022) showed how hybrid ML models, such as ARIMAX combined with neural networks, reduced cash conversion cycles by better aligning forecasts with procurement cycles. These findings support the use of ML in domains with complex, repeatable but volatile demand signals, typical of the heavy-duty filter industry.

This study uses historical and recent demand data to uncover consistent behavioral patterns and develop forecasting models tailored to product-level customer orders. Although earlier sales data from 2018–2020 is used to detect major disruptions (e.g. COVID-19), the primary modeling scope focuses on the more stable 2021–2025 period. Feature engineering techniques are applied, such as moving average smoothing, lag features, and seasonal flags, to improve the predictive power of the models. Two machine learning algorithms, Random Forest and LightGBM, are implemented and compared across multiple accuracy metrics.

The goal is to provide a practical and scalable machine learning framework that improves the reliability of the forecast at the SKU level. Although the study does not extend to full inventory optimization, it provides the foundational information to integrate demand forecasts into broader supply chain planning systems.

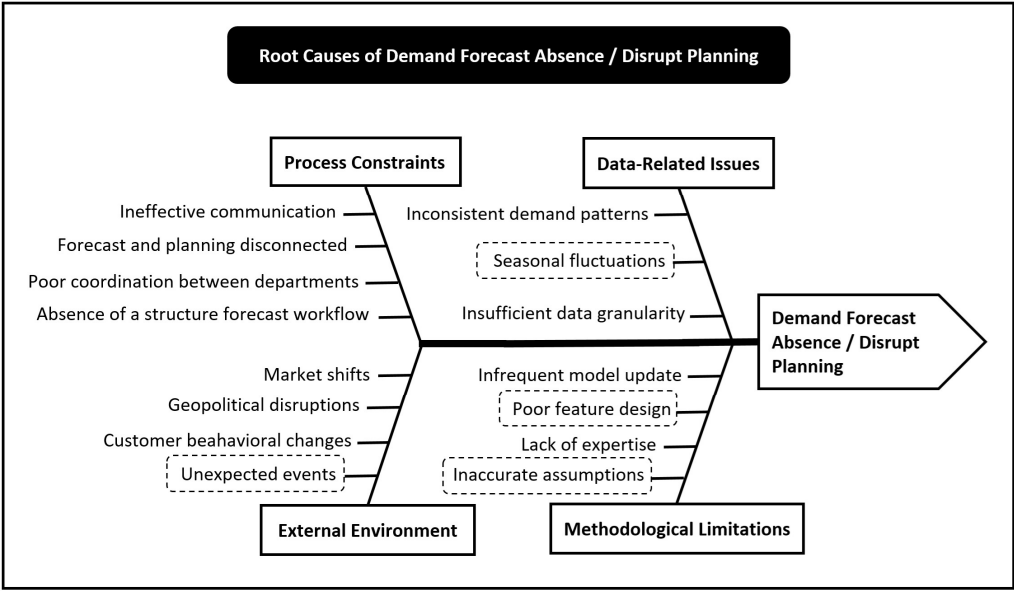


Figure 1.1: Fishbone Diagram – Root Causes of Demand Forecast Absence / Planning Disruption

To better understand the underlying challenges in demand forecasting within manufacturing settings, this study uses a Fishbone (Ishikawa) diagram (Figure 1.1) to visualize the root causes behind the absence of reliable forecasts and the resulting disruption in planning activities. The

diagram categorizes the sources of forecast failures into four domains: data-related issues, methodological limitations, process constraints, and external environmental factors.

Data-related problems include inconsistent demand patterns, insufficient historical granularity, and unaddressed seasonality, all of which distort model inputs and hinder forecast accuracy. On the methodological side, firms often rely on basic models, such as moving averages or naive extrapolation, which fail to capture nonlinear interactions, lag effects, or multidimensional feature relationships. These are further worsened by infrequent model updates and limited expertise in selecting or tuning appropriate models.

Process-related constraints arise from organizational silos, lack of coordination between departments such as sales and operations, and the absence of a structured forecasting workflow. As a result, any forecast that is produced may be disconnected from execution or decision-making processes. Meanwhile, external environmental factors, including market volatility, geopolitical changes, customer behavior changes, and one-off events such as pandemics, further undermine the stability of forecast models.

