Predicting semiconductor component lead-time for an oil and gas company:
A dynamic safety stock model with machine learning

by

Chong Zhang
Master of Industrial Engineer, Tokyo Institute of Technology 2009

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Signature of Author: ____________________
Department of Supply Chain Management
May 10, 2024

Certified by: ____________________
Dr.Thomas Koch
Postdoctoral Associate, Center for Transportation and Logistics
Capstone Advisor

Accepted by: ____________________
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Predicting semiconductor component lead-time for an oil and gas company:

A dynamic safety stock model with machine learning

by

Chong Zhang

Submitted to the Program in Supply Chain Management
on May 10, 2024, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

In the oil and gas industry, the fluctuations of semiconductor component delivery time significantly impact Printed Circuit Board Assembly (PCBA) production planning. The discrepancies between the suppliers' quoted lead-time and actual delivery lead-time present a substantial challenge, as the sponsor company`s MRP system lacks a robust safety stock model to mitigate component shortages. To enhance the predictability of both the delivery lead-time and the standard deviation of lead-time, this project introduces a machine learning-based framework. To mitigate semiconductor component shortages, this project also introduces a dynamic safety stock model, which incorporates the predicted delivery lead-time and standard deviation of lead-time. Through a comparative analysis of model performance, the tree model emerged as the most effective in predicting delivery lead-time. The dynamic safety stock model also demonstrated improvements in inventory management and production planning. The average waiting time, which was caused by component shortages, was reduced by more than 50% when compared with static safety stock model. These improvements significantly reduce the risk associated with semiconductor supply chain variability and strengthen the company`s operational resilience.

Capstone Advisor: Dr.Thomas Koch
Title: Postdoctoral Associate, Center for Transportation and Logistics

ACKNOWLEDGMENTS

I am grateful to Dr. Elenna R. Dugundji and Dr. Thomas Koch for their support, guidance, and kindness throughout the project. Their deep insights into data science and machine learning, combined with their extensive academic and industrial experience, have been invaluable to the success of this project.

I would also like to thank Toby Gooley and Dr. Chris Featherman who are CTL and WCC Lecturers. Their expertise in professional writing has greatly enhanced my English skills and helped rectify numerous writing errors.

Additionally, I appreciate Philippe Bansept and Goh MingRui Marie from the sponsor company. They provided the necessary data, offered insightful suggestions, and were always available to answer my questions during the project.

I am also thankful to my MIT CTL classmates for their friendship and support. Completing the capstone project alongside this cohort has made it a memorable and meaningful experience.

Finally, I extend my thanks to my family and friends for their endless love and support.

TABLE OF CONTENTS

1. INTRODUCTION

Printed Circuit Board Assembly (PCBA) is the backbone of an application which integrates various mechanical and electronic components onto a single bare board to create a highly functional system. The semiconductor components, which include integrated circuits (IC), transistors, resistors, and capacitors, are critical electronic components within the assembly. They fundamentally define and achieve the specified functions, capabilities, and reliability of the final application, which ranges from simple consumer products to complex systems used in automotive, aerospace, and particularly the oil and gas sector.

One of the major supply chain challenges in the production of PCBA is the stretching of semiconductor lead-time. This trend began around 2017, driven by increasing demand particularly from automotive and industrial applications. Since then, lead-time has risen significantly, with some devices experiencing a jump from a typical lead-time of 8 weeks to as long as 24-30 weeks (Roos 2017). Lead-time refers to the period from when a purchase order is placed to when the components are received at the factory. It is crucial for manufacturers to forecast lead-time to efficiently plan and execute production schedules. The components shortages due to extended lead-time stretching can cause significant production delays and increase overall production costs.

1.1 Motivation

The sponsor company is a multinational oilfield services company headquartered in Houston, Texas, United States. It designs, produces, and provides a wide range of products and services for the oil and gas industry, serving both downhole and surface systems. The sponsor company's business covers various stages of the oilfield lifecycle, including exploration, drilling, reservoir characterization, production, and reservoir management. The company serves a diverse range of customers, including national oil companies and independent operators. The sponsor company is dedicated to delivering innovative technologies that help customers optimize their exploration and production activities while addressing the world's energy challenges.

The company currently conducts business in over 120 countries. Its products and spare parts are designed, produced, and assembled mainly at 25 manufacturing centers worldwide, then shipped to distribution centers or field locations for oil and gas services.

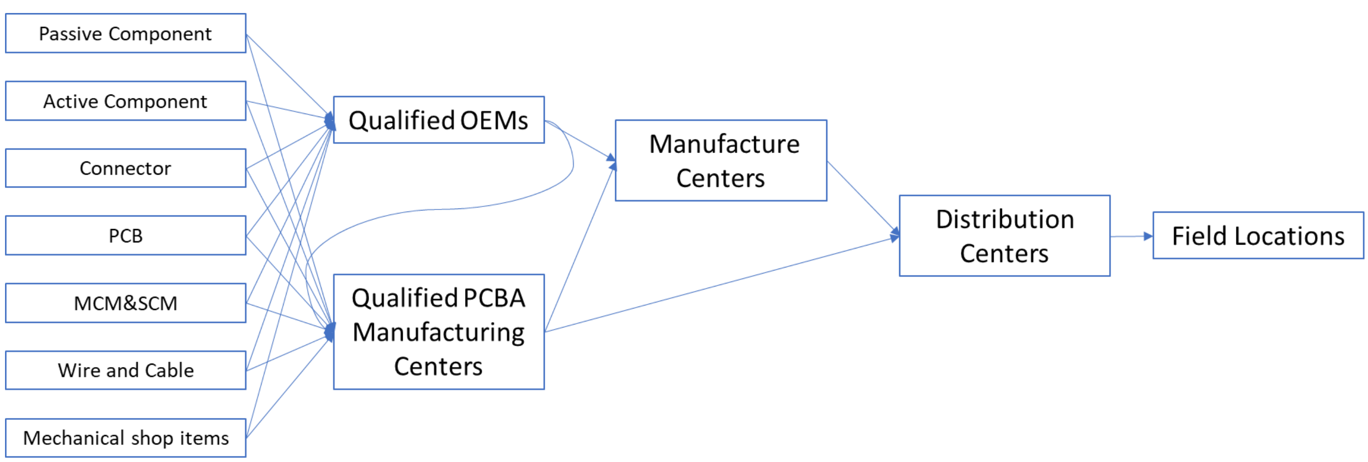
1.2 Background- Sponsor Company's Supply Chain for PCBA

