Spare Parts Demand Forecasting and Inventory Management Using Machine Learning Models: A Comprehensive Application

by

Zeynep Karaca Bektaş

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Industrial Engineering

KOÇ ÜNİVERSİTESİ

January 17, 2025

Spare Parts Demand Forecasting and Inventory Management Using Machine Learning Models: A Comprehensive Application

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Zeynep Karaca Bektaş

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Metin Türkay (Advisor)

_____

Assist. Prof. Kübra Tanınmış Ersüs

_____

Assoc. Prof. Eda Yücel

Date: _____

To my beloved family,

who have stood by me through every challenge and celebrated every success

ABSTRACT

Spare Parts Demand Forecasting and Inventory Management Using Machine Learning Models: A Comprehensive Application
Zeynep Karaca Bektaş
Master of Science in Industrial Engineering
January 17, 2025

This thesis was conducted at Türkiye's first and largest integrated air conditioning factory, focusing on spare parts management. The study addresses key challenges in demand forecasting and inventory management of spare parts, with the goal of minimizing the risks of overstocking and stock shortages. A comprehensive framework from Bacchetti and Saccani (2012), merging spare parts classification, demand forecasting, inventory management, and performance evaluation, was applied. Aligning with the forecasting framework proposed by Boylan and Syntetos (2010), the study encompassed pre-processing, processing, and post-processing phases. In the pre-processing phase, demand was classified into categories such as Erratic, Intermittent, Lumpy, and Smooth, with Intermittent being the most prevalent. Demand forecasting methods, including Naïve Forecast, Holt-Winters' Seasonal Method, Croston Modification (SBA), ARIMA, MLR, MNLR, SVR, ANN, Random Tree, REP Tree, Random Forest Regressor, XGBoost, and LightGBM, were evaluated for their suitability. In the post-processing phase, total inventory cost calculations and the optimal service level ratio were determined using the Newsvendor model. Additionally, a data-driven approach employing Sample Average Approximation was utilized for optimization inspiring from Huber et al. (2019). XGBoost outperformed all other models by achieving the minimum cost while meeting the target service level. The main contribution of this study lies in incorporating demand classification as a feature in forecasting models, emphasizing the balance between service levels and costs, and demonstrating its practical significance through real-world application in spare parts management.

Keywords: Spare parts management, demand forecasting, inventory optimization, intermittent demand, newsvendor model, machine learning, demand classification

ÖZETÇE

Yüksek Lisans Tez Başlığı
Zeynep Karaca Bektaş
Endüstri Mühendisliği, Yüksek Lisans
17 January 2025

Türkiye'nin ilk ve en büyük entegre klima fabrikasında gerçekleştirilen tez çalışmasında, yedek parça yönetimine odaklanılmıştır. Stok fazlası ve stok eksiği risklerini azaltmak amacıyla yedek parça talep tahmini ve stok yönetimi konusundaki temel zorluklar ele alınmıştır. Bacchetti and Saccani (2012)'nin yedek parça talep sınıflandırma, talep tahmini, stok yönetimi ve performans değerlendirme kapsamını içeren entegre modeli ile Boylan and Syntetos (2010)'ın tahminleme öncesi, tahminleme ve tahminleme sonrası aşamalarını kapsayan modeli kullanılmıştır. Tahminleme öncesi aşamasında talepler Düzensiz (Erratic), Aralıklı (Intermittent), Yığınsal (Lumpy) ve Düzgün (Smooth) olarak dört kategoriye ayrılmış ve en yaygın olanın Aralıklı talepler olduğu sonucuna ulaşılmıştır. Naïve Tahmin, Holt-Winters'in Mevsimsel Metodu, Croston Modifikasyonu (SBA), ARIMA, Çok Değişkenli Doğrusal Regresyon (MLR), Çok Değişkenli Doğrusal Olmayan Regresyon (MNLR), Destek Vektör Regresyonu (SVR), Yapay Sinir Ağları (ANN), Random Tree, REP Tree, Random Forest Regresyonu, XGBoost ve LightGBM gibi talep tahmini yöntemlerinin uygunluğu değerlendirilmiştir. Tahminleme sonrası aşamasında, toplam stok maliyeti hesaplamaları ve optimum hizmet seviyesi oranı, Newsvendor modeli kullanılarak belirlenmiş ve optimizasyon için Huber (2019)'den alınan ilham ile Örneklem Ortalama Yaklaşımına (Sample Average Approximation) dayalı veri odaklı bir yöntem uygulanmıştır. XGBoost, hedef servis seviyesini en az maliyetle karşılayan model olmuştur. Bu tez, talep sınıflandırmasını tahmin modellerinde bir özellik olarak kullanması, hizmet seviyeleri ile maliyetler arasındaki dengeyi vurgulaması ve yedek parça yönetiminin gerçek hayattaki uygulamasını göstermesi ile literatüre katkıda bulunmuştur.

Anahtar Kelimeler: Yedek parça yönetimi, talep tahmini, stok optimizasyonu, aralıklı talep, newsvendor modeli, makine öğrenimi, talep sınıflandırması

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Prof. Dr. Metin Türkay, for the remarkable opportunities he has provided me, which have significantly contributed to my personal and academic growth. His support, guidance, and patience throughout this thesis process have been truly essential in its completion.

I am also sincerely grateful to my company, the case company for this thesis, for their constant support and contribution to my academic development. I deeply appreciate their understanding and occasional support during my academic journey over the last three years. Their collaboration and encouragement have been essential to the completion of this work.

Finally, I would like to extend my heartfelt thanks to my husband and my family for their endless support, belief in me, and patience during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
ADI	Average Demand Interval
CV^2	Coefficient of Variation Squared
SBA	Syntetos-Boylan Approximation
ARIMA	Auto-Regressive Integrated Moving Average
MLR	Multivariate Linear Regression
MNLR	Multivariate Nonlinear Regression
SVR	Support Vector Regression
ANN	Artificial Neural Networks
REP Tree	Reduced Error Pruning Tree
Random Tree	Decision Tree-Based Model
Random Forest Regressor	Ensemble Learning Method
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine

Chapter 1

INTRODUCTION AND PROBLEM DESCRIPTION

1.1 Background of the Research

The service parts industry for products such as household appliances, automobiles, and copy machines has grown into a business worth more than 200 billion USD worldwide (Gallagher et al., 2005, as cited in Bacchetti and Saccani, 2012). Efficient spare parts inventory management is crucial for many companies, particularly those that stock a diverse array of parts, as they face challenges in balancing availability while minimizing inventory holding costs (Boylan and Syntetos, 2010). The challenge lies in ensuring a sufficient supply of specific items to fulfill anticipated demand patterns, while simultaneously achieving a prudent balance between inventory holding costs and the potential cost like loss of the sales and goodwill in case of stock-outs (Kannan et al., 2020). The trade-off between investing in inventories and experiencing shortages can be understood as balancing the costs related to inventory and the level of service achieved (der Auweraer et al., 2019).

Spare parts are typically associated with demand patterns that are erratic, lumpy, or intermittent, featuring extended no demand periods and conventional time-series techniques often struggle to accurately estimate these demand trends due to their emphasis on the most recent data points (Pinçe et al., 2021). Altay et al. (2012) highlighted that spare parts typically display intermittent demand, characterized by fluctuations in both the demand volume and the timing of demand. This complexity has drawn significant academic attention to forecasting and inventory control of intermittent demand.

1.2 Business Context of the Case Company

The company where the thesis was performed has been manufacturing residential and commercial air conditioners since 1999. It is the first and the largest integrated air conditioning factory of Türkiye, that serves more than 60 countries.

The company deeply committed not only to delivering top-tier products but also to providing enduring relationships with its customers through after-sales support. Beyond providing air conditioning solutions, the company prioritizes enhancing customer satisfaction, recognizing that service quality and brand perception are crucial for maintaining market leadership and brand loyalty. The company has set the department's target service level at 95% to ensure customer satisfaction.

In ensuring comprehensive customer satisfaction and upholding its commitment to service excellence, the company is responsible for providing spare parts for its products for 10 years after installation. Forecasting spare part demands and implementing a robust inventory policy are crucial elements in fulfilling this commitment effectively.

Between 2013-2024, the company has produced an extensive range of 604 product SKUs, accompanied by a substantial 8,517 spare part codes in their service bills of materials (SBOMs). However, this large number of spare parts, combined with volatile demand, presents a significant challenge. Currently, the company's inventory policy relies on a combination of customer forecasts and historical data. Utilizing these sources, the new inventories to be ordered are manually calculated and determined. New orders are placed monthly with a fixed review period, and this format is preferred to be maintained for its operational and managerial efficiency.

1.3 Research problems - Motivation - Purpose

Effective forecasting and inventory management becomes crucial to managing the complexity due to extensive range of material SKUs and unpredictable demand. The thesis addresses the challenge of forecasting in spare part management for manufacturing facilities. The shortage of spare parts causes customer dissatisfaction and free of charge replacement of the main product to prevent negative brand percep-

tion. However, maintaining excessive spare part inventory as a preventive measure is impractical due to warehouse constraints and financial costs, including the high holding costs. Therefore, the core issue lies in maintaining an optimum inventory level to ensure sufficient spare parts availability for meeting customer needs, while simultaneously minimizing inventory holding and replacement costs.

The author of this study currently manages the spare parts planning team of the company and will consider well-crafted demand forecasting and inventory management strategy to reduces risks of overstocking and stock-outs. To achieve the aim of the thesis, the questions listed below will be addressed:

1. Which demand forecasting methods are most suitable for the case company? Are there different demand patterns? If so, can these demand patterns be useful for forecasting models?
2. What is the most appropriate inventory strategy for the case company?
3. Additionally, Muniz et al. (2021) highlighted critical questions from previous studies in spare parts management. The questions, including 'How many spare parts should be ordered? When should they be (re)ordered? And how many items should be (re)ordered?' will be addressed in this study. Only the fixed reorder period will continue to be kept as 1 month, as required.

1.4 Scope and Outline of the Research

In the quantitative approach, the upcoming value of the variable is determined by mathematical calculations considering past data (Gokcel, 2009, as cited in İfraz et al., 2023). To address the previously mentioned inquiries, this thesis is structured to integrate quantitative research within the framework of a case study. Specifically, the thesis will center on a case company, with all datasets being collected from. Utilizing quantitative research methodologies, the study will examine business problems of the case company and develop targeted solutions through empirical data analysis.

The case company is a joint venture that comprises two primary entities. For the simplicity, the research is restricted to the products and spare parts of one of the parent companies, accounting for approximately 95% of the total spare part

shipments. The primary limitation of the research lies in the reliability and completeness of the datasets. Thus, the most trustworthy datasets available are used for the research. For instance, it is observed that shipment data exhibits greater accuracy compared to previous demand data. Therefore, shipment data is utilized as a proxy for demand, given the typically short lag between demand and shipment. The period under consideration for the research covers the years from 2013 to 2024.

Bacchetti and Saccani (2012) claimed that the literature lacks approaches that integrate spare parts classification, demand forecasting, inventory management models, and performance measurement. Their study suggested that future research should focus on developing comprehensive frameworks that encompass all these aspects.

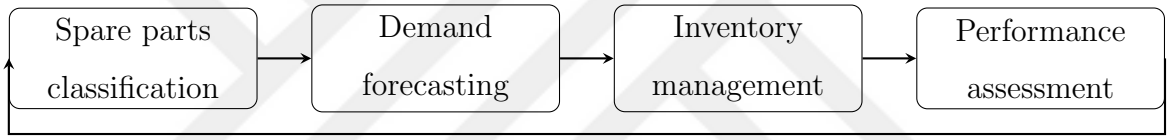

Figure 1.1: An integrated approach to spare parts management (Bacchetti and Saccani, 2012)

This thesis comprises six sections, offering a comprehensive approach to spare parts management, inspired by the framework proposed by Bacchetti and Saccani (2012). The first section encompasses the introduction and problem description, including the background of the research, business context of the case company, research problems, motivation, purpose, limitations, and outline of the study. Section 2 explores recent literature on spare part management, including demand classification, demand forecasting and inventory management. The methodologies for demand classification and forecasting models are covered in Section 3, while those for inventory management models are presented in Section 4. The implementation, results, and discussions are demonstrated in Section 5. The final section concludes the thesis and offers suggestions for the future researches.

Chapter 2

LITERATURE SURVEY

In this section, a detailed review of current literature and methodologies on demand forecasting and inventory management for spare parts will be provided.

In inventory management, the necessity for precise forecasting continues to rise. Accurate forecasting in advance can effectively reduce both overstocking and out-of-stock scenarios, consequently minimizing associated costs. However, the distinct demand patterns of spare parts cause challenging demand forecasting particularly in this domain (İfraz et al., 2023). This section will first review the literature on demand classification methods to clarify demand characteristics, then provide an review of demand forecasting models for intermittent demands, and conclude with a review of inventory management literature.

2.1 Demand Classification

The unique characteristic of spare parts is found in their demand pattern, setting them apart from numerous other products (der Auweraer et al., 2019). While certain spare parts may exhibit consistent or high demand, the majority experience intermittent demand, featuring alternating intervals with no demand and occasional non-zero demands. Furthermore, demand sizes can vary significantly, leading to erratic behavior (der Auweraer et al., 2019).

Quantitative classification approaches leverage historical data to establish relationships among relevant variables, aiding decision-making processes (Pinçe et al., 2021). Williams et al. (1984, as cited in Pinçe et al., 2021) conducted one of the pioneering studies on this topic, where demand is classified into sporadic, slow-moving, or smooth categories by analyzing the causal factors behind lead-time demand variance. The Croston method (as cited in Boylan et al., 2008) was specifically

designed for intermittent demand by focusing on the mean inter-demand intervals. Building on Croston, Johnston and Boylan's interpretation (1996, as cited in Boylan et al., 2008) introduced a new perspective on 'intermittence,' identifying mean inter-demand intervals as a categorization parameter. In their study, Syntetos (2005, as cited in Boylan et al., 2008) formulated a demand categorization scheme grounded in theory, based on critical metrics: the average interval between demands (ADI) and the squared coefficient of variation in demand sizes (CV^2). The scheme categorizes demand as erratic, smooth, lumpy and intermittent (Boylan et al., 2008). Since the framework of Syntetos can be easily understood and applied in practical settings, it continues to be utilized in newest studies like Chien et al. (2023).

2.2 Demand Forecasting Methods

Qualitative and quantitative approaches are the two main categories of demand forecasting methods (Pinçe et al., 2021). Qualitative forecasts are created subjectively by specialists, relying on non-mathematical techniques, while quantitative approaches predict future values through mathematical calculations with historical data (İfraz et al., 2023).

Time series methods emphasize analyzing past event data to construct statistical models derived from trends, and applying these models to anticipate future results (İfraz et al., 2023). A substantial body of literature focuses on time-series methods, widely used and practical for estimating spare parts demand (Pinçe et al., 2021). Naïve forecasting, one of the simplest demand forecasting techniques, estimates future values based on the most recent period of previous demand (Chien et al., 2023). Syntetos-Boylan Approximation, Holt-Winters' seasonal method, and ARIMA are commonly used time series models for forecasting of intermittent demand (Chien et al., 2023).

Holt-Winters' seasonal method models data using a local mean, a local trend, and a local seasonal factor, and each of these components is updated through exponential smoothing (Chatfield and Yar, 1988). Exponential smoothing demonstrated strong robustness as a forecasting method. The model responses rapidly to shifts in

the demand process and is commonly utilized. (Romeijnders et al., 2012). However, Croston (1972, as cited in Romeijnders et al., 2012) demonstrated that traditional forecasting models like exponential smoothing and moving averages, are less effective to predicting intermittent demand and suggested separately updating the demand volume and interval employing exponential smoothing, during only positive demand. Wright (1986, as cited in Altay et al., 2008) modified Holt's double exponential smoothing to handle data with changing reporting frequencies and irregular time intervals and suggested estimating intercepts and slopes using single exponential smoothing. This modification, unlike Croston's stationary mean assumption, offers greater flexibility and potential performance improvements (Altay et al., 2008). Croston's method is criticized by Snyder (2002, as cited in Boylan and Syntetos, 2010) for being inconsistent, and by Shenstone and Hyndman (2005, as cited in Boylan and Syntetos, 2010) for casting doubt on the existence of any stochastic demand model for which it is optimal.

Croston (1972, as cited in Bacchetti and Saccani, 2012) conducted a seminal study in intermittent demand forecasting and introduced a method that became a foundation for subsequent research. Croston's method (1972, as cited in Pinçe et al., 2021) has undergone numerous modifications by scholars over the years, and it has become a crucial standard model for assessing methods used in forecasting spare parts demand. Schultz (1987, as cited in Xu et al., 2012) introduced a forecasting procedure primarily derived from Croston's method and proposed a novel inventory policy that incorporates the delays. Notably, Schultz (1987, as cited in Xu et al., 2012) suggested using separate smoothing parameters for demand sizes and demand intervals, differing from Croston's original approach, that assumed a single smoothing parameter. By incorporating bootstrapping techniques, Snyder (2002, as cited in Teunter and Duncan, 2009) introduced more complex variations of the Croston method. Similarly, Willemain et al. (2004, as cited in Teunter and Duncan, 2009) and Porras Musalem (2005, as cited in Teunter and Duncan, 2009) proposed bootstrapping approaches. The bootstrapping approach of Willemain et al. (2004, as cited in Zhou and Viswanathan, 2011) determines safety stock by generating periods of positive demand utilizing a two-state Markov model and simulating observed

demand volume based on historical data. Willemain et al. (2004, as cited in Zhou and Viswanathan, 2011) indicate that their method outperforms both exponential smoothing and Croston's method in forecasting demand distribution over a fixed lead time. Viswanathan and Zhou (2008, as cited in Zhou and Viswanathan, 2011) introduced an enhanced bootstrapping method, demonstrating its superiority over Willemain's model. The key innovation of their method lies in generating positive demand arrivals based on the historical distribution of inter-demand intervals. Hasni et al. (2018) assessed bootstrapping methods in spare parts management, questioning their suitability for intermittent demand forecasting. The study emphasizes that, although the "distribution-free" nature of bootstrapping is valuable, it is not sufficient on its own to ensure accurate forecasts. They highlighted the need for advanced approaches specifically tailored to tackle the unique complexities of intermittent demand in spare parts management (Hasni et al., 2018).