The core insight from this diagnostic view is that the inaccuracy of the forecast is not based on a single root cause but rather on the interaction of data, process, model, and external dimensions. Addressing these requires a holistic strategy that integrates data quality improvements, organizational alignment, and predictive modeling capabilities. This study proposes machine learning-based forecasting as a response to these challenges, supported by engineered features and structured time series inputs. The goal is to develop a more reliable and adaptive forecasting system that improves SKU-level accuracy and supports responsive inventory and production planning.

1.2 Research Problem

Heavy-duty filter product manufacturers operate in environments characterized by high product variety, irregular ordering behavior, and limited visibility into downstream demand. Accurate forecasting at the product code level remains a persistent challenge due to fluctuating seasonal patterns, macroeconomic influences, and the lack of granular end-consumer data.

Traditional forecasting techniques, such as moving averages and exponential smoothing, often struggle with the complexity of manufacturing sales data. These models are built on assumptions of stability, linearity, and low dimensionality, conditions that do not reflect the dynamic nature of real-world customer ordering behavior. Consequently, these methods are

insufficient to capture the temporal interactions and nonlinearities inherent in the demand at the product level.

Although machine learning (ML) has emerged as a powerful alternative, capable of learning from high-dimensional data and modeling nonlinear demand patterns, there is still limited consensus on which features and model types are most effective for forecasting customer orders using structured historical data alone. Most prior studies assume access to downstream consumption or retail-level signals, which are often unavailable in upstream manufacturing contexts.

This raises a critical research problem: *Can machine learning models accurately forecast incoming customer orders at the product level using only historical sales order data, and which engineered features are most influential in improving predictive accuracy?*

Addressing this question is vital for manufacturers seeking to improve forecast reliability and support more responsive procurement and inventory planning through data-driven methods.

1.3 Research Objectives

In response to the identified research problem, this study aims to improve the accuracy and reliability of customer order forecasting at the product level (SKU) in the heavy-duty filter manufacturing sector by applying machine learning techniques. The specific objectives are as follows.

1. Identify which time-based, seasonal, and exogenous features significantly contribute to the predictive accuracy of machine learning models when applied to structured sales order data.
2. To evaluate the impact of feature engineering and feature selection on forecast performance, particularly in addressing irregular demand patterns and enhancing model adaptability.
3. To compare the prediction performance of the Random Forest, Linear Regression, and LightGBM models across different feature scenarios, with respect to accuracy, robustness, and interpretability.
4. Benchmark the performance of machine learning-based models against traditional baseline methods such as moving averages and naive forecasts to assess the practical

added value of the proposed approach.

1.4 Research Question

To guide the research process and ensure alignment with the stated objectives and hypothesis structure, this study is framed by the following research questions:

1. **RQ1:** Which time-based, seasonal and exogenous features significantly enhance the accuracy of machine learning models for customer order forecasting?
2. **RQ2:** How does the inclusion of engineered features affect the prediction accuracy of machine learning models compared to the use of raw features alone?
3. **RQ3:** How do different machine learning algorithms, specifically Random Forest, Linear Regression, and LightGBM compare in their ability to predict incoming customer orders?
4. **RQ4:** To what extent do machine learning models outperform traditional forecasting methods in the context of product-level demand prediction?

1.5 Hypothesis

Based on the research objectives and guided by the corresponding research questions, the following hypotheses are formulated and tested in this study:

- **H1:** The inclusion of time-based, seasonal, and exogenous features significantly improves the predictive accuracy of machine learning models in the forecasting of customer orders.
- **H2:** Forecasting models developed using engineered features achieve higher accuracy than those using raw features alone.
- **H3:** Random Forest provides more accurate forecasts than LightGBM for most product categories, except in item types where demand volatility is dominant.
- **H4:** Machine learning-based models (Random Forest, Linear Regression, and LightGBM) significantly outperform traditional forecasting methods such as moving averages in terms of MAE, MAPE, RMSE, and R^2 .