The unpredictable nature of oil and gas exploration and production makes it difficult to forecast demand for the sponsor company's products and spare parts. This uncertainty is a challenge for the supply chain supporting the Printed Circuit Board Assembly (PCBA) commodity used in the sponsor company's products. The PCBAs include active components, such as transistors, integrated circuits, and op-amps, as well as passive electronic components, such as resistors, capacitors, and diodes.

The sponsor company's PCBAs, particularly those used in downhole systems, are qualified for highly reliable service in high-temperature and high-pressure environments. The manufacturers responsible for assembling these PCBAs must meet the IPC-A-610 Class 3 standard, which is specifically intended for PCBAs with stringent reliability requirements, such as those used in aerospace, medical, and military systems. Due to this high reliability requirement, the sponsor company has a limited number of in-house manufacturing centers and external original equipment manufacturers (OEMs) that are qualified for assembling PCBAs used in downhole systems (Figure 1).

Figure 1

The Sponsor Company's PCBA Supply Chain Model

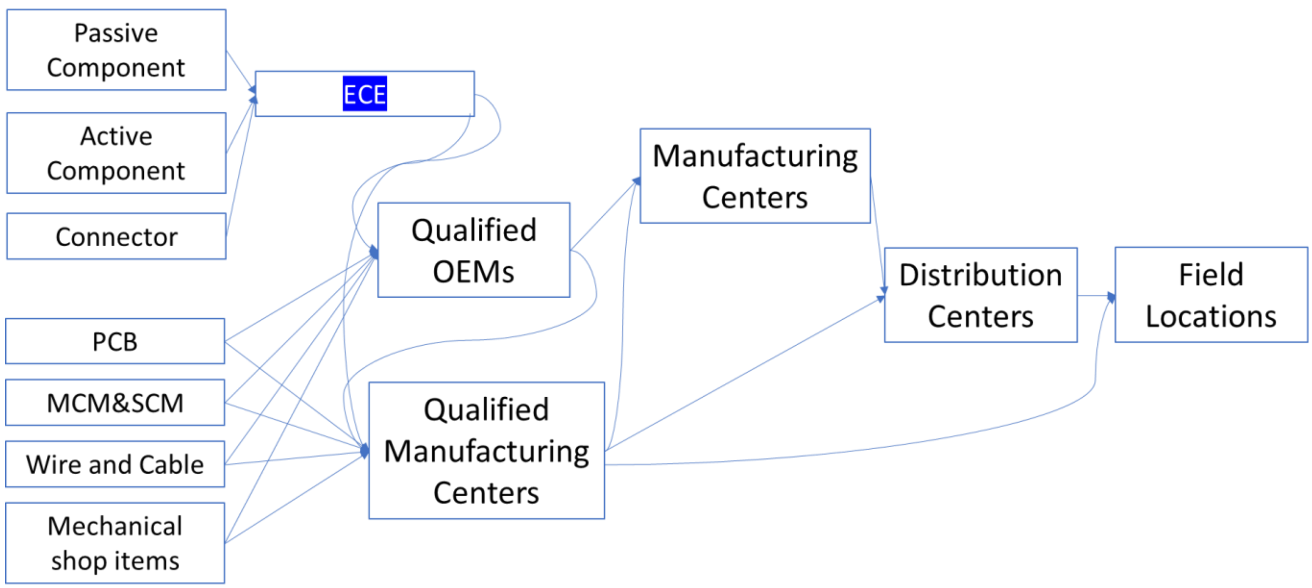


In the sponsor company's current supply chain for the PCBA commodity, each manufacturing center and external OEM is procuring electronic components independently, with limited cooperation among them. Moreover, each of the sponsor company's manufacturing centers generates a different level of demand for the passive/active components. This contributes to the demand variability for those products, which, in turn, affects inventory costs and availability.

One of the most powerful tools for addressing demand variability is the concept of risk pooling. By aggregating and managing the common components for different PCBA items at a centralized location, the sponsor company can reduce demand variability and lower supply chain costs without hurting product availability. Based on the risk pooling concept, the sponsor company's Electronics Center of Excellence (ECE) established the Stocking Strategy Program (SSP) (Figure 2) in 2019 to reduce demand variability. The idea is to establish a database for PCBA commodities and identify components' usage demand by referring to the PCBA Bill of Materials (BOM) and SAP production plan. ECE consolidates demand data from 15 manufacturing centers, then provides forecasts to distributors at the company level. About 2,000 electronic components were selected as the first cohort for the SSP program.