The bias of Croston's estimator was demonstrated by Syntetos and Boylan (2001, as cited in Boylan and Syntetos, 2010), who proposed a bias-adjusted method known as the Syntetos–Boylan Approximation (2005, as cited in Boylan and Syntetos, 2010). Additionally, Boylan and Syntetos (2003, as cited in Boylan and Syntetos, 2010) and Shale et al. (2006, as cited in Boylan and Syntetos, 2010) discussed adjustment parameters to address the bias of Croston's approach. Teunter and Duncan (2009) conducted a comparative study of forecasting approaches for intermittent demand analysing a wide dataset of UK Royal Air Force. They recommended the implementation of the standard Croston method, the Syntetos–Boylan Approximation or Bootstrapping, incorporating a modified calculation of lead-time demand and order-up-to levels with Lognormal demand (Teunter and Duncan, 2009). While bootstrapping demonstrated comparable performance, it was noted to be more challenging to implement.

The Autoregressive Integrated Moving Average (ARIMA) model, introduced by Box et al. (1970, as cited in Chien et al., 2023), is derived from autoregressive models, moving average models, and their combination, making it a widely used approach for time series forecasting. ARIMA demonstrated effectiveness as a forecasting method in various applications through the analysis of past data (Chien

et al., 2023). ARIMA was one of the models used by Vargas and Cortes (2017, as cited in İfraz et al., 2023) to forecast monthly spare part demands in the automobile industry in Mexico, by Boukhtouta and Jentsch (2018, as cited in İfraz et al., 2023) to compare the potential of SVM in anticipating the demand for spare parts belonging to the Canadian Armed Forces, and by Menezes et al. (2015, as cited in İfraz et al., 2023) as a benchmark model to evaluate the performance of feed-forward and recurrent neural networks in demand forecasting of spare parts.

Time series methods (weighted moving average, exponential smoothing, and Holt models) are considered basic approaches to demand forecasting, since they do not account for external variables influencing demand. In contrast, regression analysis, Box-Jenkins models, and artificial intelligence-based methods are more effective in capturing relationships within historical data (Aktepe et al., 2021). Alalawin and Arabiyat (2021) used real-world data to project the demand and price of spare parts for hybrid vehicles, employing a multiple linear regression (MLR) model for analyzing the relationship between demand and price and validating the results through a case study. Aktepe et al. (2021) forecasted the demand for construction machinery spare parts by applying multiple linear regression, multiple nonlinear regression, artificial neural networks, and support vector regression models in a factory. Aryudha and Hasibuan (2024) highlighted that spare part distributors, such as GMT, face challenges in managing fluctuating inventories and require algorithms to predict spare part demand accurately and used linear regression to predict spare part demand, capturing the connection between independent and dependent variables.

Wang et al. (2024) highlighted that machine learning approaches like support vector machines and neural networks have been applied to intermittent demand forecasting. Hua and Zhang (2006, as cited in Bacchetti and Saccani, 2012) proposed Support Vector Machines (SVM) regression to forecasting spare parts demand. Huang et al. (2010, as cited in Aktepe et al., 2021) introduced a method that integrates grey relational analysis with Support Vector Machines (SVM) to improve the estimation of spare parts usage. This approach is particularly valuable when accurate quantification of influencing factors is difficult. Riwayadi et al. (2024) evaluated the effectiveness of Support Vector Regression (SVR) in forecasting spare parts of

coolers and freezers in the service and maintenance industry. achieved reliable results through grid search optimization for parameter tuning.

Gutierrez et al. (2008, as cited in Kourentzes, 2013) suggested a neural network (NN) method for estimating lumpy demand time series and it outperformed both Croston's method and the Syntetos-Boylan Approximation in their experiments, though it has significant limitations. Nasiri Pour et al. (2008, as cited in Zhuang et al., 2022) proposed a model combining a neural network with four input variables to predict demand occurrence, along with exponential smoothing for forecasting non-zero demand levels. Kourentzes (2013, as cited in Jiang et al., 2019) suggested a model for forecasting sporadic demand using two constituent elements, similar to Croston's approach, but replaced simple exponential smoothing with a neural network (NN) method to update demand size and demand interval. NN method achieves a higher service level than the SBA, even though its forecasting accuracy is lower and does not require an increase in stockholding volume (Jiang et al., 2019). Following Kourentzes, researchers extensively explored segmented forecasting by separately predicting demand occurrence and demand quantity (Zhuang et al., 2022). Menezes et al. (2015, as cited in İfraz et al., 2023) evaluated the performance of feed-forward and recurrent neural networks in forecasting spare part demand. They compared their results with the Croston and ARIMA methods, and artificial neural networks (ANN) outperformed these statistical methods in three out of four datasets (İfraz et al., 2023). Lolli et al. (2017, as cited in Tian et al., 2021) employed feed-forward single-hidden layer neural networks for intermittent demand forecasting and analyzed various aggregation levels. Amirkolaii et al. (2017) applied neural networks (NN) to project aircraft spare parts demand and improved forecasting accuracy for intermittent demand by utilizing aggregated demand features. Their approach outperformed traditional models like Croston modifications, moving average, and single exponential smoothing (Amirkolaii et al., 2017). Vargas and Cortés (2017) assessed different forecasting methods, involving moving averages, exponential smoothing, ARIMA, ANN, and hybrid ARIMA-ANN models, for estimating monthly demand for automobile spare parts in Mexico. Even though none of the applied models consistently outperformed in all cases, artificial neural networks

(ANNs) demonstrated strong performance while ARIMA models exhibited greater stability in post-sample periods. These findings underscore the need to enhance forecasting methodologies for the industry to facilitate more effective planning and decision-making.

İfraz et al. (2023) performed a case study on spare part demand forecasting, applying various methods, including regression-based approaches (MLR, MNLR, Gaussian process regression, additive regression, regression by discretion, SVR), rule-based methods (decision table, M5Rule), tree-based methods (random forest, M5P, Random tree, REPTree), and artificial neural networks (ANN). Among these models, SVR, DT, and M5P performed best within their respective categories, while ANN provided the most accurate forecasts overall.

Random forest, a machine learning algorithm that uses bagging to combine the outputs of multiple decision trees and reduce overfitting (Breiman, 2001, as cited in Chien et al., 2023), was applied by Choi and Suh (2020 as cited in Chien et al., 2023) to forecast military aircraft spare part demand in South Korea. Their study demonstrated that random forest outperformed other decision tree algorithms in reducing operation and maintenance costs (Chien et al., 2023). Yoon and Kim (2015, as cited in İfraz et al., 2023) carried out a study on demand forecasting to improve the availability of spare parts for ships in the Korean Navy, utilizing Linear Regression (LR), Random Tree (RT), Random Forest (RF), and Artificial Neural Network (ANN) methods. Their evaluation revealed that RF provided the most accurate results (İfraz et al., 2023). Kannan et al. (2020) performed a comparative study with spare parts demand dataset of a producer. Machine learning models, such as XGBoost and Random Forest, were found to provide more accurate predictions and generally outperformed ARIMA in their results. In evaluating the overall performance against benchmark models (Naïve and Zero Forecast), they observed that the benchmark methods outperformed both the advanced time series models and machine learning approaches. (Kannan et al., 2020).

Gradient Boosting Decision Trees (GBDT), developed by Friedman (2001, as cited in Zhuang et al., 2022), represent a foundational boosting-class algorithm. Chen and Guestrin (2016, as cited in Zhuang et al., 2022) developed the XGBoost

algorithm to address limitations in handling large data volumes, complex features, and high response levels in industrial applications. Pawar and Tiple (2019, as cited in Chien et al., 2023) applied multiple machine learning techniques to forecast Anti-Aircraft missile spare parts demand. They demonstrated that XGBoost and Multi-layer Perceptron achieved higher forecasting accuracy compared to other conventional methods. However, XGBoost required careful hyperparameter tuning and significant computational resources for large datasets (Chien et al., 2023). LightGBM, an open-source algorithm framework developed by Microsoft, builds upon XGBoost with optimizations in three main aspects: the Histogram Algorithm, Gradient-based One-Side Sampling, and Exclusive Feature Bundling (Zhuang et al., 2022). Zhuang et al. (2022) introduced the Intermittent Demand Combination Forecasting method, which integrates internal and external data to improve spare parts demand forecasting. They used machine learning-based feature engineering, with classification models for predicting demand occurrence and regression models to estimate non-zero demand quantities. This method, with LightGBM as the base model, demonstrated significant improvements in forecasting accuracy of real-world automotive after-sales data (Zhuang et al., 2022). Chien et al. (2023) developed a stacking ensemble approach to improve forecasting accuracy by classifying demand patterns and constructing corresponding models. Baseline models, including Holt-Winters' seasonal method, Syntetos-Boylan Approximation, ARIMA, and XGBoost, were combined using a linear regression meta-learning model. An empirical study conducted at a leading aftermarket component manufacturer in Taiwan. Chien et al. (2023) demonstrated that the stacking ensemble outperformed benchmark models across various demand patterns, reducing forecast errors and total costs.

Wang et al. (2024) introduced an innovative method for combining probabilistic forecasts to address intermittent demand, emphasizing a holistic view of not only forecasting accuracy but also inventory efficiency. The study utilized several individual forecasting models, including Quantile Regression, Poisson, Negative Binomial, the Bootstrap Method of Willemain, the Bootstrap Method of Zhou and Viswanathan, LightGBM, LSTM (Long Short-Term Memory), ARIMA, and Exponential Smoothing (Wang et al., 2024).

2.3 Inventory Management

For service and maintenance providers, a well-organized spare parts inventory management system is crucial. Keeping a sufficient stock of spare parts enables companies to achieve their operational targets by providing timely shipment to customers (Atakay et al., 2022).

The two main and most commonly used inventory policies are continuous review and periodic review. A continuous review policy involves constant checking of spare parts stock levels, with two common approaches: the (s, S) policy, where inventory is replenished to the order-up-to level (S) once it falls below the reorder point (s), and the (Q, R) policy, where a fixed order size (Q) is issued whenever stock reaches to the reorder point (R) (Zhang et al., 2021). The (R, S) policy is one of the most common periodic review policies. Orders are issued at the beginning of each fixed ordering cycle (R) to bring the stock level to determined level (S) (Zhang et al., 2021).

Roberts (1962, as cited in Sani and Kingsman, 1997) used renewal theory to approximate the optimal values of (s) and (S) for the (s, S) policy and defined ordering cost and backlogged cost as limits. Then, Naddor (1975, as cited in Sani and Kingsman, 1997) proposed heuristic methods to approximate solutions for the (s, S) policy, needing only the demand's mean and variance. After that, Ehrhardt (1979, as cited in Sani and Kingsman, 1997) developed a new (s, S) system called the power approximation, based on the approximation methods from Roberts' paper, requiring only the mean and variance of demand.

Sani and Kingsman (1997) compared ten periodic inventory policies, including the standard (R, S) policy, using real-world data of vehicles and agricultural machinery. They highlighted that (s, S) inventory control system is considered the theoretically optimal approach for managing items with low and intermittent demand. Additionally, they showed that low and intermittent demand practitioners can use all of the inventory policies confidently (Sani and Kingsman, 1997). Syntetos et al. (2015) stated that the periodic order-up-to-level inventory control system is commonly employed in real-world and preferred for its simplicity, as it requires

optimizing only the order-up-to-level parameter. In this study, the (R, S) inventory policy will be applied, as it is the most relevant inventory policy for managing current real-world operations of the company.

The newsvendor problem is a fundamental and widely utilized inventory model, commonly employed in various fields to address inventory management challenges (Oroojlooyjadid et al., 2020). The model seeks to balance overage costs, incurred from unsold goods, and underage costs, resulting from stockouts and lost sales. The goal of the Newsvendor model is to reduce the total cost by identifying the optimal order quantity (Oroojlooyjadid et al., 2020). Natarajan et al. (2017) performed computational experiments on automotive spare parts, highlighting the asymmetry and ambiguity of the newsvendor model, which assumes a known probability distribution for random demand—an assumption often unrealistic due to limited or unreliable distribution information. Natarajan et al. (2017) introduced an innovative information set for the distributionally robust multi-item newsvendor problem, employing second-order partitioned statistics to address asymmetry. Huber et al. (2019) proposed a framework including demand estimation, inventory optimization, and integrated estimation and optimization. They evaluated the resulting costs associated with the newsvendor decision (Huber et al., 2019). In the same paper, Huber et al. (2019) also introduced data-driven approaches to the newsvendor problem, utilizing machine learning and quantile regression to solve it without a specific demand distribution.

Syntetos and Boylan (2006, as cited in Syntetos et al., 2015) and Syntetos et al. (2010, as cited in Syntetos et al., 2015) emphasized the weak relationship between forecast accuracy metrics, such as mean error, and inventory management metrics, including inventory investment and customer service. Syntetos et al. (2010, as cited in Snyder et al., 2012) recommend focusing on stock control metrics rather than relying solely on point forecasts when making inventory decisions. Gardner (1990, 2006, as cited in Syntetos et al., 2015) recommended the use of tradeoff curves to evaluate the balance between inventory investment and customer service. Additionally, Sani and Kingsman (1997) suggested assessing performance through average regret metrics, offering a complementary perspective on improving inven-

tory management practices. Eaves and Kingsman (2004) compared the forecasting performance of Syntetos-Boylan Approximation (SBA) with Exponential Smoothing (ES), Croston's method, a 12-month moving average, and a simple previous year average method. Leveraging comprehensive demand and renewal lead-time data from the UK's Royal Air Force spare parts, Eaves and Kingsman (2004) introduced an alternative forecasting measure based on Wemmerlöv's methodology. This measure evaluated forecasting performance through implied stock-holdings and calculated the exact safety margin necessary to achieve zero stock-outs. While the method provided a key performance metric in terms of differences in stock-holding requirements with monetary implications, it did not offer specific guidance on the stock levels to maintain (Eaves and Kingsman, 2004).

Altay et al. (2008) compared the Modified Croston's method Syntetos–Boylan Approximation (SBA) and Wright's method, an approach is based on Holt's double exponential smoothing. The study utilized an (s, S) inventory policy and Naddor's heuristic. They assessed the effectiveness of the forecasting methods by implementing three forecast accuracy metrics and four accuracy implication metrics including average inventory, proportion of orders immediately fulfilled from stock, average percentage of orders satisfied, and total cost (Altay et al., 2008). The results indicated that the Syntetos–Boylan Approximation (SBA) is preferable for minimizing inventory levels, while the Modified Holt's method is superior for achieving high customer service levels (Altay et al., 2008).

Syntetos and Boylan (2010) emphasized the equal importance of forecast accuracy and variability (sampling error of the mean) in achieving required service levels and minimizing inventory costs. Their study analyzed four prominent intermittent demand estimation methods in terms of their variance and bias properties. Analytical expressions for the sampling error of the mean and bias were presented, enabling the calculation of the Mean Squared Error (MSE) for these estimators. The findings emphasize the importance of evaluating both the accuracy and variability of forecasts to fully understand their impact on stock control.

Zhou and Viswanathan (2011) compared a bootstrapping method with the parametric approaches of Babai and Syntetos (2007, as cited in Zhou and Viswanathan,

2011), who used the Syntetos–Boylan Approximation (SBA) to forecast the mean of the lead time demand and the corrected Mean Squared Error (MSE) for its variance. According to Babai and Syntetos (2007, as cited in Zhou and Viswanathan, 2011) the optimum order-up-to level S_t at period t is the lowest S_t which meets the condition where the cumulative probability of meeting the demand, calculated using the probability density function of total lead time demand, is greater than or equal to the target cycle service level. Zhou and Viswanathan (2011) measured the performance considering the total inventory-related cost, including holding and penalty costs, under a 95% service level.

Altay et al. (2012) examined the impacts of different types of correlation, including auto-correlated demand size, auto-correlated inter-arrival time, and cross-correlated interval and size integration, on forecasting and stock control for intermittent demand items. Employing the Syntetos–Boylan Approximation to estimate the average and variance of the demand, along with the Power Approximation to calculate the re-order point and order-up-to levels in a periodic replenishment policy, it has been demonstrated that correlation in intermittent demand significantly impacts forecast accuracy and inventory control efficiency.