1.6 Research Scope and Limitation

1.6.1 Scope of the research

This research focuses on developing machine learning-based forecasting models for incoming customer orders in a manufacturing environment that produces heavy-duty vehicle filter products. The scope and boundaries of the study are defined as follows:

1. **Industry Focus:** The study is limited to the manufacturing domain, with an emphasis on product-level forecasting for heavy-duty filter SKUs. Unlike a CRM, it focuses exclusively on predicting incoming customer orders to support upstream planning and procurement. Internal consumption, inventory control, and distribution logistics are beyond the scope.
2. **Technological Scope:** The forecasting models developed in this study are based on supervised machine learning algorithms, including Random Forest, LightGBM and Linear Regression. The primary objective is to evaluate and compare their effectiveness in capturing demand patterns using structured historical data. Technologies such as Optical Character Recognition (OCR) or other components of artificial intelligence are excluded from the scope.
3. **Forecasting Focus:** The forecasting task involves predicting monthly aggregated incoming order quantities at the SKU level. Feature engineering includes time-based variables (e.g., lags, moving averages), seasonal markers (e.g., holidays, year-end) and selected external indicators (e.g., macroeconomic indices). The study does not extend to inventory optimization or downstream replenishment strategies.
4. **Performance Evaluation:** Forecasting models are evaluated using standard statistical accuracy metrics, including mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) and R^2 . Although the results are interpreted in the context of planning effectiveness, the study does not directly measure inventory KPIs, such as stockouts or service levels.
5. **Data and Time Frame:** The dataset comprises internal sales order records spanning 2018 to 2025. Although 2018-2025 data are used to understand long-term historical patterns and disruptions, core modeling and validation activities are based on the more

stable 2021–2025 period. Real-time system integration or deployment is not included in this research.

1.6.2 Limitation of the Research

Although this study provides valuable information on the use of machine learning to forecast incoming customer orders, several limitations should be acknowledged.

1. **Sector-Specific Scope:** The research is confined to the heavy-duty filter manufacturing industry and focuses exclusively on SKU-level forecasting. It does not explore forecasting approaches in other sectors, such as retail, distribution, or discrete manufacturing with different product structures.
2. **Model and Technology Limitation:** This study evaluates only three supervised machine learning models: Random Forest, LightGBM, and Linear Regression for monthly demand forecasting. Other advanced algorithms (e.g., neural networks, deep learning) and operational technologies (e.g., IoT, RFID, blockchain) are outside the research scope.
3. **Single-Case Dataset:** All modeling and evaluation are based on historical sales order data from a single-case company, spanning 2018-2025. Although this supports depth of analysis, it limits generalizability to organizations with different demand structures or business environments.
4. **Data Quality Dependency:** Model accuracy depends on the quality and granularity of historical order data. Issues such as missing records, abnormal demand spikes, or irregular ordering cycles can affect the reliability and transferability of the models to other datasets.
5. **No Real-Time Deployment:** The study is conducted as an offline simulation and does not cover real-time integration into enterprise systems or planning workflows. Operational impacts, such as user adoption or system performance under live conditions, are beyond their scope.
6. **Exclusion of Human and Organizational Factors:** The research does not assess organizational readiness, user behavior, or planner acceptance. Although machine learning may reduce manual effort, change management and behavioral responses to forecast recommendations are not examined.

1.7 Significance of Study

This research is significant in both academic and industrial contexts, as it addresses a persistent challenge in manufacturing: the absence of accurate data-driven forecasting methods to anticipate incoming customer orders at the product level. By applying and evaluating machine learning models in a real-world manufacturing case, the study contributes valuable insights into the design and implementation of predictive forecasting systems.

1. Academic Contribution:

- The study extends the literature by evaluating and comparing machine learning algorithms (Random Forest, LightGBM) to forecast customer order volumes in a B2B industrial manufacturing setting, an area where empirical research remains limited.
- Advances academic understanding of how feature engineering and selection—including seasonal and exogenous variables—affect the accuracy of forecasting.
- The research bridges theoretical concepts such as seasonality, demand uncertainty, and temporal dynamics with measurable outcomes, such as forecast error and model interpretability.

2. Practical Contribution:

- The findings offer manufacturing planners and analysts a systematic, evidence-based approach to selecting machine learning models for demand forecasting.
- The study provides clear insights on the characteristics that drive the accuracy of the forecast, allowing organizations to improve their data preparation and forecasting processes using readily available order data.
- Although inventory optimization is not modeled directly, improved visibility into future demand supports better informed upstream planning, procurement, and production readiness.