Figure 2

The Sponsor Company's PCBA Supply Chain ECE SSP Model



Note: A Stocking Strategy Program (SSP) was established by ECE to mitigate the impact of demand variability. However, the appropriate Safety Stock model has not been validated yet.

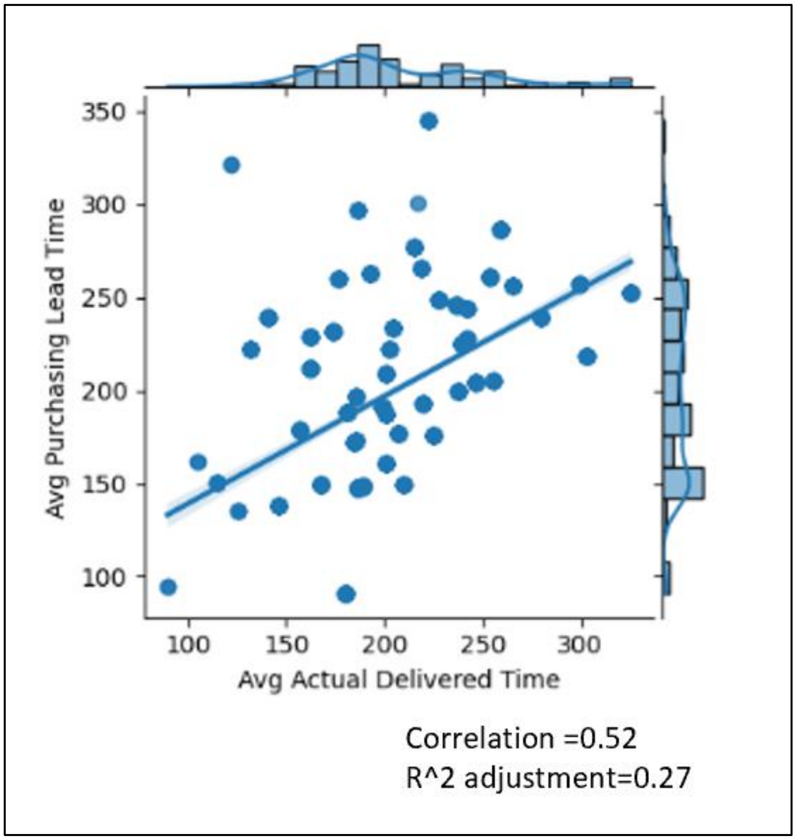
1.3 The Challenges of the SSP Program

The Stocking Strategy Program (SSP) is expected to significantly reduce demand variability, two important complicating factors are making the SSP less effective than desired. First, the SSP program currently does not consider or implement proper Safety Stock logic. However, a properly calibrated safety stock policy could help to reduce both product shortages and excess inventory, along with their associated costs.

Second, the active and passive PCBA components are creating a bottleneck due to their long and highly variable lead-times. Since 2018, the sponsor company has experienced price inflation and lead-time extension for active/passive components, which is likely driven by the rise in demand for electronic components from the mobile device and automotive industries. For some components, the quoted manufacturing lead-time is longer than 20 months, but the average actual delivery time is 150% of the quoted lead-time, with a 50% coefficient of variation (CV). Figure 3 illustrates that only 27% of the variability in the actual delivery time can be explained by the supplier`s quoted lead-time.

Figure 3

Correlation Between Quoted Lead-time and Actual Lead-time, Capacitor Commodity Year 2022



Note: 27% R-Squared value indicates there is significant variability between the actual delivery lead-time and the supplier quoted lead-time

The market competition has led the company to face uncertainty where the main feature is to offer a higher level of product customization (Corti et al. 2006) by shifting to a make-to-order (MTO) system. The sponsor company`s manufacturing center, which adheres to a MTO production system, exclusively produces customized products. In the MTO manufacturing system, it is crucial to determine an accurate due date by considering both the suppliers` quoted lead-time and the in-house manufacturing lead-time. Providing customers with an accurate due date is critical because any deviation from the expected due date would lead to a higher inventory cost for the manufacturing center (Gordon et al., 2002) and decrease the sponsor company`s competitiveness due to missing the customer`s expectation.

The extended and fluctuated supplier-quoted lead-times create ambiguity for due date assignment in the sponsor company`s supply chain model for PCBA manufacturing. When purchasing lead-times are inconsistent, it becomes challenging to forecast the availability of components for production. Lead-time variability also affects inventory management. For example, delivery delays for semiconductor components cause other components within the same BOM to become excess inventory for the sponsor company. Thus, the extended and uncertain lead-times for active/passive electronic components impact the sponsor company's supply chain, not only from a cost perspective, but also in terms of the availability of final products and services. Eventually, this situation impacts the customer satisfaction level. The sponsor company is therefore motivated to both incorporate a safety stock program and more accurately predict lead-times in order to reduce inventory costs and enhance the availability of active/passive components.

1.4 Problem Statement and Research Questions

The SSP program seeks to reduce the overall inventory cost, mitigate shortage risks, and enhance the availability of active/passive components in the sponsor company's PCBA supply chain model by aggregating demand from different OEMs and the sponsor company`s manufacturing centers, and managing components stock at a centralized location. However, the SSP program underperforms because it does not consider or implement safety stock. In addition, the SSP program's service level is reduced due to the high coefficient of variation (CV) value of historical component delivery times, and the fact that the lead-times quoted by electrical suppliers are not reliably accurate.