Kourentzes (2013) explored the application of neural networks for forecasting intermittent time series and compared them to established methods, such as simple moving averages, exponential smoothing, Croston's method and its modifications. The study used a commonly applied order-up-to policy with a monthly review time and an order-up-to level S, calculated using demand over the lead time, forecast error variance, and a safety factor derived from the normal distribution for achieving the desired service level. The performance was evaluated using forecasting accuracy measures, including Mean Error and Mean Absolute Error, as well as inventory indicators such as service levels and holding stock. The findings revealed a conflict between forecasting accuracy measures and inventory indicators. While NNs showed poor performance in accuracy, it excelled in inventory simulations. Simulation outcomes, derived from actual service levels and trade-off curves, indicated that NN models outperformed Croston's method and its modifications, demonstrating their effectiveness for intermittent demand applications (Kourentzes, 2013).

do Rego and de Mesquita (2014) conducted research focused on inventory management for automotive spare parts. They simulated 17 combined policies that integrate various time buckets, demand forecasting models including Simple Moving Average (SMA), Syntetos–Boylan Approximation (SBA), and Bootstrapping—and demand distribution models over lead time. These combined inventory policies were implemented under a continuous review system with multiples of a fixed-order quantity. do Rego and de Mesquita (2014) evaluated the performance of each policy using total costs and fill rate as metrics, ultimately recommending the best-performing policies for practical application.

Hahn and Leucht (2015) compared the generalized hurdle Negative Binomial model against existing models, such as the hurdle Poisson and the Negative Binomial, for inventory control under intermittent demand. They evaluated their suitability and accuracy based on inventory metrics, including target service level and actual stock. Hahn and Leucht (2015) also utilized the robust worst-case approach by determining the expected service level to assess the risk of decreases in service level when demand data quality was inadequate. They determined the optimal order-up-to level by minimizing the stock level while ensuring that the expected shortage remains within a specific proportion of the expected demand per period.

Syntetos et al. (2015) performed a holistic comparison of parametric and bootstrapping methods for predicting intermittent demand, aiming to evaluate their effects on inventory control performance. The study assessed three key performance measures: total inventory investment, achieved cycle service level, and total backorders. Inventory investment was calculated by summing the unit costs of all SKUs, achieved cycle service level reflected the percentage of replenishment cycles where demand was met directly from stock, and backorders were averaged over time for each SKU. The findings revealed that simple parametric methods, such as exponential smoothing, performed well, and raised questions about the added complexity of bootstrapping methods and their necessity for inventory management (Syntetos et al., 2015).

Amirkolaii et al. (2017) utilized neural networks (NN) and Mean Square Error (MSE) to model and analyze the aircraft spare parts. The results were compared

with well-known forecasting methods for intermittent demand, like Croston and its modifications, as well as traditional approaches like moving average and single exponential smoothing (Amirkolaii et al., 2017). Their findings suggest that neural networks incorporating a larger number of features significantly enhance the accuracy of demand forecasts while mitigating related financial impacts in intermittent demands. The study also computed positive and negative errors to quantify forecast accuracy, while positive errors (overstocking) increase inventory costs, negative errors (understocking) critically impact customer satisfaction (Amirkolaii et al., 2017).

Hasni et al. (2018) explored the use of bootstrapping in inventory review systems for predicting critical parameters, such as reorder points or order-up-to levels, subject to service level condition. Service level is the most widely used measure in bootstrapping literature. Both service level and fill rate are key measures, with fill rate being particularly relevant for assessing inventory performance from a customer perspective, despite receiving less focus. (Hasni et al., 2018). In the same study, Hasni et al. (2018) also defined S as the smallest value at ensuring that the empirical distribution function provides or exceeds the service level threshold under the order-up-to-level policy and presented mathematical expression of fill rate for both periodic and continuous review systems.

Wang et al. (2024), in their probabilistic intermittent demand forecasting study, proposed an inventory-cost-based optimal weighting approach under a newsvendor framework to enhance inventory management. This method calculates optimal weights for probabilistic forecasts by minimizing a cost function that balances inventory holding costs and lost sales costs. The combined forecasts are generated as a finite mixture of individual probabilistic forecasts, ensuring they maintain the properties of probability distributions. Wang et al. (2024) defined the target customer service levels using the newsvendor framework, with the service levels calculated based on the cost ratio of inventory holding and lost sales. They highlighted that inventory evaluation should incorporate both forecasting performance and economic impacts.

Chapter 3

DEMAND FORECASTING METHODS FOR SPARE PARTS

Boylan and Syntetos (2010) presented a valuable framework for categorizing contributions across three key phases of forecasting: pre-processing, which involves rules for classifying spare parts according to average time between demands and the coefficient of variation of demand sizes; processing, which focuses on applying the appropriate forecasting method; and post-processing, which includes adjustments made by the forecast user. In this study, demand classification will serve as the pre-processing phase, demand forecasting models will be explored during the processing phase, and inventory management will be implemented in the post-processing phase.

Figure 3.1: Phases of forecasting (Boylan and Syntetos, 2010)

3.1 Demand Classification

To select appropriate forecasting models for the case dataset, the demand classification method of Syntetos (2005, as cited in Chien et al., 2023) will be applied in this study as part of the pre-processing phase of demand forecasting. The average inter-demand interval (ADI) and the coefficient of variation of demand (CV^2) constitute the two parameters of the two-dimensional matrix (Chien et al., 2023). The Average Demand Interval (ADI) represents the average time between two sequantial

demands for a spare part, reflecting the intermittency of demand as outlined:

$$ADI = \frac{T}{N} \tag{3.1}$$

In equation 3.1, N denotes the number of periods with non-zero demand while T is defined as the total period of demand (Chien et al., 2023). The coefficient of variation of demand, CV^2, standardizes demand variability by relating the variance ratio of demand data to the square of the mean demand, as expressed in Equation 3.2:

$$CV^2 = \frac{\sum_{t=1}^{T}(X_t - u)^2/T}{u^2} \tag{3.2}$$

Here, X_t denotes the size of the demand at time t, u is defined as the mean value of the demand, and T represents the total period of demand (Chien et al., 2023). Based on the demand-driven categorization framework suggested by Boylan et al. (2008), developed following the work of Syntetos (2005, as cited in Boylan et al., 2008), the recommended threshold value for ADI is 1.23, and for CV^2 is 0.49.

Demand categorization of Boylan et al. (2008) (after Syntetos (2005, as cited in

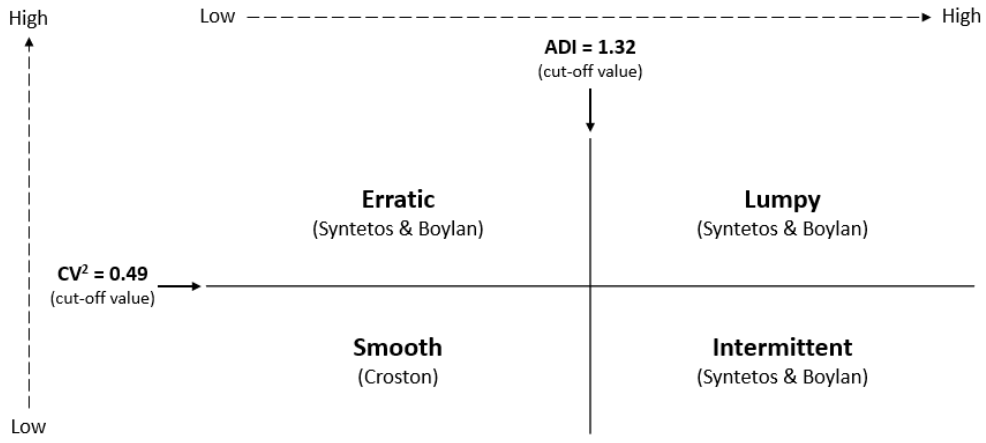

Figure 3.2: Demand-base Categorization for Forecasting

Boylan et al., 2008)) is shown in Figure 3.2. The demand is categorized as "Erratic" when CV^2 exceeds and ADI falls below the respected cut-off values. The "Lumpy" category is defined by both CV^2 and ADI being greater than their respective cut-offs. Conversely, when both parameters are below the cut-off values, the demand

is classified as "Smooth". Finally, the "Intermittent" category is characterized by a lower CV^2 and a higher ADI than the established cut-off values (Boylan et al., 2008).

Classification is considered one of the most extensively studied problems by researchers in the fields of machine learning and data mining (Baradwaj and Pal, 2012. as cited in Soofi and Awan, 2017). In this study, K-Means Algorithm is applied for demand classification in addition to demand categorization of Boylan et al. (2008).

The K-Means algorithm is an unsupervised clustering method based on division, commonly used in data mining and pattern recognition (Li and Wu, 2012). M_j and S_j, the mean vectors for each class, are mathematically expressed as:

$$M_j = \frac{1}{N_j} \sum_{x \in S_j} x \quad \text{and} \quad N_j = |S_j| \tag{3.3}$$

where $\mathcal{X} = \{X_1, X_2, \ldots, X_N\}$ are the sample pattern set, which is categorized into C classes. These classes are denoted as $S_1, S_2, \ldots, S_c$. In Equation 3.3, N_j refers to the number of samples in class S_j. Then, the clustering criterion function J is expressed in Equation 3.4:

$$J = \sum_{j=1}^{c} \sum_{x \in S_j} \|x - M_j\|^2 \tag{3.4}$$

Here, J represents the sum of squared inaccuracies between each sample and its corresponding mean vector across all classes (Li and Wu, 2012).

3.2 Demand Forecasting Methods

3.2.1 Holt-Winters' Seasonal Method

Winters (1960) introduced a forecasting method designed to account for trend, seasonality, and overall sales levels. Forecasting models can incorporate either additive or multiplicative seasonal effects, determined by the relationship between the sales level and the seasonal pattern. An additive model is ideal for situations where the seasonal pattern's intensity remains independent of the sales level, while a multiplicative model is more appropriate when the intensity of the seasonal pattern changes in proportion to the sales level (Winters, 1960). For the multiplicative model, the

formulas for adjusting level, trend and seasonal factor, L_t, T_t, and I_t respectively, are:

$$L_t = \alpha \left(\frac{X_t}{I_{t-p}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1}) \tag{3.5}$$

$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \tag{3.4}$$

$$I_t = \delta \left(\frac{X_t}{L_t} \right) + (1 - \delta)I_{t-p} \tag{3.6}$$

where α, γ, δ represent the smoothing parameters, p indicate the count of observations within a seasonal cycle and X_t is the observed time series. $\hat{X}_t(k)$ represents the updated at time t for the value k periods ahead, as defined by Equation 3.7. (Chatfield and Yar, 1988).

$$\hat{X}_t(k) = (L_t + kT_t)I_{t-p+k} \tag{3.7}$$

3.2.2 Croston Modification: Syntetos-Boylan Approximation (SBA)

Croston (1972, as cited in Pinçe et al., 2021) analytically demonstrated biasness of exponential smoothing for estimating intermittent demand and proposed to splitting the demand process as inter-demand time and demand size. Each component is forecasted separately using exponential smoothing, with updates occurring only during periods of positive demand. The forecasts are subsequently merged to predict the demand for each unit of time (Pinçe et al., 2021).

If $z_t = 0$:

$$\hat{s}_t = \hat{s}_{t-1} \tag{3.8}$$

$$\hat{z}_t = \hat{z}_{t-1} \tag{3.9}$$

If $z_t \neq 0$:

$$\hat{s}_t = \alpha s_t + (1 - \alpha)\hat{s}_{t-1} \tag{3.10}$$

$$\hat{z}_t = \alpha z_t + (1 - \alpha)\hat{z}_{t-1}, \tag{3.11}$$

where s_t and z_t are inter-demand time and demand size of period t and $\hat{s}_t$ and $\hat{z}_t$ are the next period forecasts of them. The demand forecast $\hat{Y}_t$ is outlined in Equation

3.12 (Pinçe et al., 2021):

$$\hat{Y}_t = \frac{\hat{z}_t}{\hat{s}_t}. \tag{3.12}$$

An analytical approximation of the bias in Croston's method was provided by Syntetos and Boylan (2005, as cited in Pinçe et al., 2021), who also proposed an alternative forecasting approach, Syntetos-Boylan approximation (SBA). According to SBA, the demand forecast $\hat{Y}_t$ is corrected by a coefficient $\left(1 - \frac{\alpha}{2}\right)$:

$$\hat{Y}_t = \left(1 - \frac{\alpha}{2}\right) \frac{\hat{z}_t}{\hat{s}_t} \tag{3.13}$$

where α is the smoothing constant (Pinçe et al., 2021).

3.2.3 AutoRegressive Integrated Moving Average (ARIMA)

ARIMA is a well-known and commonly applied statistical technique for time series forecasting. It combines autoregression (AR), which models the relationship with past values, moving averages (MA), which account for error terms from both current and past periods, and integration (I), which involves differencing to make the data stationary, enabling it to capture complex relationships in stochastic time series (Kannan et al., 2020). The formula of ARIMA model is expressed in Equation 3.14. (Kannan et al., 2020).

$$\hat{F}_t = c + \phi_1 \hat{F}_{t-1} + \phi_2 \hat{F}_{t-2} + \cdots + \phi_p \hat{F}_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \tag{3.14}$$

where:

$\hat{F}_t$: predicted value at time t,

c: a constant term,

$\phi_1, \phi_2, \ldots, \phi_p$: the autoregressive coefficients

$e_t, e_{t-1}, \ldots$: the error terms,

$\theta_1, \theta_2, \ldots, \theta_q$: the moving average coefficients.

3.2.4 Regression-based Methods: Multivariate Linear Regression (MLR) and Multivariate Nonlinear Regression (MNLR)

Linear regression is a suitable algorithm for identifying the connection between independent and dependent variables, as it models the connection between predictors

(independent variables) and response variables (dependent variables) under the assumption of linearity (Aryudha and Hasibuan, 2024).

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n \tag{3.15}$$

where:

Y: Dependent variable

a: Intercept (constant value), as introduced in Equation 3.16.

b: Regression coefficient, as introduced in equation 3.17.

X: Independent variable

$$a = \frac{\sum Y(\sum X^2) - \sum X \sum XY}{n(\sum X^2) - (\sum X)^2} \tag{3.16}$$

where:

$\sum Y$: Total of all Y values

$\sum X^2$: Sum of squares of all X values

$\sum X$: Total of all X values

$\sum XY$: Sum of products between all X and Y values

n: Number of data points

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \tag{3.17}$$

where:

b: Regression coefficient

n: Number of data points

$\sum XY$: Sum of products between all X and Y values

$\sum X$: Total of all X values

$\sum Y$: Total of all Y values

$\sum X^2$: Sum of squares of all X values.

Nonlinear regression analysis models a curvilinear relationship between variables. Here, X represents the independent variables or input data and Y represents the dependent variables or output data, a_0 represents the constant value and b_1, b_2, b_3,...,

b_n represents weighted coefficients of independent variables. A multiple nonlinear regression equation is illustrated in equation 3.18. (Aktepe et al., 2021).

$$Y = a_0 + b_1X_1 + b_2X_2^2 + b_3X_3^3 + \cdots + b_nX_n^n \tag{3.18}$$

3.2.5 Support Vector Regression (SVR)

Support Vector Regression (SVR), based on statistical learning theory, is commonly used in forecast models (Aktepe et al., 2021). Balasundaram and Prasad (2020, as cited in Riwayadi et al., 2024) stated that Support Vector Machines (SVM), originally developed for classification, can be adapted for regression tasks by utilizing the ϵ insensitive loss function:

$$L_\epsilon(y) = \begin{cases} 0 & \text{if } |y - f(x)| \leq \epsilon, \\ |y - f(x)| - \epsilon & \text{otherwise.} \end{cases} \tag{3.19}$$

where the function f and convex optimization problem in linear cases are expressed as below (Riwayadi et al., 2024).

$$f(x) = \mathbf{w}^T\mathbf{x} + b, \quad \mathbf{w} \in \mathbb{R}^d,\, b \in \mathbb{R}. \tag{3.20}$$

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to } \begin{cases} y_i - (\mathbf{w}^T\mathbf{x}_i + b) \leq \epsilon, \\ (\mathbf{w}^T\mathbf{x}_i + b) - y_i \leq \epsilon. \end{cases} \tag{3.21}$$

and where the optimization equations in nonlinear cases is expressed in Equation 3.22 (Riwayadi et al., 2024).

$$\max \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_i^+ - a_i^-\right)\left(a_j^+ - a_j^-\right)k(X_i, X_j) - \epsilon\sum_{i=1}^{N}\left(a_i^+ - a_i^-\right) - \sum_{i=1}^{N} y_i\left(a_i^+ - a_i^-\right) \tag{3.22}$$

Subject to:

$$\sum_{i=1}^{N}\left(a_i^+ - a_i^-\right) = 0, \quad 0 \leq a_i^+ - a_i^- \leq c \tag{3.23}$$

Selecting the appropriate kernel function such as polynomial, normalized polynomial, radial basis, and Pearson VII (PUK) and its optimal parameters, is critical for

effective SVM regression (Aktepe et al., 2021). Riwayadi et al. (2024) highlighted the critical role of parameter tuning focusing on cost (c) and epsilon (ϵ) in enhancing the precision and efficiency of the SVR model, demonstrating its potential for practical applications in forecasting and decision-making.