3. Industrial Relevance:

- This research supports digital transformation in the manufacturing sector by demonstrating how machine learning can be used effectively in environments characterized by high SKU complexity and volatile demand.
- It provides a tested and replicable forecasting framework for industries where accurate demand prediction is critical, but traditional data sources (e.g., end-consumer or real-time signals) are limited.

4. Policy and Strategic Relevance:

- The study provides strategic insights for companies that seek to modernize their demand forecasting practices, particularly under the constraints of limited visibility and long lead times.
- By integrating the forecasting framework in real operational data, it offers a scalable foundation for integrating predictive analytics into strategic planning, without requiring major structural changes to existing systems.

Table 1.1: Summary of Research Contributions

Aspect	Significance
Academic	Evaluates and compares machine learning models (Random Forest, LightGBM, Linear Regression) for customer order forecasting.
Theoretical	Links demand uncertainty and seasonality to measurable improvements through feature engineering.
Practical	Provides a forecasting framework for manufacturers to improve prediction accuracy using structured features.
Industrial Impact	Supports data-driven forecasting in B2B manufacturing with high SKU complexity and limited downstream data.
Strategic Relevance	Informs forecasting strategy and analytics adoption in manufacturing environments with volatile demand patterns.

Table 1.1 summarizes the key contributions of this research in five dimensions: academic, theoretical, practical, industrial, and strategic.

From an academic perspective, the study contributes to the growing literature on machine learning applications in manufacturing by evaluating and comparing forecasting models,

specifically Random Forest and LightGBM, using real-world sales order data from a heavy-duty filter manufacturer. The research addresses a relatively underexplored area within demand forecasting at the SKU level in B2B industrial settings.

Theoretically, the study establishes a direct link between demand uncertainty, seasonal fluctuations, and the effectiveness of predictive models enhanced by structured feature engineering. It moves beyond general discussions of machine learning by empirically examining how specific time-based, seasonal, and exogenous variables influence forecasting outcomes.

On a practical level, the research provides manufacturing planners and analysts with a forecasting framework that is grounded in actual order behavior. The study offers actionable insights into which features most significantly affect model performance, thereby supporting data-driven decision-making in environments with high product complexity and fluctuating demand.

From an industrial impact point of view, the research promotes the adoption of machine learning forecasting tools in manufacturing sectors that lack downstream visibility or real-time consumption data. It demonstrates how accurate forecasts of incoming customer orders can strengthen upstream planning, even in the absence of comprehensive supply chain digitization.

Strategically, the study provides evidence to support organizations in modernizing their forecasting policies by leveraging historical order data and machine learning models. It sets the foundation for data-informed scalable planning practices that improve response to cyclical and volatile demand conditions.

1.8 Thesis Structure

The rest of this thesis is organized as follows.

1. **Chapter 1** This chapter introduces the background and motivation for improving demand forecasting in a manufacturing environment using machine learning. The study presents the research problem, the objectives, the research questions, the hypotheses, the scope, the limitations, and the significance of the study, with a particular focus on forecasting incoming customer orders for heavy-duty filter products.
2. **Chapter 2** This chapter reviews relevant literature on demand forecasting, time series modeling, and machine learning applications in manufacturing. Highlights the

limitations of traditional forecasting methods, discusses the role of feature engineering, and identifies research gaps that justify a machine learning-based approach.

3. **Chapter 3** This chapter outlines the research methodology, including data sources, preprocessing steps, feature engineering techniques, model selection and evaluation metrics. It describes how various machine learning models, such as Random Forest and LightGBM, are developed and tested using historical sales order data.
4. **Chapter 4** This chapter presents the results of the forecasting experiments, including model performance metrics and feature importance analysis. It compares the effectiveness of different models in capturing seasonal and nonseasonal demand patterns and discusses the findings in relation to the research questions.
5. **Chapter 5** This chapter summarizes the key findings and contributions of the study, acknowledges its limitations, and provides recommendations for future research. It also outlines how the results can inform forecasting strategies in manufacturing environments with volatile customer order behavior.