The objective of this project is to improve the SSP program's service level by developing a simulation model for safety stock and by utilizing machine learning techniques to predict future lead-times for active and passive electronic components. The questions to be answered include:

- What is a safety stock model that can be implemented in the SSP program, and how would it enhance service levels? How much inventory would the safety stock model save?
- What are the key factors that impact active/passive components' lead-times, and how can the sponsor company predict potential extensions in lead-times for these components?

We hypothesize first, that a safety stock level that incorporates both demand variance and component lead-time variance could enhance the service level for the SSP program (reduce stock out event) and reduce inventory costs (decrease overall system wait time). Second, we hypothesize that the high market demand for components in the automotive manufacturing and general electronic equipment industries surpasses manufacturing capacity, and thus leads to long and uncertain lead-times. Lastly, we also want to study whether component part types (transistor, op-amp, integrated circuit, resistor, capacitor, and diode) influenced components' lead-times in the past 10 years, and judge whether these variables can be used for future lead-time prediction.

1.5 Scope: Project Goals and Expected Outcomes

The project's overall goal is to improve the effectiveness of the SSP program by establishing a simulation model and informing decision making regarding safety stock level and lead-time prediction for active/passive components in the sponsor company's supply chain.

The deliverables to the SSP program will include:

- A methodology for determining an optimal safety stock for active/passive components under the SSP program.
- A dashboard for communication about potential lead-time extension compared with suppliers' quoted lead-times.

After implementing the methodology, the sponsor company expects to benefit from on-hand inventory cost reduction, and fewer shortage events affected by electronic components.

2. State of the Practice

As outlined in Chapter 1, the stretched and fluctuating delivery lead-time of electronic components impacts production planning and causes higher inventory costs across the overall production line. It is crucial to avoid shortages of electronic components, and maintaining safety stock is one methodology currently utilized to prevent such shortages. To calculate the necessary quantity of safety stock, a methodology to predict component delivery lead-time needs to be established.

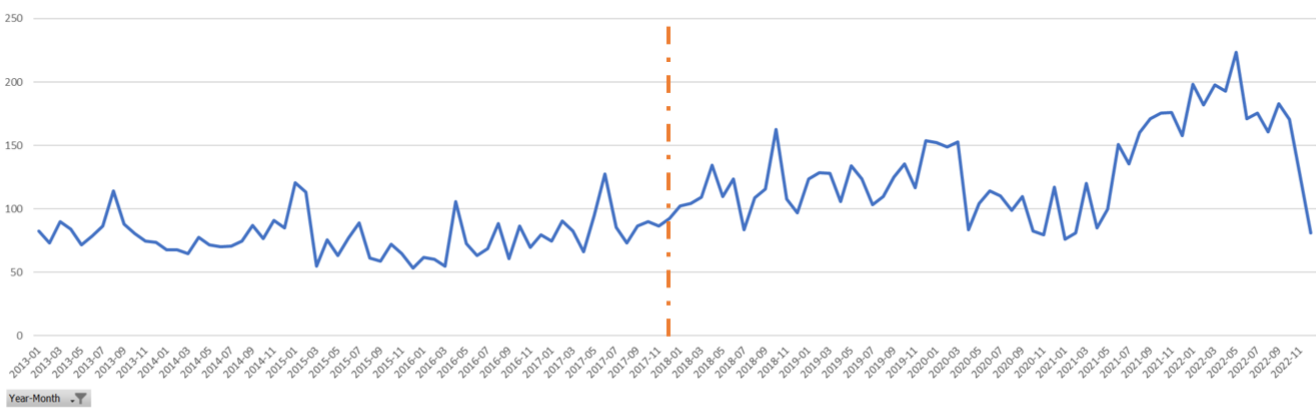
Machine learning has become a game changer in forecasting sales and lead-time prediction. While the concept of using historical data to predict trends has been around for centuries, machine learning started to revolutionize the field in the late 20th century. Unlike traditional analytical methods which struggle with complex and non-linear relationships, machine learning algorithms can discern patterns and dependencies in complex data, facilitating more accurate and reliable forecasts. To address the problem, we reviewed literature in several areas. First, we explored different machine learning models utilized for predicting lead-time across various industries. Second, we examined different safety stock models used in environments where lead-time fluctuates significantly.

2.1 Delivery Lead-time – Target Variable

Delivery lead-time is defined as the interval between the issuance of a purchase order to a supplier and the moment when an order is delivered to plant. A preliminary analysis of active and component purchasing lead-time fluctuations is depicted in Figure 4 and Figure 5. These figures clearly illustrate that the average purchasing lead-times have experienced more significant fluctuations since 2017, particularly during the Covid-19 period.