3.2.6 Artificial Neural Networks

Neural Networks (NNs) are adaptable, nonlinear, data-centric models that are highly suited to forecasting. Unlike conventional statistical time-series methods, that usually struggle to captures patterns driving the data due to restrictive functional assumptions, NNs adapt directly to the data-generating process (Kourentzes, 2013). The feedforward Multilayer Perceptron (MLP) represents the most commonly utilized type of NN for forecasting (Kourentzes, 2013). Zhang et al. (1998, as cited in Kourentzes, 2013) provided a comprehensive explanation of these models and their application to forecasting.

$$Y_t' = \beta_0 + \sum_{h=1}^{H} \beta_h g \left(\gamma_{0h} + \sum_{i=1}^{I} \gamma_{hi} X_i \right), \tag{3.24}$$

where:

Y_t': Describes the one-step ahead forecast,

p_i: Represents the lagged inputs,

β_0: Bias for the output layer,

β_h: Weight coefficients for the output layer,

g: Non-linear transfer function, often sigmoid or hyperbolic tangent (TanH),

γ_{0h}: Bias for the hidden layer,

γ_{hi}: Weights connecting input nodes to hidden nodes,

H: Number of hidden nodes,

I: Number of input nodes,

X_i: Input variables.

3.2.7 Random Forest Regressor (RF)

The random forest, a machine learning ensemble technique, is applied to time series analysis by constructing multiple decision trees using randomly selected subsets of

variables. This process, called bagging, combines the outputs of all trees to produce a final estimation, enhancing accuracy and robustness (Panda and Mohanty, 2023).

The decision trees are refined through a recursive optimization process after constructing. For each split, a random subspace of k features is chosen, where $k \leq l$ and l denotes the total number of features. This randomness in feature selection introduces diversity among the decision trees, improving the ensemble's robustness. The selection of the optimal variable for splitting is determined using variance reduction as a criterion (Panda and Mohanty, 2023). The variance at a given node N is defined as:

$$Var(N) = \frac{1}{|N|} \sum_{y \in N} (y - \bar{y})^2, \tag{3.25}$$

where $|N|$ is the number of samples at node N, y is the target value, and $\bar{y}$ is the mean prediction.

To quantify the improvement obtained through the split and identify the optimal variable for partitioning, the r decrease in variance obtained by dividing the node into left N_l and right N_r child nodes is expressed in equation 3.26 (Panda and Mohanty, 2023).

$$Reduction = Var(N) - \left(\frac{|N_l|}{|N|} Var(N_l) + \frac{|N_r|}{|N|} Var(N_r) \right) \tag{3.26}$$

The random forest model determines its ultimate output by combining the outputs of all D decision trees within the ensemble. By aggregating these results, the model reduces variance and improves generalization, thereby enhancing the overall predictive performance. The aggregated prediction $\hat{y}_t$ is mathematically expressed as:

$$\hat{y}_t = \frac{1}{D} \sum_{d=1}^{D} T_d(x_t), \tag{3.27}$$

where T_d illustrates the estimation of the d-th decision tree, x_t is the input data at time t, and D denotes the total number of trees (Panda and Mohanty, 2023).

3.2.8 XGBoost and LightGBM

The XGBoost algorithm is a gradient boosting framework that integrates multiple weak classifiers to form a robust classifier, supporting Classification and Regres-

sion Trees (CART) and linear classifiers in the role of base learners. By employing a second-order Taylor expansion of the cost function, XGBoost captures more detailed information, significantly enhancing its computational efficiency and predictive performance (Zhang et al., 2021).

The Gradient Boosting algorithm iteratively constructs weak prediction models to form the final predictive model F. At each step m, the model F_m is updated by adding a new base learner $h_{m+1}(x)$, which approximates the residuals between the true target y and the current model's prediction F_m. The updated model at step $m+1$ is expressed in equation 3.28 (Zhang et al., 2021).

$$F_{m+1} = F_m + h_{m+1}(x) \tag{3.28}$$

The objective is to determine $h_{m+1}(x)$, where the negative gradient of the loss function serves as the residual for constructing the base learner $h(x)$.

XGBoost minimizes the residuals by iteratively appending weak regression trees to the ensemble. The predicted value $\hat{y}_i$ is given in Equation 3.29 (Zhang et al., 2021).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F \tag{3.29}$$

where:

K: number of regression trees,

F: set of all regression trees,

f_k: k-th regression tree .

To ensure accurate predictions while maintaining generalization, the objective function of minimization $Obj^{(t)}$ is expressed in Equation 3.30 (Zhang et al., 2021).

$$Obj^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^{t} \Omega(f_t) + constant \tag{3.30}$$

where:

$l(y_i, \hat{y}_i^{(t)})$: loss function, demonstrating the error between forecasted and actual values,

$\Omega(f_t)$: regularization term, restricting model complexity.

The regularization term is shown in Equation 3.31:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{3.31}$$

where:

T: the number of leaf nodes,

w: the leaf weights,

γ: the complexity penalty for the leaf nodes,

λ: the regularization parameter.

XGBoost applies a second-order Taylor expansion of the loss function to optimize the objective (Zhang et al., 2021). The simplified objective function is expressed in Equation 3.32:

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{3.32}$$

where:

g_i: the first-order derivative of the loss function,

h_i: the second-order derivative of the loss function.

Mathematical expressions of g_i and h_i are given in Equation 3.33:

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \tag{3.33}$$

In a specific split, the optimal weight w_j^* of a leaf node j is determined by equating the derivative of the objective function as zero:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{3.34}$$

where:

I_j: the set of sample indices for the leaf node j.

Finally, the optimal objective value $L^{(t)}(q)$ for the split is computed as:

$$L^{(t)}(q) = -\frac{1}{2} \left[\sum_{j=1}^{T} \frac{\left(\sum_{i \in I_{Rj}} g_i \right)^2}{\sum_{i \in I_{Rj}} h_i + \lambda} + \frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} \right] - \gamma \tag{3.35}$$

This iterative approach allows XGBoost to minimize the objective function while balancing prediction accuracy and model complexity through regularization (Zhang et al., 2021).

Even though XGBoost improved the Gradient Boosted Decision Trees (GBDT) framework and achieved success in machine learning competitions, its need to scan the entire dataset for each feature makes it computationally expensive (Deng et al., 2021). Ke et al. (2017, as cited in Deng et al., 2021) proposed Gradient-based One-Side Sampling and Exclusive Feature Bundling, leading to the development of LightGBM. It significantly enhances GBDT efficiency. LightGBM accelerates gradient calculations using a Histogram-based approach by building statistical histograms for each feature to determine optimal split points. Additionally, histograms for child nodes are derived from their parent and sibling nodes, further enhancing computational efficiency (Deng et al., 2021). LightGBM employs a Leaf-wise growth strategy for regression trees and adopts Gradient-based One-Side Sampling to sample and assign weights to data points. LightGBM, as described by Deng et al. (2021), is an efficient framework for implementing Gradient Boosted Decision Trees, providing rapid training, efficient memory utilization, and high precision.

Chapter 4

INVENTORY MANAGEMENT FOR SPARE PARTS

4.1 Inventory Policies

The timing and quantity of orders, determined by inventory policies, can be classified into continuous or periodic review systems, typically defined by parameters such as the order-up-to level (S), review period (R), order quantity (Q), and reorder point (s) (Vandeput, 2020).

The reorder point s, the fixed order quantity Q^*, and the safety stock S_s for the continuous review policy (s, Q) are expressed in Equations 4.1, 4.2 and 4.3, respectively (Vandeput, 2020).

- The optimal order quantity Q^*:

$$Q^* = \sqrt{\frac{2kD}{h}} \tag{4.1}$$

where:

k: the fixed transaction costs,

D: demand,

h: the holding costs.

- The reorder point s:

$$s = d_L + S_s \tag{4.2}$$

where:

d_L: demand during the lead time,

S_s: safety stock.

- Safety Stock S_s:

$$S_s = z_\alpha \sqrt{(\mu_L)\sigma_d^2 + \mu_d^2 \sigma_L^2} \tag{4.3}$$

where:

μ_L: mean lead time,

σ_d: standard deviation of demand,

μ_d: mean demand,

σ_L: standard deviation of the lead time,

z_α: factor corresponding to the desired service level,

α: cycle service level calculated as $\Phi(z_\alpha)$.

The review period R, the order-up-to level S, and the safety stock S_s for the periodic review policy (R, S) are defined in Equations 4.4, 4.5, and 4.6, respectively (Vandeput, 2020).

- The review period R:

$$R = 2^k T_B \tag{4.4}$$

where:

k: the fixed transaction costs,

T_B: the base time period for the review cycle.

- The order-up-to level S:

$$S = d_L + d_R + S_s \tag{4.5}$$

where:

d_L: demand during the lead time,

d_R: demand during the review period,

S_s: safety stock.

- Safety Stock S_s:

$$S_s = z_\alpha \sqrt{(\mu_L + R)\sigma_d^2 + \mu_d^2 \sigma_L^2} \tag{4.6}$$

where:

μ_L: mean lead time,

R: review period,

σ_d: standard deviation of the demand,

μ_d: mean demand,

σ_L: standard deviation of the lead time,

z_α: factor corresponding to the desired service level,

α: Cycle service level calculated as $\Phi(z_\alpha)$.

The above calculations are based on a unified safety stock model designed to account for two key sources of variation: demand fluctuations throughout the lead time and across the average lead time. This model enables the development of robust inventory policies that can handle stochastic demand, variable lead times, and review periods, while assuming demand and lead times are continuous, normally distributed, and independent of each other, with excess demand being backordered and orders allowed to cross (Vandeput, 2020).

4.2 Newsvendor Model

The newsvendor model, a widely explored inventory model in academic studies, addresses seasonal demand while considering two primary cost components: overage costs (c_o), incurred from excess inventory and underage costs (c_u), resulting from insufficient inventory (Vandeput, 2020). The objective of the newsvendor model is to minimize the overall cost, expressed as:

$$\min_{q \geq 0} \mathbb{E}\left[c_u(D-q)^+ + c_o(q-D)^+\right], \tag{4.7}$$

where q is the order quantity, D is the random demand, c_u is the underage cost, and c_o is the overage cost (Huber et al., 2019).

The optimal solution to this problem involves choosing the order quantity q^* that aligns with the quantile of the CDF F, satisfying:

$$q^* = \inf\left\{p : F(p) \geq \frac{c_u}{c_u + c_o}\right\}, \tag{4.8}$$

Here, the optimal service level is $\frac{c_u}{c_u+c_o}$ (Huber et al., 2019). The service level indicates the likelihood of satisfying a demand within a specified period. In most real-world scenarios, the demand distribution F is not known, but past records

are usually available. This problem can be addressed through traditional model-based optimization or data-driven optimization with Sample Average Approximation (SAA), where both approaches utilize point forecasts and historical estimation errors to guide inventory decisions (Huber et al., 2019).

In the model-based approach, the order quantity $q(\mathbf{x})$ is determined by combining the mean forecast $\hat{y}(\mathbf{x})$ with the service level quantile:

$$q(\mathbf{x}) = \hat{y}(\mathbf{x}) + \inf\left\{p : \hat{F}(p, \hat{\theta}) \geq \frac{c_u}{c_u + c_o}\right\}. \tag{4.9}$$

where the forecast error distribution $\hat{F}$ is assumed (e.g., normal with parameter $\hat{\theta}$) (Huber et al., 2019).

In the data-driven optimization using Sample Average Approximation (SAA), the error distribution $\bar{F}$ is estimated empirically from forecast errors $\epsilon_1, \ldots, \epsilon_n$, without requiring any distributional assumption. The empirical distribution is expressed in Equation 4.10.

$$\bar{F}(p) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\epsilon_i \leq p). \tag{4.10}$$

The order quantity $q(\mathbf{x})$ is determined by optimizing it to align with the quantile of the service level provided from the empirical distribution and subsequently adding it to the mean forecast. $\hat{y}(\mathbf{x})$, resulting in Equation 4.11 (Huber et al., 2019).

$$q(\mathbf{x}) = \hat{y}(\mathbf{x}) + \inf\left\{p : \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\epsilon_i \leq p) \geq \frac{c_u}{c_u + c_o}\right\}. \tag{4.11}$$

In this study, the order quantity will be optimized using the mean forecast $\hat{y}$ obtained from various demand forecasting models and then will be combined with a data-driven optimization approach based on Sample Average Approximation (SAA). The (R, S) inventory policy will be applied without safety stock, and the optimized order quantity $q(\mathbf{x})$ will serve as the order-up-to level S.

Chapter 5

EMPIRICAL RESULTS AND DISCUSSIONS

This chapter begins with an overview of the data and proceeds to explain the demand classification process. Next, it discusses data aggregation and feature analysis. The classification data for each part and year, derived from demand classification, is incorporated as a feature in the demand forecasting models. Following this, the implementation of 13 demand forecasting models is presented, along with their results and a detailed discussion. Models with more than 5,000 unique part codes and R^2 values greater than 0.1 are selected for inventory optimization, while the others are excluded. Finally, the chapter concludes with the application of inventory management techniques and an analysis of their outcomes for the remaining 8 demand forecasting models. All analyses and models in the study are implemented using Python.

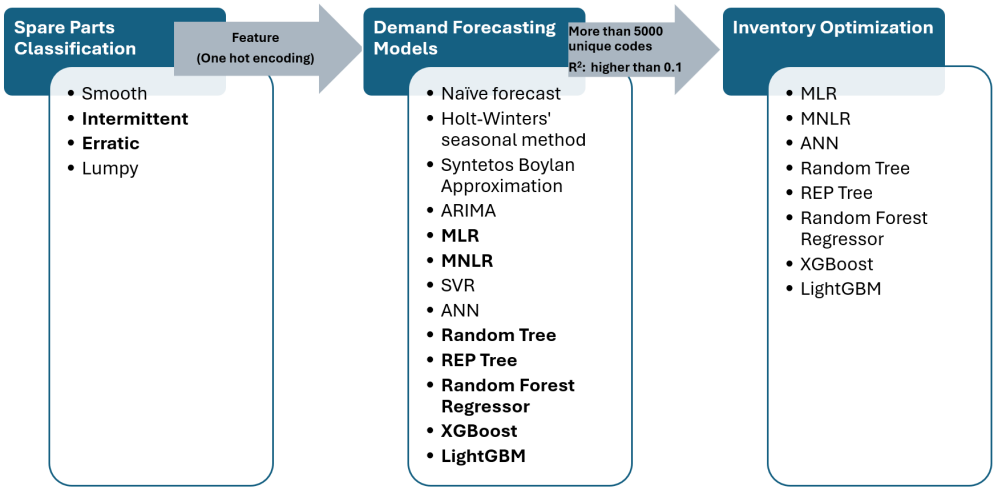

Figure 5.1: Data Flowchart

5.1 Data Preparation

The dataset comprises 8,517 unique spare part SKUs shipped between 2013 and 2024. It was collected in a monthly format from the case company and originally contained 94,839 rows, representing the concatenated non-zero shipment quantities for each SKU's time series. To ensure continuity, the periods between the first and last non-zero shipment for each SKU were filled with zeros, resulting in a final dataset of 442,385 rows.

The shipment quantities have a mean of 9.6, a minimum of 0, a maximum of 76,838, and a standard deviation of 345.4. The sum of shipped quantities from the previous year and the sum of shipped quantities from the previous two years are used as separate features in the demand forecasting models.

Each spare part in the data set is originally listed in the bill of materials (BOM) for a final product. BOMs link the information about the product and the part and was obtained from the case company. In addition, case company provided SKU-level data on air conditioner installation quantities from 2011 to 2023. Finally, the installed quantity, in other words, the parts in use, was aggregated by combining information from the BOM and installation quantity data. For example, in July 2023, five different products, each including one unit of spare part 5400061404, were installed with quantities of 51, 36, 14, 3, and 2, respectively. By summing the installation quantities across these products, the total installation quantity for this specific part in July 2023 amounted to 106 units. The sum of installation quantities from the previous year and the sum of previous two years are used as separate features in the forecasting models.

The demand forecast of the final product could potentially be used as a feature in the model. However, it is not utilized initially, as the need for spare parts is relatively low during the early stages of a specific product's lifecycle. Rather than incorporating it immediately, the demand for the spare part will be included as a feature in the following year once it begins.

5.2 Demand Classification

To implement the demand classification method proposed by Boylan et al. (2008), as detailed in Chapter 3.1, the average inter-demand interval (ADI) and the coefficient of variation of demand (CV^2) are implemented for each part and each year. Based on these metrics, SKUs for each year are classified as smooth, intermittent, erratic, or lumpy, using the recommended threshold values: 1.23 for ADI and 0.49 for CV^2.

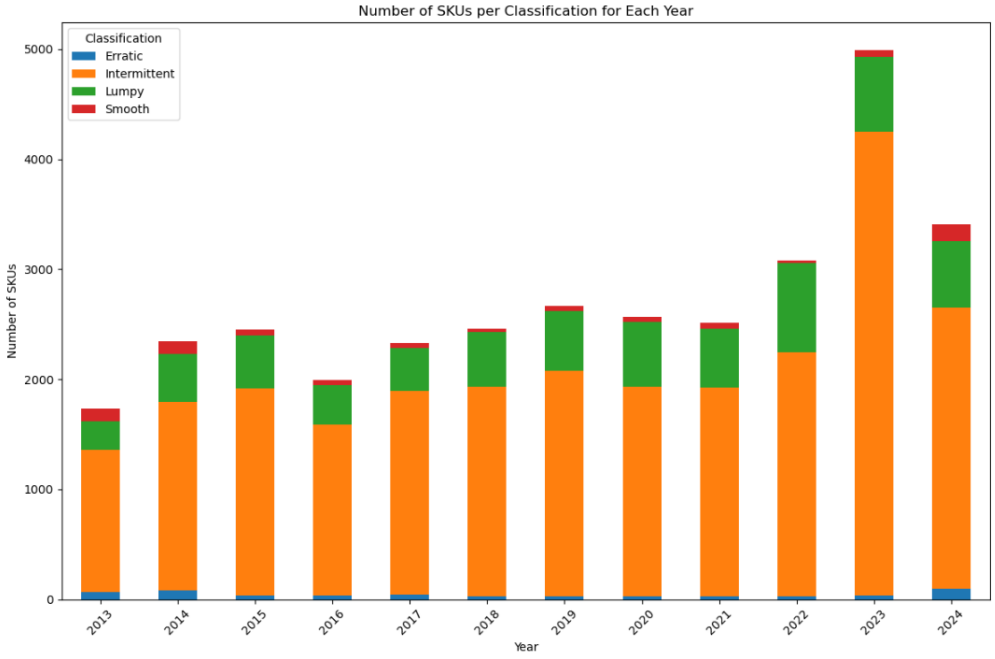

Figure 5.2: Demand classification per year

Figure 5.2 shows the distribution of SKUs among demand classifications for each year. The results show that the majority of parts exhibit intermittent demand characteristics, followed by those classified as lumpy across all years.