Figure 4

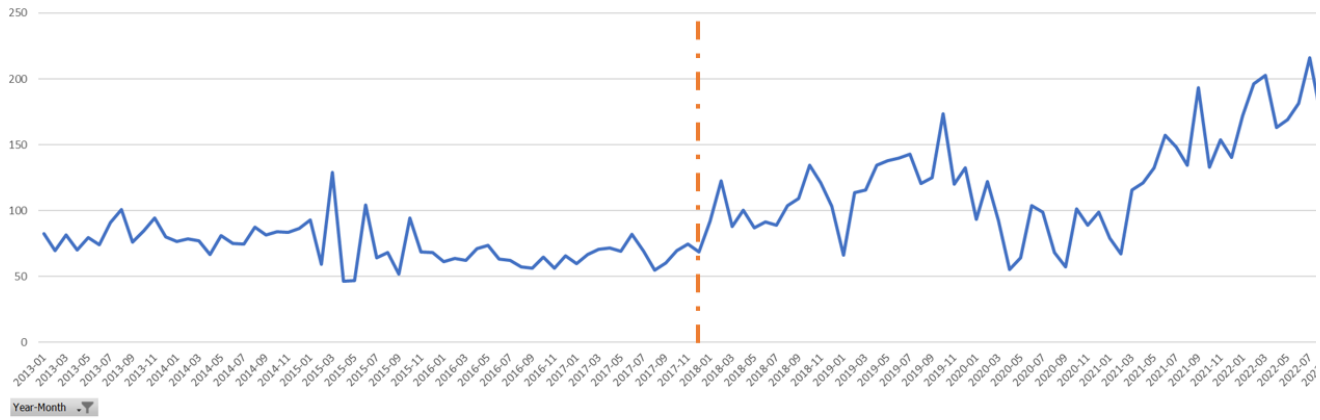
Average Monthly Lead-time of Passive Components (2012-2022)



Note: The average delivery lead-time of passive components has shown increased fluctuation since year 2018

Figure 5

Average Monthly Lead-time of Active Components (2012-2022)



Note: The average delivery lead-time of active components has shown increased fluctuation since year 2018

The variability in purchasing lead-time can significantly affect manufacturing performance. As noted in Chapter 1, when purchasing lead-time is inconsistent, it becomes challenging to forecast the availability of components for production. The variability of purchasing lead-times also affects inventory management. Delivery delays from semiconductor components cause other components within the same Bill of Materials (BOM) to become excess inventory for the sponsor company.

2.2 Supervised Machine Learning Linear and Tree Based Model

As outlined Chapter 1, the sponsor company is experiencing significant fluctuations in purchasing lead-time for both active and passive components. Predicting purchasing lead-time and proactively informing the procurement team about potential lead-time extensions can bring several benefits, including enhancing manufacturing performance, improving inventory management, and ultimately enhancing customer satisfaction.

Supervised machine learning is a subset of machine learning algorithms that focus on developing models that learn from labelled datasets. The datasets contain training data samples, with each sample consisting of input features paired with a target label. Supervised machine learning includes a variety of algorithms that analyze the relationship between input features and the corresponding output labels. The goal of supervised machine learning is to infer a function from labelled training data and make predictions for new, unseen data.

Recent studies have demonstrated the effectiveness of supervised machine learning algorithms in predicting lead-time in various industries. For instance, Shen et al. (2022) conducted an analysis of purchasing history data from semiconductor manufacturing by using eight machine learning algorithms and found that light gradient boosting algorithms exhibit solid predictive performance (0.91 as adjusted R squared value, and less than two weeks mean average error).

Oliveira et al. (2021) conducted a purchasing lead-time prediction for a pharmaceutical company by using machine learning and compared five machine learning algorithms, concluding that the support vector machine approach obtained the best performance with an average error of less than two days.

Sethi (2020) conducted a purchasing lead-time prediction for an Oil & Gas company by using machine learning and compared eight machine learning algorithms using Performance Metrics Mean Absolute Scaled Error (MASE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Normalized Root-Mean-Square Error (NRMSE). Sethi found that Random Forest (RF) and Linear Regression (LR) performed similarly across all five metrics and have higher accuracy compared to other machine learning algorithms.

Moreover, machine learning algorithms are also used for internal lead-time prediction. Lingitz et al. (2018) applied machine learning for a semiconductor manufacturer to predict internal lead-time and found that Random Forest and bagging Random Tree model outperformed compared with other models. According to their findings, Work in Process (WIP) appears to be the most important variable. Gyulai et al. (2018) conducted a machine learning experiment for an optics company, using material type, shape, and quantity as independent variables (features) and compared six machine learning algorithms. Gyulai et al. found that the Random Forest model is the best choice from an accuracy perspective, but in cases where simplicity is most important, the Linear Regression method can provide accurate enough prediction performance.

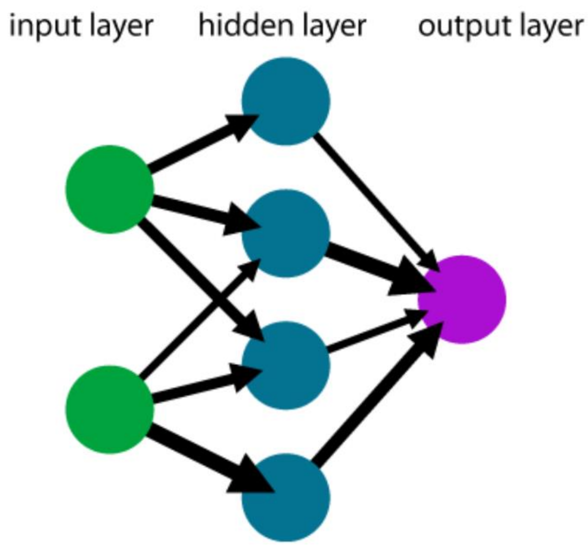
2.3 Supervised Machine Learning Neural Network (NN) and Deep Neural Network (DNN)

Both Neural Network (NN) and Deep Neural Network (DNN) are types of artificial intelligence based on the structure and function of the human brain (Figure 6). An NN is composed of interconnected neurons which are organized in layers. A typical NN has one input layer, one or two hidden layers, and one output layer. Each neuron in the layer is connected to several neurons in the next layer. The connections between neurons have weight that is adjusted during the machine learning training process. By updating the weights among neurons, the NN aims to minimize the difference between the actual output and the predicted output. The DNN is a

Neural Network with a greater number of hidden layers. DNN can improve model performance with techniques such as updating activation, dropout, and batch normalization.

Figure 6

A Simple Neural Network



Note: Special thanks to Murx for contributing this figure to Wikipedia

Hsu and Sha (2007) developed a Neural Network (NN) with backpropagation rule for due date prediction and concluded that NN has better performance compared with regression models. Asadzadeh et al. (2011) created a Neutral Network (NN) model for manufacturing lead-time prediction and concluded that NN outperformed compared with regression models in terms of relative error (Mean Absolute Percentage Error). Wang and Jiang (2019) conducted a Deep Neutral Network (DNN) model for in-house order completion time prediction by using real-time data from RFID devices. They concluded that DNN has better performance compared with NN for large-scale datasets.

2.4 Safety Stock Model for Inventory Management

This research project also addresses the analysis of a safety stock model specifically designed for components with highly fluctuating lead-times. Currently, the sponsor company is implementing a constant safety stock quantity approach. However, research studies have shown the limitations of this approach in responding to erratic demand patterns.

Kanet et al. (2010) ran a linear program by using Excel and concluded that a constant safety stock quantity is not a suitable approach to responding when there is an erratic demand pattern. Becker et al. (2013) proposed and validated a dynamic safety stock formula for components with a high deviation of purchasing delivery lead-time. Bahroun and Belgacem (2019) demonstrated that dynamic safety stock can minimize safety stock quantity, improve service levels, and decrease the probability of stock-outs when there is lead-time fluctuation. This capstone project plans to examine the fundamental dynamic safety stock model $SS = \sqrt{LT\sigma_D^2 + \delta_L^2 D^2}$ and study if inventory amounts can be reduced at the whole manufacturing center level.

From the above literature review, we conclude that the supervised machine learning models, which include regression, tree-based, and neural network models, have been shown to improve the lead-time prediction in other industries. The regression model offers simplicity and interpretability, while the tree-based model, such as Random Forest, offers robustness against overfitting and can handle non-linear data. The Neural Network models, particularly the Deep Neural Network model, show better performance for large-scale datasets. These models could help the sponsor company with predicting delivery lead-time for electronic component . We also looked at the potential safety stock model when dealing with fluctuated lead-time environments. By incorporating machine learning results into the dynamic safety stock model, this capstone project aims to help the sponsor company to reduce inventory at the whole manufacturing center level. Chapter 3 will explain the data from the sponsor company and how we use these methods and data for delivery lead-time study.

3. Methodology

In the previous chapter, we delved into the utilization of machine learning models for the prediction of lead-time across diverse industries. We also conducted a literature review about the safety stock model applied for highly fluctuating lead-time environments. In this chapter, we explain the data source from the sponsor company, outline the proposed machine learning models to be employed in this capstone project, and delineate the simulation model to analyze safety stock level.

3.1 Data Source

The primary source of data for this capstone project is the sponsor company `s purchasing historical report between year 2012 to year 2022 (order date). This dataset includes Passive / Active Category, Type, Manufacturer Name, Standard, High Temperature Grade, Special Material Used, PO (purchase order) lines, PO

Order Date, Received Date, and Purchasing Lead-time. The Actual Delivered Time, which is calculated from Receipted Date subtracted from the PO Order Date, is treated as the target variable. Also, if the Actual Delivery Time is shorter than 56 days, then the PO line is considered to be in the “Yes” category of the “on-hand or not” line, which is excluded from the model studies (Table 1).

Table 1

Features from Sponsor Company Purchasing Historical Report

Feature Name	Remark
Passive/Active	Distinguishing between Passive and Active components
Type	Component Commodity
Manufacture Name	Producer of the Component
Standard	Specify the standard that Component is designed for
High Temperature Grade	The operation range of the Components
PO lines	Unique identifier of purchasing transaction
PO Order Date	The date on which the purchase order was placed
Receipt Date	The date on which the purchase order was closed
Purchasing Lead Time	The Lead Time Supplier quote for the Components
Actual Delivery Time	Difference between Receipt Date and PO Order Date
On Hand or Not	Yes, if Actual Delivery Time is shorter than 56 days

In the data cleansing process, this capstone project focuses on a specific subset of the sponsor company`s data to ensure data quality. The data cleansing process involved several exclusions:

- Catalog components: Only catalog components were included in the study. The customized electronic components designed for the sponsor company were excluded.
- Inter-manufacturing center purchase: The purchase orders reflecting the transactions among the sponsor company`s manufacturing centers were excluded. The capstone project focuses on external purchasing.
- Obsolete components: The purchase orders for obsolete components purchasing were excluded.
- Data errors: The components which have data errors within the sponsor company`s dataset was excluded. See the example provided in Table 2. Once component number has multiple MPN (Manufacture part number) be specified. One MPN has spec and standard be specified, while the other does not.

Table 2

Example of Data Error

Component Number	TYPE	Manufacture Name	MPN	Special Material	Feature
000-*****	Resistor	KOASPEER	RK73H2ATTE****	Tin, Nickel	Military: -55C to +155C
000-*****	Resistor	VISHAY	CRCW08051****		--

Note: Same component number with different MPN, however, one MPN is missing spec information

The sponsor company operates a high-mix, low-volume production system, resulting in a limited number of PO lines for machine learning studies. Seven types of electronic components are selected for analysis in the Delivery Time study. Details regarding component type, remark, and total number of PO lines for each component type are provided in Table 3.