Table 5.1 presents the yearly percentages of each demand classification along with their averages. Intermittent demand is defined by an Average Inter-Demand Interval (ADI) higher than 1.23 while a coefficient of variation of demand (CV^2) lower than 0.49, representing an average of 76.17% of the total data. Smooth demand occurs when both the ADI and CV^2 are low, specifically less than 1.23 and 0.49, respectively. This category accounts for 2.63% of the total data. Erratic demand

Table 5.1: Yearly Percentages of Demand Classification with Average

Year	Erratic (%)	Intermittent (%)	Lumpy (%)	Smooth (%)
2013	3.92	74.19	15.26	6.62
2014	3.63	73.01	18.53	4.82
2015	1.51	76.64	19.72	2.12
2016	1.70	77.87	17.83	2.60
2017	1.73	76.16	20.38	1.73
2018	1.18	77.43	20.13	1.26
2019	1.09	76.75	20.47	1.69
2020	1.21	74.10	23.09	1.60
2021	1.15	75.52	21.34	1.99
2022	0.96	71.00	26.11	1.92
2023	0.70	86.46	12.00	0.84
2024	2.88	74.93	17.85	4.34
Average	1.80	76.17	19.39	2.63

is observed when the ADI is low (less than 1.23), but the CV^2 is high, exceeding 0.49. Meanwhile, lumpy demand is classified when both the ADI and CV^2 are high, exceeding 1.23 and 0.49, respectively. Lumpy demand makes up 19.39% of the total data, while erratic demand represents the smallest proportion at just 1.8%.

Examples of intermittent, smooth, erratic, and lumpy demand patterns are presented in Figures 5.2, 5.3, 5.4, and 5.5, respectively.

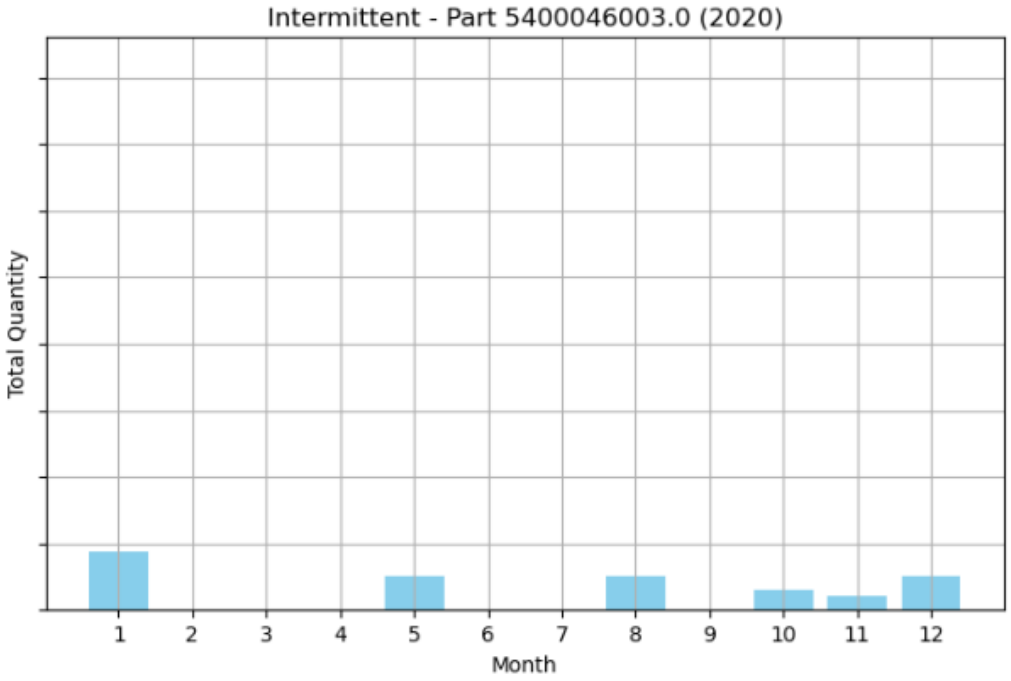

Figure 5.3: Intermittent demand pattern

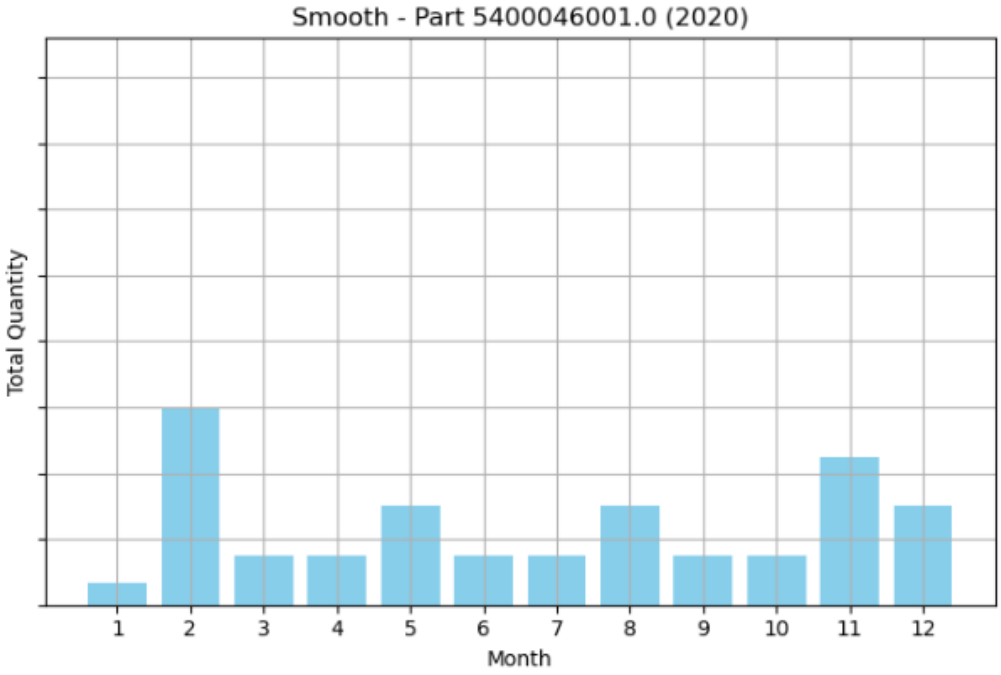

Figure 5.4: Smooth demand pattern

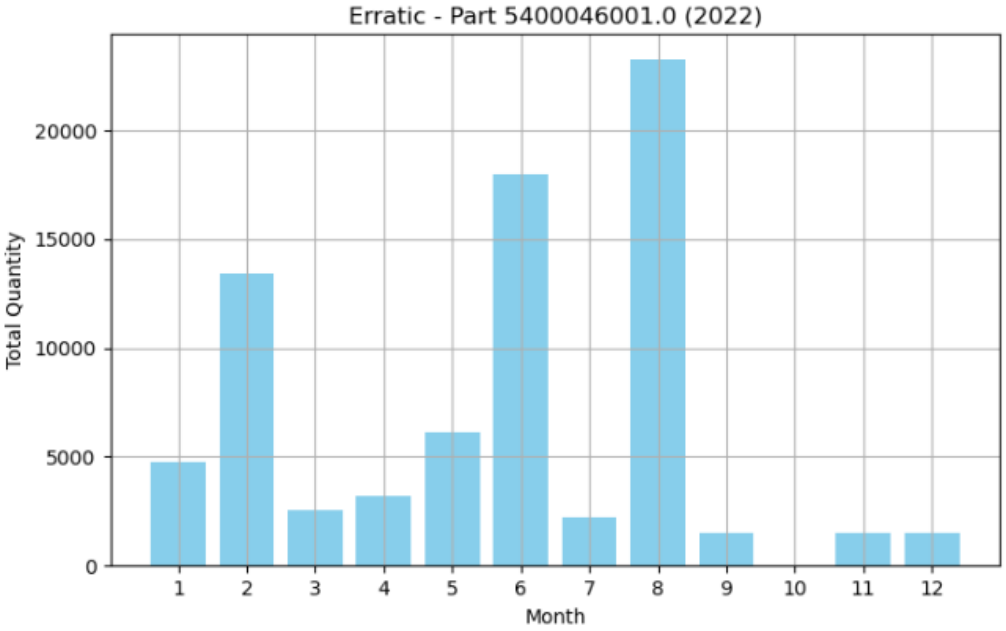

Figure 5.5: Erratic demand pattern

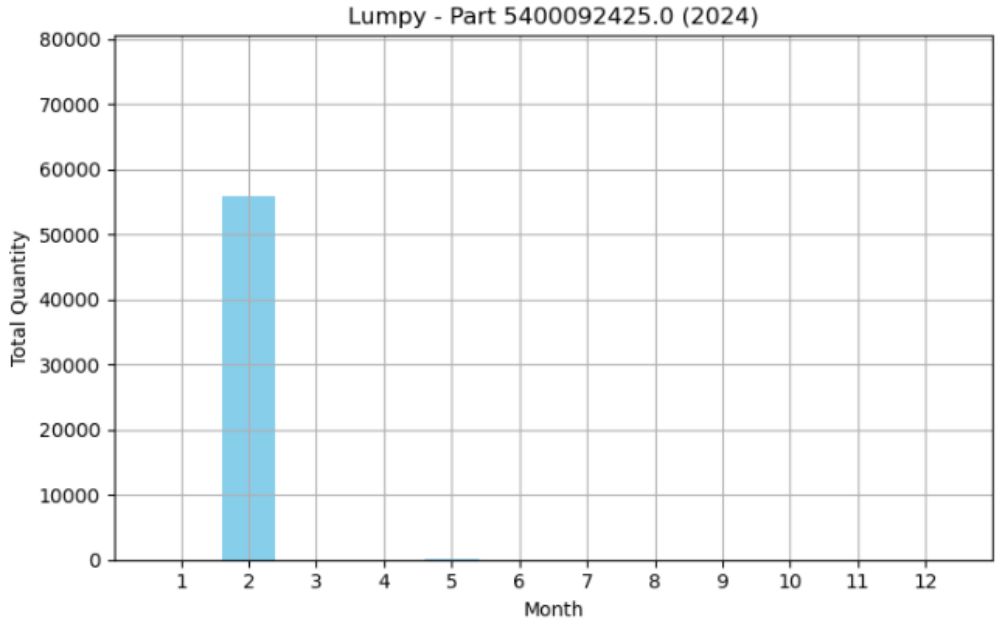

Figure 5.6: Lumpy demand pattern

Additional to classification by the thresholds recommended by Boylan et al. (2008), the demands are categorized using K-means clustering. It is applied to the dataset with 4 clusters for each year by predicting the cluster labels based on the ADI and CV^2 values. Figure 5.7 illustrates the distribution of SKUs across clusters for each year. In the traditional classification approach, the majority of parts exhibit an intermittent demand characteristic. In contrast, K-means clustering reveals no predominant demand pattern across the 4 clusters.

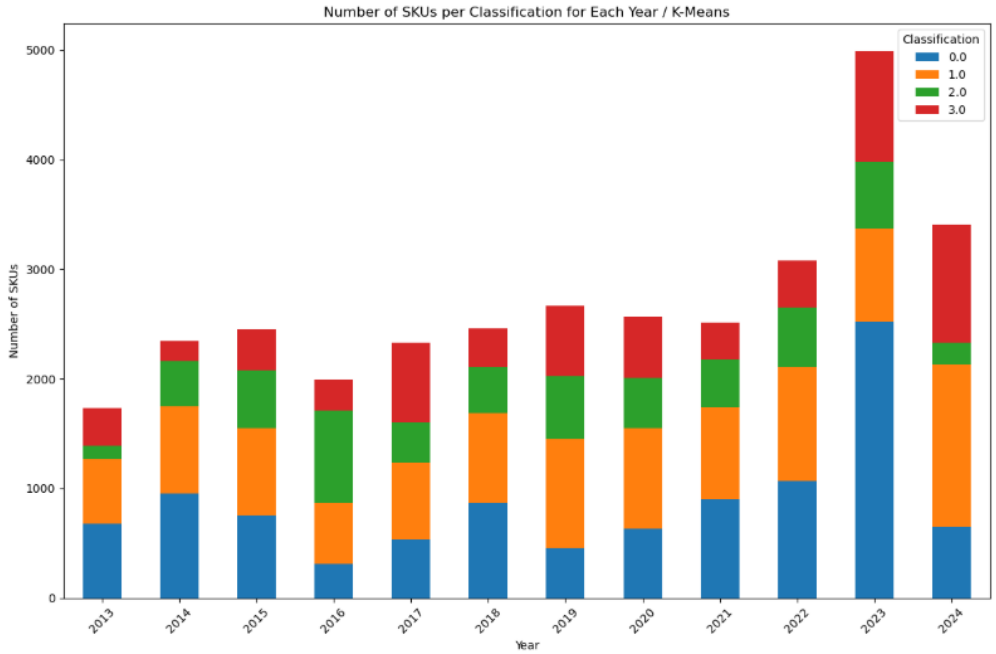

Figure 5.7: Demand classification per year / K-means

Cluster	**ADI**	CV^2
0.0	4.85	0.36
1.0	10.05	0.09
2.0	4.70	0.36
3.0	3.04	0.47

Table 5.2: Cluster averages for ADI and CV2.

Table 5.2 presents the average values of ADI and CV^2 for each cluster. Cluster 1.0

exhibits the highest ADI and the lowest CV^2, like the erratic demand characteristic, while cluster 3.0 displays the lowest ADI and the highest CV^2, similar to intermittent demand characteristic. The values for clusters 0.0 and 2.0 are similar to each other. Furthermore, all average ADI values exceed the recommended threshold of 1.23, while all average CV^2 values fall below the recommended threshold of 0.49. Figures 5.8 and 5.9 compare two classification methods: Figure 5.8 shows the classification result of Boylan et al. (2008) (erratic, lumpy, smooth, and intermittent) based on predefined ADI and CV^2, while Figure 5.9 shows K-means clustering, illustrating a different approach to grouping the data based on the ADI and CV^2 values.