Table 3

Component Type for Delivery Lead-time Study

Type	Passive/Active	Remark	PO lines Qty
Chip Capacitor	Passive	Surface mounting capacitor, store electrical power	6K~
Film Resistor	Passive	Surface mounting resistor to control current flow	4K~
OPAMP	Active	Function as a voltage amplifier	3K~
Power Supply IC	Active	Integrated Circuit (IC) designed to regulate voltage level	2K~
Rectifier diode	Passive	Convert alternating current (AC) to direct current (DC)	1K~
MOSFET Transistor	Active	Amplifying or switching electronic signals	1K~
IC Interface	Active	Convert data signal between components	1K~

3.2 Exogenous Features

As noted in Section 3.1, the sponsor company has a high-mix, low-volume production system. The characteristic result is a small number of purchase orders (PO lines). Consequently, the historical purchasing report provided by the sponsor company offered limited features for the electronic component delivery lead-time study.

To address this challenge, the capstone project adopted the approach of incorporating exogenous variables. Exogenous variables represent external factors that can influence the target variable (in our capstone project, it is demand for electronic components) but are not directly controlled by the sponsor company. By incorporating exogenous variables, the project aims to improve the accuracy and performance of the machine learning model for the delivery lead-time study. There is research supporting the use of exogenous variables in the forecast studies from other industries. For instance, Alvarez-Chaves et al. (2021) successfully incorporated exogenous variables such us public holidays, academic calendars, and sporting events into a predictive model for patient flow in an emergency department. Their study demonstrated these exogenous variables improved model performance.

There are 15 market indexes incorporated as exogenous variables in this capstone project. Details are provided in Table 4.

Table 4

The Exogenous Variables Incorporated in this Capstone Project

Market Index	Description	Remark
PHLX Semiconductor Index	A NASDAQ market index that track performance of semiconductor manufacturer	Includes companies such as Intel, AMD, and NVIDIA, etc.
S&P 500-Aerospace & Defense	A S&P500 market index that track performance of aerospace and defense manufacturer	Features companies like Boeing and Lockheed Martin, etc.
S&P 500-Oil & Gas Equipment & Services	A S&P500 market index track the Oil and Gas service service company	Comprises companies such as Schlumberger (SLB) and Halliburton, etc.
S&P 500-Automobile Manufacturers	A S&P500 market index track the performance of automobile company	Contains companies such as General Motors (GM) and Tesla, etc.
S&P 500-Communications Equipment	A S&P500 market index track the performance of communication equipment	Includes companies like CISCO and Motorola Solutions,etc.
S&P 500-Health Care Equipment	A S&P500 market index that track the performance Health Care equipment	Features companies such as GE Healthcare and Medtronic,etc.
S&P 500-Semiconductor Materials & Equipment	A S&P500 market index that track the performance of Semiconductor equipment manufacture	Comprises companies such as Teradyne and KLA Corporation,etc.
S&P 500-Semiconductors	A S&P500 market index that track the performance of Semiconductor company	Includes companies like Analog Devices and Intel,etc.
S&P 500-Technology Hardware, Storage & Peripherals	A S&P500 market index that track the performance of IT Hardware Company	Contains companies such as HP Inc. and Apple,etc.
TOPIX Transportation equipment	A Tokyo Stock Exchange market index that track the Japanese automotive manufactures and part suppliers	Includes companies like Toyota and Honda,etc.
TAIEX	Taiwan Stock Exchange Index, which include to capture the semiconductor manufactures in Taiwan	Stock Index
FTSE	London Stock Exchange Index, which include to capture the 100 major companies on London Stock market	Stock Index
DAX	Deutscher Stock Exchange Index, which include to capture the 40 major company on Frankfurt Stock market	Stock Index
Nikkei	Tokyo Stock Exchange Index, which include to capture the 225 major company on Tokyo Stock market	Stock Index
GSCPI	A index that measure the level of pressure on global supply chain	Supply Chain Pressure Index

3.3 Machine Learning Models Selected for Lead-time Prediction

Based on the literature that we reviewed, we identified and will utilize the following supervised machine learning algorithms to predict electronic component delivery lead-times.

- Linear Regression is a statistical approach that models the relationship between input features and output. The input features are called the independent variables, and the output is called a dependent variable.
- Ridge Regression is a variant of linear regression that incorporates regularization to handle multicollinearity (high correlation) among the input features and mitigate overfitting. It can help improve the generalization performance of the model by shrinking the coefficient of the correlated independent variables.
- Lasso Regression is a regression analysis method like ridge regression but has a different type of regularization that encourages sparsity in the model by setting some regression coefficients to zero. Lasso might select one and ignore other independent variables.
- Extreme Gradient Boosting (XGB) is based on decision tree algorithms. It starts with a simple tree model, then trains additional models to correct the error of the previous one. This helps to reduce both bias and variance, leading to more accurate prediction.
- Random Forest (RF) is a machine learning model that makes predictions by aggregating the results from multiple decision trees. Each tree is built from a random subset of the data and a random selection of the features. It takes the average of prediction results from multiple trees to mitigate the overfitting and underfitting while maintaining robustness.
- Deep Neural Network (DNN) is a type of artificial neural network that is composed of multiple layers between input and output layers. Each layer in the network performs a series of computations and transforms the results to the next layer. DNN can model complex non-linear relationships between input features and the output target variable. DNN can use techniques activation, dropout, and batch normalization to improve model performance.