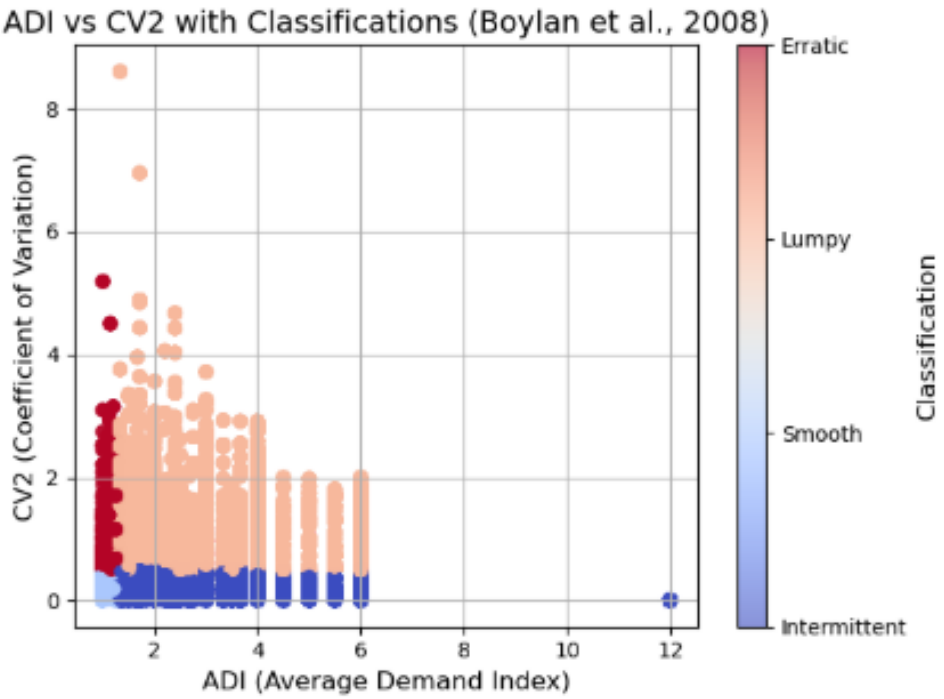

Figure 5.8: ADI vs CV^2 with Classifications of Boylan et al. (2008)

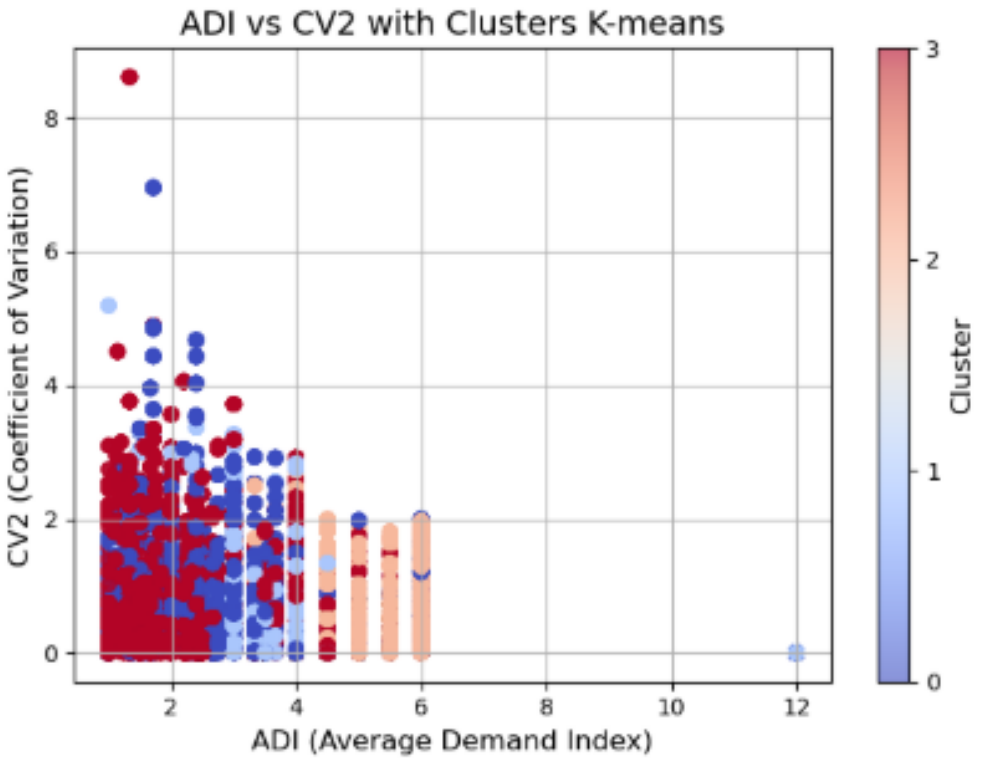

Figure 5.9: ADI vs CV^2 with Clusters K-means

Classification based on the predefined threshold of Boylan et al. (2008) is used for the modelling rather than the machine learning classification to be align with the current literature. Additionally, this demand classification is incorporated into the forecasting models as a feature using one-hot encoding to represent the different demand categories.

5.3 Data Aggregation and Feature Analysis

Multiple feature analysis approaches, including the correlation matrix, mutual information scores, feature importance from a Random Forest model, and permutation importance, were implemented to investigate how features relate to and influence the target variable.

The correlation matrix in Figure 5.10 offers important insights into the relationships between features, particularly their influence on total shipment quantity (Total). A moderate correlation is observed between Total and previous shipped

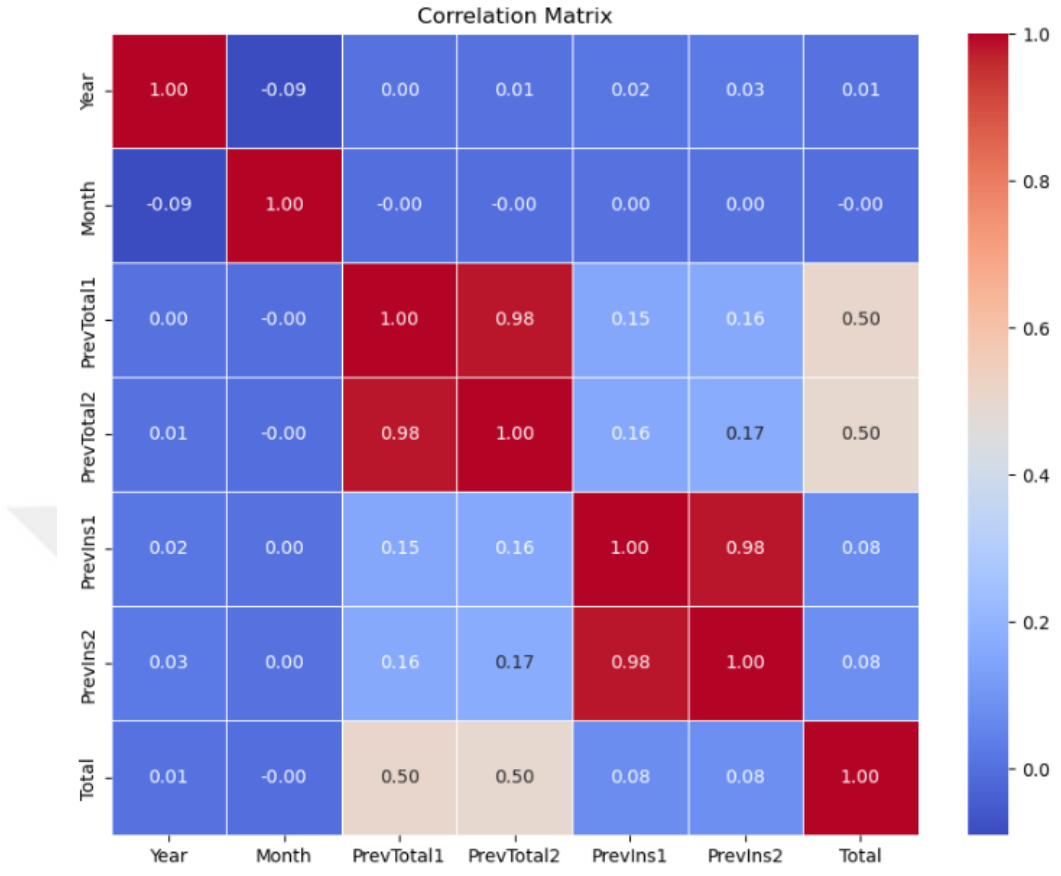

Figure 5.10: Correlation Matrix

quantities (PrevTotal1 and PrevTotal2), indicating that historical shipment data plays a significant role in predicting current shipments. In contrast, the correlation between Total and previous installed quantities (PrevIns1 and PrevIns2) is lower. Additionally, the correlations involving year and month with other features are minimal. It shows that direct relationship of month and year with shipment quantities is less prominent. Overall, the analysis highlights the strong contribution of historical shipment quantities while revealing varying levels of influence from other features.

The Mutual Information (MI) scores measure how much knowledge one random variable provides about another (Alduailej et al., 2022). The scores, as shown in Figure 5.11, closely align with the insights from the correlation matrix. MI highlights the significant influence of historical shipment quantities on total shipment quantity. Both analyses identify the total shipped quantity from the previous year (PrevTo-

tal1) and the sum of the previous two years (PrevTotal2) as the most impactful features and emphasize the critical role of cumulative historical shipment data in predicting current shipments. Additionally, the total installed quantities from the previous year (PrevIns1) and the sum of the previous two years (PrevIns2) demonstrate notable contributions in the MI scores. The feature "Year," that showed minimal correlation in the matrix, shows moderate importance in the MI scores while "Month" maintains its low significance across both analyses. Among the classification features, Classification Intermittent emerges as the most impactful, followed by Classification Lumpy, Classification Erratic, and Classification Smooth.

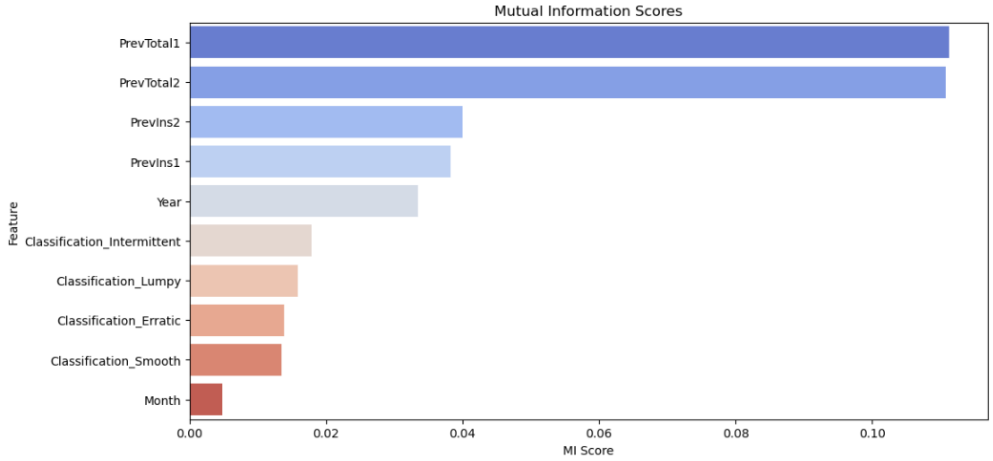

Figure 5.11: Mutual Information Scores

Random Forest (RF) builds a group of weakly correlated classification or regression trees and combines their outputs for improved predictions. A notable aspect of RF is its capacity to calculate feature importance metrics, such as the Gini index and the mean decrease in accuracy (Iranzad and Liu, 2024). The feature importance results from the RandomForestRegressor, as shown in Figure 5.12, offer both complementary and contrasting insights compared to the correlation matrix and Mutual Information (MI) scores. The feature importance from Random Forest identifies "Month" as the most significant feature and contrasts with its minimal dependency observed in the correlation matrix and MI scores. The consistent importance of the sum of shipped quantities from the previous year (PrevTotal1) and

the last two years (PrevTotal2) across all methods reinforces their critical role in forecasting. The feature "Year", which showed minimal correlation and moderate MI scores, gains higher importance in Random Forest. The total installed quantities (PrevIns1 and PrevIns2) maintain moderate importance across all methods, while demand classification features show lower but consistent contributions.

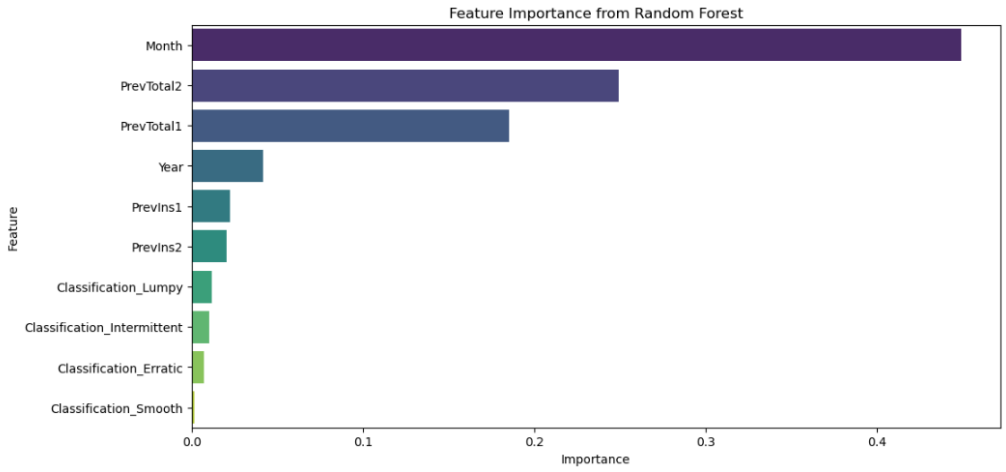

Figure 5.12: Feature Importance from Random Forest

Permutation importance, developed by Altmann et al.(2010, as cited in Iranzad and Liu (2024)), is a technique for selecting relevant predictors by permuting the response variable. It achieves this by permuting the response variable and fitting Random Forests and allows the selection of predictors that maintain their importance despite the permutation (Iranzad and Liu, 2024). The results, as shown in Figure 5.13, provide additional insights to findings from the Random Forest feature importance, correlation matrix, and Mutual Information (MI) scores. Consistent with the previous analyses, the sum of shipped quantities from the last two years (PrevTotal2) and the previous year (PrevTotal1) emerge as the most critical features. "Month" also remains a key contributor, consistent with the feature importance results, and indicates the significance of seasonal patterns, on contrary the correlation matrix and MI scores. The feature "Year" demonstrates moderate importance, while the total installed quantities (PrevIns1 and PrevIns2) retain their lower importance. Among the classification features, Classification Erratic shows

slightly more relevance compared to other categories, but all classification features contribute less overall. These results collectively confirm the pivotal role of historical shipment quantities and highlight the seasonal influence of "Month".

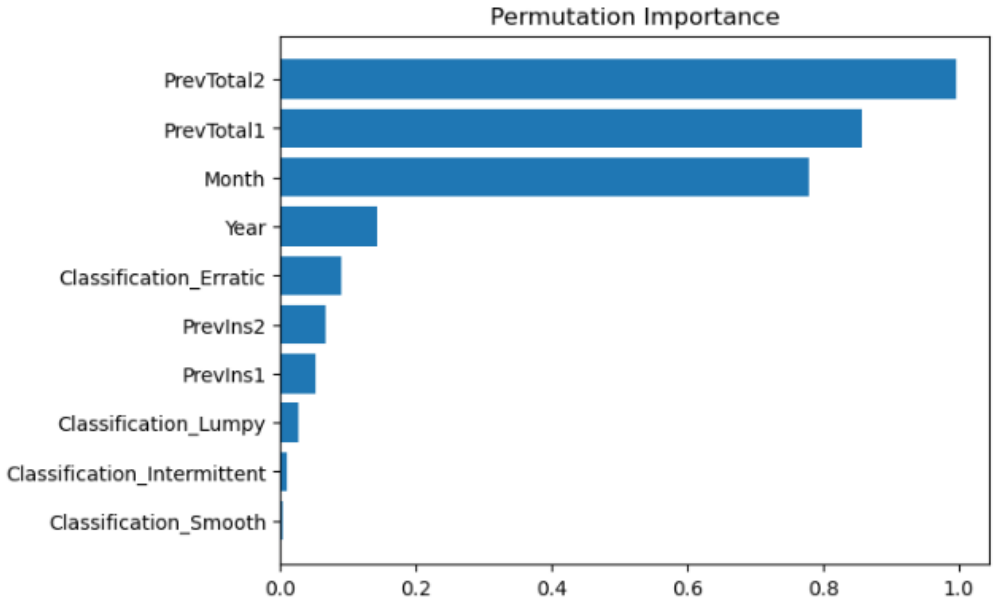

Figure 5.13: Permutation Importance

Based on the findings from all analyses that emphasized the importance of different features in various ways, all features will be included in the models. This approach ensures a comprehensive representation of the factors influencing total shipment quantities.

5.4 Implementation of Demand Forecasting Models

13 different forecasting models, as described in Chapters 2.2 and 3.2, are applied to the data series. For each model, the processed part data frame is prepared based on the required minimum time period data. The processed parts are then divided into two as training and testing sets, with 80% of the data allocated for training and 20% for testing.

The Naïve forecast is implemented as the baseline model for parts with more than five time periods of data. A total of 2,208 unique parts are excluded due to

insufficient data, while the remaining 6,309 parts are used for prediction using the Naïve method. The Mean Squared Error (MSE) of the test set is 479,624.

The Holt-Winters seasonal method is applied to parts with data covering more than two full seasons. A total of 3,816 unique parts are excluded due to insufficient data, while the remaining 4,701 parts are used. The model parameters are selected to capture the seasonal and structural characteristics of the shipment data. An additive seasonal component with an annual cycle (seasonal='additive', seasonalperiods=12) is implemented to account for consistent seasonal patterns. Initial values of level, trend, and seasonality are automatically estimated (initializationmethod='estimated'). Bias correction (removebias=True') is incorporated to enhance forecast accuracy, and predictions are restricted to non-negative values to represent shipment quantities accurately. The Mean Squared Error (MSE) of the test set is 458,563 under the additive seasonal component. However, when a multiplicative seasonal component is used, the MSE drastically increases to 1.23×10^{11}, making it the highest MSE observed among all the models. Given the significantly higher MSE observed with the multiplicative seasonal component, only the results from the additive seasonal model will be included in the summary tables.

Syntetos-Boylan Approximation (SBA) is applied to parts with more than five time periods of data. A total of 2,208 unique parts are excluded due to insufficient data, while 6,309 parts are used for forecasting. The smoothing parameter α is set as 0.1. As cited in Altay et al. (2008), Croston employed a unified smoothing constant to produce both forecasts and advised using low values for it, such as between 0.1 and 0.2. The smoothed demand level and interval between non-zero demands are iteratively updated, with forecasts determined by the ratio of smoothed demand to the interval and scaled by $\left(1 - \frac{\alpha}{2}\right)$ to correct biases. Forecast values are constrained to be non-negative, reflecting the nature of shipment quantities. The test set Mean Squared Error (MSE) is $521,477$.

The auto ARIMA model is applied to 4,077 parts with sufficient data, utilizing seasonal adjustments and optimized parameter selection to capture trends and seasonality. A total of 2,208 parts are excluded due to insufficient data. A log transformation ($\log(1 + x)$) is used on the target variable to preserve non-zero val-

ues in the inventory data. Key parameters include exploring AR (Auto-Regressive) terms and MA (Moving Average) terms up to a maximum of 2 (`max_p=2`, `max_q=2`), as well as seasonal AR and MA terms up to a maximum of 1 (`max_P=1`, `max_Q=1`). Seasonal adjustments are enabled (`seasonal=True`) with a seasonal period of 12 months (`m=12`), and a stepwise search method (`stepwise=True`) is used to optimize parameter selection efficiently. The Mean Squared Error (MSE) of the test set is $448,704$. Then, the XARIMA (Exponential Autoregressive Integrated Moving Average) model is applied. While the unique parts are same with the ARIMA model, the Mean Squared Error (MSE) of the test set, $575,286$ is higher the previous result, $448,704$. Therefore, only result from the ARIMA model will be presented in summary tables.

The Multivariate linear regression model is applied to 6,268 parts, with 2,249 parts excluded due to insufficient data. The Mean Squared Error (MSE) of the test set is $394,4925$, reflecting the model's ability to effectively capture relationships between historical demand, installed quantities, and shipment quantities. Non-Negative Least Squares is used to ensure that all coefficients are non-negative and aligned with real-world demand patterns.

The Multivariate Nonlinear Regression with Polynomial Features model combines the flexibility of polynomial transformations with non-negativity constraints. The model uses a polynomial degree of 2 to capture higher-order and interaction effects among features. It is applied to 5,926 unique parts, with 2,591 parts excluded due to insufficient data. The Mean Squared Error (MSE) of the test set is $323,827$.

The Support Vector Regression (SVR) model with Radial Basis Function (RBF) kernel is implemented to forecast shipment quantities for each part. The model employs grid search for hyperparameter tuning to optimize the model's performance. The model's key parameters include the regularization parameter (C), with tested values of 0.01, 0.1, 1, and 10, and the epsilon parameter (ϵ), with tested values of 0.5, 1.0, 5.0, and 10. A log transformation ($\log(1 + x)$) is applied to the target variable to ensure non-negative predictions. After the model is trained, predictions are transformed back to the original scale using $\exp(x) - 1$. The SVR model is applied to 6,111 unique parts, with 2,406 parts excluded due to insufficient data.

The Mean Squared Error (MSE) of the test set is $505,829$.

The Artificial Neural Network (ANN) model, implemented using the Multi-Layer Perceptron Regressor (MLPRegressor), is configured with two hidden layers containing 64 and 32 neurons, respectively, and uses the ReLU (Rectified Linear Unit) activation function. The Adam optimizer is employed for gradient-based optimization with a learning rate of 0.001, ensuring efficient training. The model is trained for a maximum of 500 iterations to achieve convergence. Before training, features are standardized using StandardScaler. The negative predictions are clipped to zero to maintain non-negativity, reflecting the nature of shipment quantities. The model is applied to 6,309 unique parts, with 2,208 parts excluded due to insufficient data or errors during processing. The Mean Squared Error (MSE) of the test set is $431,131$.

The Random Tree model is implemented using the DecisionTreeRegressor. The splitter parameter is set as "random" to improve generalization and reduce overfitting. The model is allowed to grow fully by setting "max_depth" to "None" and captures intricate patterns in the data. To ensure reproducibility, the "random_state" parameter is fixed at 42. The model is applied to 6,309 unique parts, with 2,208 parts excluded due to insufficient data. The Mean Squared Error (MSE) of the test set is $455,197$.

The Reduced Error Pruning Tree (REPTree) model is also implemented using the DecisionTreeRegressor. The model applies cost-complexity pruning to prevent overfitting and improve generalization. Cost-complexity pruning parameter, ccp_alpha, set to 0.01 to balance the tree's complexity and its ability to fit the data. The minimum number of samples, min_samples_split, required to split an internal node is set to 5 . The "max_depth" parameter is set to "None", allowing the tree to grow fully within the constraints of pruning. To ensure reproducibility, the "random_state" parameter is fixed at 42. The model is applied to 6,309 unique parts, with 2,208 parts excluded due to insufficient data. The Mean Squared Error (MSE) of the test set is $448,315$.

The Random Forest model, implemented using the RandomForestRegressor, is configured with 100 decision trees (n_estimators=100) to aggregate predictions from multiple trees, which provide lower variance and higher accuracy. The "max_depth"

parameter is set to "None", enabling each tree to grow fully and capture detailed patterns in the data. The "random_state" parameter is fixed at 42 for the reproducibility. The model is applied to 6,309 unique parts, with 2,208 parts excluded due to insufficient data. The Mean Squared Error (MSE) of the test set is $446,403$.

The XGBoost model is implemented using the XGBRegressor and configured with 100 estimators (n_estimators=100). A maximum depth of each tree is set to 6 (max_depth=6). To balance convergence speed and prediction accuracy, the learning rate is set to 0.1 (learning_rate=0.1). Higher generalization and reduced overfitting are ensured through subsampling of the training data (subsample=0.8) and columns (colsample_bytree=0.8) during training. The (random_state) is set to 42. The target variable is log-transformed using $\log(1+y)$ to have non-zero target values. Predictions generated in the log-transformed scale are then transformed back to the original scale using $\exp(x)-1$. The model is applied to 6,309 unique parts, with 2,208 parts excluded due to insufficient data. The Mean Squared Error (MSE) of the test set is $461,856$.

The LightGBM model is implemented using the LGBMRegressor. It is configured with 100 boosting iterations (n_estimators=100) and a learning rate of 0.1 (learning_rate=0.1) to minimize errors. The model employs the gradient boosting decision tree algorithm (boosting_type='gbdt') with no restrictions on tree depth (max_depth=-1). The parameter (min_data_in_leaf=5) ensures that each leaf node contains at least five samples, and (min_data_in_bin=1) allows finer-grained data splits. The (force_row_wise=True) parameter enforces row-wise threading for efficiency, while (verbose=-1) suppresses unnecessary output for cleaner processing. A log transformation ($\log(1+y)$) is applied to the target variable to ensure non-zero results. Predictions are generated in the log-transformed scale and transformed back to the original scale using $\exp(x)-1$. The model is applied to 6,309 unique parts, with 2,208 parts excluded due to insufficient data or errors during processing. The Mean Squared Error (MSE) of the test set is $441,382$.

The performance of the forecasting models is summarized in Tables 5.2 and 5.3. As shown in Table 5.2, ARIMA, Holt-Winters, and MNLR are the least inclusive models as they incorporate fewer unique SKUs. The R^2 values are generally low, as

expected, due to the intermittency of the data. Among the models, MLR, MNLR, and ANN exhibit the best Test R^2 values, while SBA, SVR, ARIMA and Naïve show lower performance in terms of Test R^2. As shown in Table 5.3, MNLR, MLR, and ANN achieve the best MSE results, whereas ARIMA, SBA, and SVR demonstrate the worst performance. For MAE, LightGBM, XGBoost, and SVR yield the best results, while ANN, MNLR, and ARIMA perform the worst.

To ensure inclusivity and explainability, MLR, MNLR, ANN, Random Tree, REP Tree, Random Forest, XGBoost, and LightGBM are selected for inventory management optimization. Conversely, models such as Naïve, Holt-Winters, SBA, ARIMA, and SVR are excluded due to their limited inclusivity (less than 5000 unique SKUs) and / or insufficient R^2 values (below 0.1).

Table 5.3: Model Performance Summary - Part 1

Model	**Unique Parts in Test**	**Skipped Parts in Test**	**Train R^2**	**Test R^2**	**Test MSE Mean**	**Test MSE Max**
Naïve forecast	6,309	2,208	-0.28	0.08	390,919	784,002
Holt-Winters' seasonal method	4,701	3,816	0.44	0.16	362,717	549,357
Croston modification: SBA	6,309	2,208	-0.98	0.00	342,784	745,565
ARIMA	4,077	2,208	0.10	0.08	448,704	642,883
MLR	6,268	2,249	0.34	0.25	278,290	475,764
MNLR	5,926	2,591	0.42	0.18	251,712	490,676
SVR	6,111	2,406	0.12	0.04	345,101	692,806
ANN	6,309	2,208	0.37	0.18	332,827	566,814
Random Tree	6,309	2,208	0.99	0.13	311,977	577,863
REP Tree	6,309	2,208	0.71	0.14	307,490	577,883
Random Forest Regressor	6,309	2,208	0.89	0.15	304,787	572,874
XGBoost	6,309	2,208	0.91	0.12	312,244	612,276
LightGBM	6,309	2,208	0.27	0.16	301,722	544,380

Table 5.4: Model Performance Summary - Part 2

Model	**Train MSE**	**Test MSE**	**Train RMSE**	**Test RMSE**	**Train MAE**	**Test MAE**
Naïve forecast	19,432	479,624	139.40	692.55	8.88	20.40
Holt-Winters' seasonal method	8,873	458,563	94.20	677.17	7.18	19.70
Croston modification: SBA	30,068	521,477	173.40	722.13	12.06	19.20
ARIMA	16,378	575,286	127.98	758.48	8.22	21.21
MLR	10,115	394,492	100.57	628.09	7.93	20.76
MNLR	5,074	323,827	71.24	569.06	6.17	21.23
SVR	13,447	505,828	115.96	711.22	7.35	18.31
ANN	9,572	431,131	97.84	656.61	6.60	21.41
Random Tree	197	455,196	4.44	674.68	0.03	20.62
REP Tree	4,293	448,314	65.52	669.56	3.79	20.27
Random Forest Regressor	1,732	446,402	41.62	668.13	2.90	19.88
XGBoost	1,330	461,859	36.47	679.60	1.46	18.26
LightGBM	11,064	441,381	105.19	664.37	4.16	17.52

5.5 Implementation of Inventory Management

The data provided by the case company includes the part cost, the part price, the penalty / goodwill cost, and the holding cost ratio. The optimal service level for each product is calculated using the Newsvendor model as $\frac{c_u}{c_u+c_o}$, where c_u donates the underage cost per part and c_o signifies the overage cost per part. The underage cost, c_u, is the sum of the lost profit and the penalty/goodwill cost, as shown in Equation 5.1. Lost profit is calculated by subtracting the cost of the spare part from its price. In cases of spare part shortages, the company is responsible for providing a replacement product, and the price of the replacement product defined as the penalty/goodwill cost. The overage cost, c_o, is determined by the holding cost, and derived by multiplying the holding cost ratio by the part cost, as shown in Equation 5.2.

$$c_u = part\,price - part\,cost + goodwill\,cost\,(replacement\,product\,price) \quad (5.1)$$

$$c_o = holding\,cost\,ratio * part\,cost \quad (5.2)$$

The target service level of the department is set to 95% by the company to ensure customer satisfaction.

The overage and underage costs for each forecasting model are calculated using the discrepancies between forecasted and actual demand for test set, which is 20% of total data. The underage cost of model reflects the penalty associated with unmet demand and is computed by multiplying the per-unit underage cost with the shortage quantity. The overage cost per model accounts for excess inventory and is calculated by multiplying the per-unit overage cost with the surplus quantity. The total cost of the model is the sum of the underage and overage cost of the relative model, and the service level of the model is computed by dividing the underage cost by the total cost.

Overage cost, underage cost, total cost, and service level of each model are illustrated in Table 5.5. Among the eight models, MLR achieves the lowest total cost, while LightGBM demonstrates the highest service level. However, the target

service level is not met by any of these models.

Table 5.5: Overage and Underage Costs, Total Cost, and Service Level for Demand Forecasting Models

Model	**Overage Cost**	**Underage Cost**	**Total Cost**	**Service Level**
MLR	45,114,930	263,924,900	309,039,830	85.4%
MNLR	88,581,160	224,359,700	312,940,860	71.7%
ANN	63,635,350	257,353,000	320,970,350	80.2%
RT	48,398,960	279,261,400	327,660,360	85.2%
REPTree	44,360,190	284,204,500	328,564,690	86.5%
RF	44,519,400	253,348,600	297,868,000	85.1%
XGBoost	26,884,560	297,962,500	324,847,060	91.7%
LightGBM	19,009,620	309,392,100	328,401,720	94.2%

Huber et al. (2019) conducted inventory optimization using point forecasts and forecast errors. They applied model-based optimization and data-driven optimization with Sample Average Approximation, as discussed in Chapters 2 and 4. Inventory optimization method of Huber et al. (2019) is applied to improve service levels and meet the company's target by accounting for each parts cost-optimal service level. Since model-based inventory optimization requires a specific forecast error distribution, the normality of residuals for demand forecasting models is evaluated. The distributions are assessed using statistical tests (Shapiro-Wilk and Kolmogorov-Smirnov) and visualized through histograms and Q-Q plots. Table 5.6 shows that none of the models' residuals pass the normality tests, with all p-values from the Shapiro-Wilk and Kolmogorov-Smirnov tests being 0. This result indicates a significant deviation from normality in the residuals of all models.

Figure 5.10 shows the histogram and Q-Q plot for the residuals of the MLR model as a representative example among the eight models. The histogram reveals a strong concentration of residuals around zero, while the Q-Q plot highlights substantial deviations relative to the theoretical normal distribution, particularly within the tails. This results indicate that the residuals of the MLR model, similar to those

Table 5.6: Normality Test Results for Residuals

Model	**Shapiro-Wilk (Statistic, p-value)**	**Kolmogorov-Smirnov (Statistic, p-value)**
MLR	(0.013, 0.0)	(0.461, 0.0)
MNLR	(0.0159, 0.0)	(0.4586, 0.0)
ANN	(0.0124, 0.0)	(0.4626, 0.0)
RT	(0.0111, 0.0)	(0.4632, 0.0)
REPTree	(0.011, 0.0)	(0.4638, 0.0)
RF	(0.0105, 0.0)	(0.465, 0.0)
XGBoost	(0.0089, 0.0)	(0.4642, 0.0)
LightGBM	(0.0087, 0.0)	(0.4684, 0.0)

of the other models, show significant non-normality. Therefore, model-based optimization is not suitable in this case. Instead, data-driven optimization with Sample Average Approximation will be applied, leveraging the error distribution based on the empirical forecast errors.

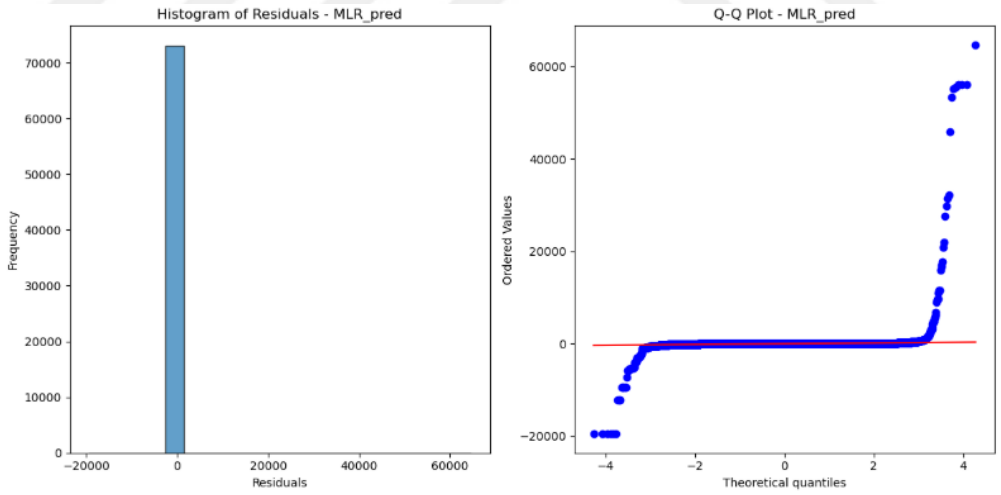

Figure 5.14: Histogram of Residuals and Q-Q Plot

To implement data-driven optimization with Sample Average Approximation to the data set, optimized inventory levels are calculated for demand forecasting models and individual parts. For each part and model, the residuals (differences between actual and predicted values) are analyzed to derive an empirical cumulative distribution function (CDF). The optimum order quantity, q^*, is sum of the forecasted

values and the respective quantile values of the corresponding model as shown in Equation 5.3 and described by Huber et al. (2019). The quantile values are obtained from the empirical cumulative distribution function (CDF) based on the residuals.

The underage and overage costs are calculated by multiplying the relative cost with the difference between the observed and the optimum order quantity, q^*. The total cost for each model is the sum of underage and overage costs of the test set, while the service level is determined as the ratio of the underage cost to the total cost, consistent with the original calculation method.

$$q^*(\mathbf{x}) = \hat{y}(\mathbf{x}) + \inf\left\{p : \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(\epsilon_i \leq p) \geq \frac{c_u}{c_u + c_o}\right\}. \tag{5.3}$$

Table 5.7: Overage and Underage Costs, Total Cost, and Service Level after Inventory Optimization for Demand Forecasting Models

Model	**Overage Cost**	**Underage Cost**	**Total Cost**	**Service Level**
MLR	61,990,020	1,589,342,000	1,651,332,000	96.25%
MNLR	52,769,820	1,575,008,000	1,803,477,000	97.07%
ANN	58,815,380	1,584,457,000	1,643,272,000	96.42%
RT	57,310,150	1,616,678,000	1,673,989,000	96.57%
REPTree	58,023,520	1,639,921,000	1,697,945,000	95.58%
RF	58,509,140	1,576,753,000	1,635,613,000	96.41%
XGBoost	58,004,050	1,572,653,000	1,630,657,000	96.44%
LightGBM	60,030,810	1,582,517,000	1,642,880,000	96.32%

Overage cost, underage cost, total cost, and service level of each model after the inventory optimization are illustrated in Table 5.7. Among the eight models, XGBoost achieved the lowest total cost, while MNLR demonstrated the highest service level. However, all models met the target serice level, with minor deviations.

Table 5.8 provides a summary of the forecast accuracy metrics, including MSE, RMSE, and MAE from Table 5.4, as well as inventory performance metrics such as total cost and service level, both before optimization (from Table 5.5) and after

Table 5.8: Model Performances and Costs Before and After Inventory Optimization

	Before the Optimization		**After the Optimization**				
Model	**Total Cost**	**Service Level**	**Total Cost**	**Service Level**	**MSE**	**RMSE**	**MAE**
MLR	309,099,830	85.40%	1,651,332,000	96.25%	394,495.22	628.09	20.76
MNLR	312,940,860	71.64%	1,803,477,000	97.07%	323,827.07	569.06	21.23
ANN	320,970,350	80.17%	1,643,272,000	96.42%	431,131.19	656.60	21.41
RT	327,660,360	85.22%	1,673,989,000	96.57%	455,916.97	674.68	20.62
REPTree	328,564,690	86.48%	1,697,945,000	95.82%	448,314.79	669.56	20.26
RF	297,868,000	85.05%	1,635,613,000	96.41%	446,402.72	668.13	19.87
XGBoost	324,847,060	91.72%	1,630,657,000	96.44%	461,855.88	679.60	18.26
LightGBM	328,401,720	94.21%	1,648,208,000	96.32%	441,381.63	664.37	17.52

optimization (from Table 5.7), for the selected eight models. While MNLR achieved the lowest MSE, it resulted in the lowest service level before optimization and the highest total cost after optimization. This result highlights that traditional performance assesment indicators, such as MSE, are insufficient to determining the best model for the case company. Additionally, the model with the highest service level before optimization did not yield the lowest total cost after optimization, since the optimization process accounted for the optimal service levels of individual parts.

Although LightGBM achieved the highest service level before optimization, XGBoost, Random Forest, and ANN meet the target service level of 95% with lower costs. Ultimately, XGBoost emerged as the model that achieved the target service level with the lowest total cost, despite being outperformed by other models in terms of MSE. Furthermore, XGBoost also outperformed all models except LightGBM in terms of MAE, and showed its suitability for the company's objectives.

Additionally, Table 5.9 illustrates the standard deviations of the test set before and after optimization. The results demonstrate the standard deviation of the q values, which is the sum of y_prediction and the related quantile values of optimization, is better than standard deviation of y_predicted values before the optimization.

This thesis aims to determine the demand forecasting method that best suits the case company by defining demand patterns and suggest the most appropriate inventory strategy, including ordering and stocking quantities. The recommended inventory strategy for the company is (R, S), aligning with the preference for monthly reviews, where the optimal order quantity, q^*, derived from the XGBoost model,

Table 5.9: Standard Deviation of Each Model Before and After Optimization in the Test Set

Model	**Before the Optimization** ($\hat{y}(\mathbf{x})$)	**After the Optimization** ($q(\mathbf{x})$)
MLR	316.8	195.9
MNLR	425.6	271.8
ANN	219.0	81.8
RT	190.2	166.9
REPTree	177.7	142.8
RF	166.5	107.2
XGBoost	149.1	141.0
LightGBM	155.1	147.3

is used as S, the order-up-to level . Since the suggested strategy meets the target service level by adding a quantile to the point forecast, the implementation of safety stock is deemed unnecessary.

It is important to highlight that a comprehensive application was implemented taking into account the company's target service level threshold of 95%. Figure 5.15 shows the sensitivity analysis of service level and the total cost. Notably, the highest service level achieved before optimization was 94.21%, obtained using LightGBM, with a total cost of 328,401,720. After optimization, the lowest total cost required to achieve the target service level is 1,630,657,000, obtained using XGBoost. Adjusting the service level threshold, which is given by a target from the company, could be considered due to the difference in total cost before and after the optimization.

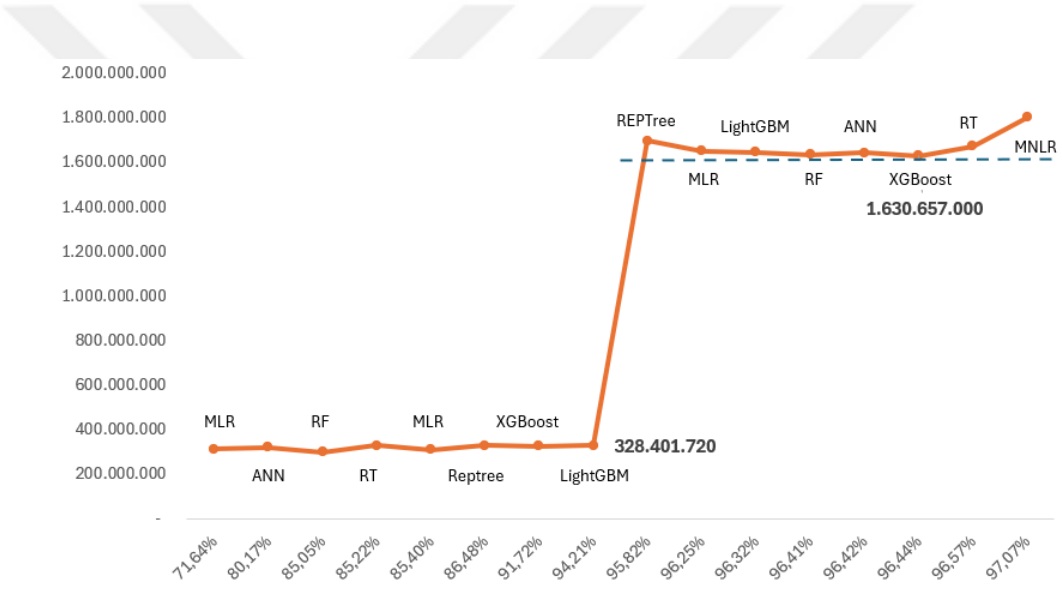

Figure 5.15: Sensitivity Analysis of Service Level and Total Cost

Chapter 6

CONCLUSIONS

This study addressed critical challenges in demand forecasting and inventory management of spare parts for the purpose of reducing the risks of overstocking and stock-outs. Demand classification and the identification of the most effective forecasting techniques were specifically examined for the case company. The study also explored the most effective inventory strategies, addressing key questions such as how many spare parts to order and which items to stock to ensure efficient operations. The comprehensive approach suggested by Bacchetti and Saccani (2012) was applied, combining spare parts classification, demand forecasting, inventory management, and performance assessment. Additionally, the study aligned with the forecasting framework proposed by Boylan and Syntetos (2010) and implemented pre-processing, processing, and post-processing phases of forecasting. Demand classification was conducted as part of the pre-processing phase, followed by the implementation of demand forecasting methods during the processing phase. Finally, optimum order quantities, q^*, were derived from the point predictions of the forecasting phase. Thus, the post-processing of the systematic approach was completed by applying inventory optimization.

Firstly, the demands were categorized across four categories: Erratic, Intermittent, Lumpy, and Smooth. On yearly average, 76.17% of the demands were identified as Intermittent, followed by 19.39% as Lumpy. These results demonstrate that the demand patterns are highly sporadic and traditional demand forecasting methods are insufficient as emphasized in the literature review in Chapter 2.

The dataset includes year, month, and total shipment quantity as baseline features. The original dataset was further extended to include total shipment quantities from the previous year, total shipment quantities from the previous two years, total installation quantities from the previous year, total installation quantities from the

previous two years, and encoded demand classification.

Secondly, thirteen demand forecasting models, covering a range of traditional, statistical, and machine learning approaches, were implemented. The Naïve Forecast, Holt-Winters' Seasonal Method, and SBA are commonly used for basic and intermittent demand forecasting. As a statistical time series model, ARIMA (Auto-Regressive Integrated Moving Average) was employed. Multivariate models (MLR and MNLR) were employed to capture relationships between multiple features. Machine learning models, (SVR and ANN), as well as tree-based models (Random Tree and Reduced Error Pruning Tree), were also evaluated. Additionally, advanced gradient boosting algorithms, (XGBoost and LightGBM), were implemented. All models were systematically compared based on R^2, the unique number of parts, MSE, and MAE. Traditional models (Naïve Forecast, Holt-Winters' Seasonal Method, SBA), along with ARIMA and SVR, were excluded from further analysis due to their low inclusivity and poor R^2 performance.

As a post-processing step in forecasting, the service level of the newsvendor model was utilized for inventory cost calculations, and a data-driven approach employing Sample Average Approximation was implemented following Huber et al. (2019). The total costs and optimum service levels of each model were compared and evaluated against the target service level defined by the case company. Key findings reveal that traditional performance metrics, such as MSE, are inadequate for identifying the most suitable model, as they fail to account for service level performance and total costs. Notably, models with the highest service levels before the optimization do not always result in the lowest total costs after optimization. This finding shows the importance of applying part-specific optimal service levels. The results before and after optimization also highlight the trade-off between achieving minimizing total costs and target service levels. Another finding shows that models excelling in MAE are closer to meeting target service levels with lower financial impacts.

Based on the findings and analyses presented in this study, a tailored inventory strategy has been developed to address the specific needs of the company. The (R, S) policy, aligning with the preference for monthly reviews, is proposed as the inventory strategy for the company. Within this framework, the optimal order quantity, q^*,

derived from the predictions of the XGBoost model and the quantile of Newsvendor optimal service level, is utilized as the order-up-to level S.

This thesis contributes to the literature by presenting a comprehensive approach that integrates demand forecasting through demand classification and inventory optimization using the Newsvendor model, offering a practical framework for researchers. The application of this framework in the air conditioner production sector underscores its practicality and suitability for industry-specific implementations. This study also emphasizes the innovative use of demand classification as an feature within forecasting models, rather than exclusively applying forecasting models to specific demand categories.

The final discussion is the trade-off between achieving targeted service levels and the corresponding increase in total costs after optimization. While optimization effectively aligns service levels with company targets, it presents the challenge of balancing operational efficiency with financial sustainability. Future research could focus on developing methods to mitigate the cost impact of reaching target service levels, such as exploring adaptive service level thresholds. Furthermore, examining the relationship between demand variability, service levels, and corresponding costs could offer valuable insights for creating more flexible and cost-efficient inventory strategies.

BIBLIOGRAPHY

Aktepe, A., Yanık, E., and Ersöz, S. (2021). Demand forecasting application with regression and artificial intelligence methods in a construction machinery company. *Journal of Intelligent Manufacturing*, 32:1587–1604.

Alalawin, A. and Arabiyat, L. (2021). Forecasting vehicle's spare parts price and demand. *Journal of Quality in Maintenance Engineering*, 27:483–499.

Alduailej, M., Khan, Q. W., Tahir, M., Sardaraz, M., and Malik, F. (2022). Machine-learning-based ddos attack detection using mutual information and random forest feature importance method. *Symmetry*, 14(6):1095.

Altay, N., Litteral, L. A., and Rudisill, F. (2012). Effects of correlation on intermittent demand forecasting and stock control. *International Journal of Production Economics*, 135:275–283.

Altay, N., Rudisill, F., and Litteral, L. (2008). Adapting wright's modification of holt's method to forecasting intermittent demand. *International Journal of Production Economics*, 111:389–408.

Amirkolaii, K., Baboli, A., Shahzad, M., and Tonadre, R. (2017). Demand forecasting for irregular demands in business aircraft spare parts supply chains by using artificial intelligence (ai). *IFAC PapersOnLine*, 50(1):15221–15226.

Aryudha, G. and Hasibuan, W. (2024). Linear regression analysis in predicting the amount of stock of hp sparepart goods in gmt. *Journal of Artificial Intelligence and Engineering Applications*, 4(1):550–556.

Atakay, B., Obaşılı, O., Özcet, S., Akbulak, , Cevher, H., Alcaz, H., Gördesli, , Paldrak, M., and Staiou, E. (2022). Spare parts inventory management system in a service sector company. In *Digitizing Production Systems*.

Bacchetti, A. and Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, 40:722–737.

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

Boylan, J., Syntetos, A., and Karakostas, G. (2008). Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, 59:473–481.

Chatfield, C. and Yar, M. (1988). Holt-winters forecasting: Some practical issues. *Statistician*, 37:129–140.

Chien, C., Ku, C., and Lu, Y. (2023). Ensemble learning for demand forecast of after-market spare parts to empower data-driven value chain and an empirical study. *Computers Industrial Engineering*, 185(109670).

Deng, T., Zhao, Y., Wang, S., and Yu, H. (2021). Sales forecasting based on lightgbm. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 1–5.

der Auweraer, S. V., Boute, R. N., and Syntetos, A. A. (2019). Forecasting spare part demand with installed base information: A review. *International Journal of Forecasting*, 35:181–196.

do Rego, J. R. and de Mesquita, M. A. (2014). Demand forecasting and inventory control: A simulation study on automotive spare parts. *International Journal of Production Economics*, 161:1–16.

Eaves, A. H. C. and Kingsman, B. G. (2004). Forecasting for the ordering and stockholding of spare parts. *Journal of the Operational Research Society*, 55(4):431–437.

Hahn, G. J. and Leucht, A. (2015). Managing inventory systems of slow-moving items. *International Journal of Production Economics*, 170:543–550.

Hasni, M., Aguir, M., Babai, M., and Jemai, Z. (2018). Spare parts demand forecasting: a review on bootstrapping methods. *International Journal of Production Research*, 57(15-16):4791–4804.

Huber, J., Müller, S., Fleischmann, M., and Stuckenschmidt, H. (2019). A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research*, 278(3):904–915.

Iranzad, R. and Liu, X. (2024). A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*. Received: 7 November 2022 / Accepted: 6 January 2024.

Jiang, A., Tam, K., and X. Guo, Y. Z. (2019). A new approach to forecasting intermittent demand based on the mixed zero-truncated poisson model. *Journal of Forecasting*, 38(7):579–594.

Kannan, B. A., Kodi, G., Padilla, O., Gray, D., and Smith, B. C. (2020). Forecasting spare parts sporadic demand using traditional methods and machine learning - a comparative study. *SMU Data Science Review*, 3(2).

Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *International Journal of Production Economics*, 143:198–206.

Li, Y. and Wu, H. (2012). A clustering method based on k-means algorithm. *Physics Procedia*, 25:1104–1109. 2012 International Conference on Solid State Devices and Materials Science.

Muniz, L. R., Conceição, S. V., Rodrigues, L. F., de Freitas Almeida, J. F., and Affonso, T. B. (2021). Spare parts inventory management: a new hybrid approach. *The International Journal of Logistics Management*, 32(1):40–67.

Natarajan, K., Sim, M., and Uichanco, J. (2017). Asymmetry and ambiguity in newsvendor models. *Management Science*, 64(7):3146–3167.

Oroojlooyjadid, A., Snyder, L. V., and Takáč, M. (2020). Applying deep learning to the newsvendor problem. *IISE Transactions*, 52(4):444–463.

Panda, S. K. and Mohanty, S. N. (2023). Time series forecasting and modeling of food demand supply chain based on regressors analysis. *IEEE Access*, 11:1–. Published April 11, 2023, Open Access.

Pinçe, , Turrini, L., and Meissner, J. (2021). Intermittent demand forecasting for spare parts: A critical review. *Omega*, 105(102513).

Riwayadi, E., Suprajitno, H., and Miswanto (2024). Enhancing spare parts management through support vector regression: A case study in the service and maintenance industry. *Moneter: Jurnal Keuangan Dan Perbankan*, 12(2):207–217.

Romeijnders, W., Teunter, R., and van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, 220(3):386–393.

Sani, B. and Kingsman, B. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48(7):700–713.

Snyder, R. D., Ord, J. K., and Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, 28(2):485–496.

Soofi, A. A. and Awan, A. (2017). Classification techniques in machine learning: Applications and issues. *Journal of Basic & Applied Sciences*, 13:459–465.

Syntetos, A. A., Babai, M. Z., and Jr., E. S. G. (2015). Forecasting intermittent inventory demands: Simple parametric methods vs. bootstrapping. *Journal of Business Research*, 68(8):1746–1752.

Syntetos, A. A. and Boylan, J. E. (2010). On the variance of intermittent demand estimates. *International Journal of Production Economics*, 128(2):546–555.

Teunter, R. H. and Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, 60(3):321–329.

Tian, X., Wang, H., and Erjiang, E. (2021). Forecasting intermittent demand for inventory management by retailers: A new approach. *Journal of Retailing and Consumer Services*, 62:102662.

Vandeput, N. (2020). *Inventory Optimization: Models and Simulations*. De Gruyter, Berlin/Boston.

Vargas, C. G. and Cortés, M. E. (2017). Automobile spare-parts forecasting: A comparative study of time series methods. *International Journal of Automotive and Mechanical Engineering*, 14(1):3898–3912.

Wang, S., Kang, Y., and Petropoulos, F. (2024). Combining probabilistic forecasts of intermittent demand. *European Journal of Operational Research*, 315:1038–1048.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.

Xu, Q., Wang, N., and Shi, H. (2012). Review of croston's method for intermittent demand forecasting. In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, pages 1456–1460.

Zhang, S., Huang, K., and Yuan, Y. (2021). Spare parts inventory management: A literature review. *Sustainability*, 13(2460).

Zhou, C. and Viswanathan, S. (2011). Comparison of a new bootstrapping method with parametric approaches for safety stock determination in service parts inventory systems. *International Journal of Production Economics*, 133(2):481–485.

Zhuang, X., Yu, Y., and Chen, A. (2022). A combined forecasting method for intermittent demand using the automotive aftermarket data. *Data Science and Management*, 5:43–56.

İfraz, M., Aktepeb, A., Ersöz, S., and Çetinyokuş, T. (2023). Demand forecasting of spare parts with regression and machine learning methods: Application in a bus fleet. *Journal of Engineering Research*, 11(100057).