3.4 Safety Stock Model Study

The sponsor company currently employs two distinct safety stock models. Each is tailored to a specific purpose.

- Safety Stock Considering Demand Variation = $\sigma_D \times \sqrt{LTq}$. Where σ_D denotes the standard deviation of demand during the lead-time and LTq represents the lead-time as quoted by the supplier.
- Safety Stock Accounting for Sales Lead-time Exceeding Cumulative Lead-time = $(LTc - LTs) \times Qm$ where LTc denotes the Cumulative Lead-time, which is the sum of the longest purchase component lead-time and the In-house Manufacturing Lead-time. LTs denotes the sales Lead-time that committed to customers. Qm is the mean of demand.

In the current models, the deviation from the delivery lead-time of electronic components is not considered. This oversight may lead to inaccuracies in safety stock calculation, potentially resulting in increased stock-out events and higher inventory cost.

The capstone project aims to investigate whether the safety stock quantity from a dynamic safety stock model (Equation 1) can decrease stock out events and reduce inventory cost. The simulation will be conducted using Python Simpy. The lead-time and the deviation of lead-time will be sourced from the validated machine learning model in the preceding section.

$$\text{Dynamic Safety Stock } = \sqrt{Max(LTq, LTp)\sigma_D^2 + \delta_L^2 D^2} \qquad (1)$$

Where:

LTq denotes the quoted lead-time from supplier.

LTp denotes the mean of delivery lead-time, as predicted by the machine learning model.

D denotes the mean of demand.

δ_L denotes the standard deviation of delivery lead-time, as predicted by the machine learning model.

σ_D is standard deviation of demand.

4. Results and Discussion

In Chapter 3, we presented the methodologies employed in this capstone project to predict the lead-time of electronic components. This project encompasses a study of lead-time prediction for seven types of electronic components, utilizing the sponsor company`s historical purchasing report with exogenous variables. In this chapter, we present the results and discuss the findings derived from the machine learning models.

As delineated in chapter 3, the objective is to develop the safety stock model, which reduces the risk of delivery shortages due to market fluctuations. The target variable for the machine learning models is the weekly average delivery lead-time of electronic components.

In order to analyze the trends associated with the average delivery lead-time and its standard deviation, four features have been engineered and incorporated into the dataset for machine learning model analysis. These features include ”AvgD lag270” and “AvgD lag360,” which represent the weekly average delivery lead-times for the prior nine months and one year, respectively. Additionally, “Dev lag270” and “Dev lag360” have been introduced to capture the standard deviation of the delivery lead-times for the corresponding period. The purpose of integrating these features is to enhance the model's capability to discern patterns and make more accurate predictions regarding delivery lead-times.

4.1 Results of Linear Models

The results of applying the Linear, Ridge, and Lasso models are provided in Table 5. It has become evident that the linear regression models employed do not successfully capture the complexities inherent in the dataset. The R squared values, when assessing both the training and the test sets, consistently reveal modest to low explanatory power. For instance, the linear model`s R squared value for Diodes is as low as 0.08 on the test set, highlighting a significant underfitting issue. This analysis result suggests the need for a more sophisticated machine learning model that can account for the non-linear relationship between the features and the target variable.

Table 5

Performance of Linear model for seven types of Electronic Components

	Resistor			Capacitor			Diode			Amplifier			Interface			Transistor			Power Supply		
	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE
Linear Model	0.42	0.48	0.26	0.48	0.42	0.17	0.3	0.08	0.38	0.51	0.19	0.26	0.4	0.3	0.38	0.43	0.27	0.35	0.31	0.19	0.32
Ridge Model	0.42	0.48	0.26	0.48	0.41	0.17	0.3	0.07	0.38	0.51	0.19	0.26	0.4	0.28	0.38	0.43	0.28	0.35	0.31	0.21	0.31
Lasso Model	0.41	0.48	0.26	0.48	0.41	0.17	0.3	0.05	0.39	0.5	0.21	0.26	0.4	0.3	0.37	0.43	0.28	0.35	0.31	0.21	0.31

Note: The underperformance of linear models, as indicated by both R-squared and MAPE metrics, suggests data is non-linear.

4.2 Results of Tree Models

The results of Regression Tree, RF, and XGB model when predicting average delivery lead-time are provided in Table 6. The table compares performance by using three types of machine learning models across seven types of electronic components. The performance metrics used are R squared for both training and test datasets, and the Mean Absolute Percentage Error (MAPE).

The RF model exhibits a higher level of R squared values for all datasets, but not proportionally high-test R squared values when comparing it to the XGB model. This indicates that while the RF model captures the training data well, it struggles to generalize its predictions to unseen data.

The XGB model shows less overfitting than the RF model, with the training R squared values generally having similar values as those of the RF model. However, the test R squared values are higher than those of the RF model. This suggests that the XGB model has more balanced performance between fitting the training data and the unseen data. The MAPE values for the XGB model are similar to those for the RF model, indicating a similar level of prediction error.

The capacitor and diode datasets show relatively low R squared values of 0.4 and 0.41, respectively, on the test data. This result indicates that the model can explain only around 40% of the variance in the unseen data of these components. Upon reviewing the dataset, it was noted that, compared to the other types of electronic components, military spec constitutes a higher proportion of the data- 98% for capacitors and 79% for diodes. These findings suggest that additional features should be investigated and potentially incorporated into the model to improve its predictive accuracy for the components used by the sponsor company.

Table 6

Performance of tree models when predicting Average Delivery lead-time for seven types of Electronic Components

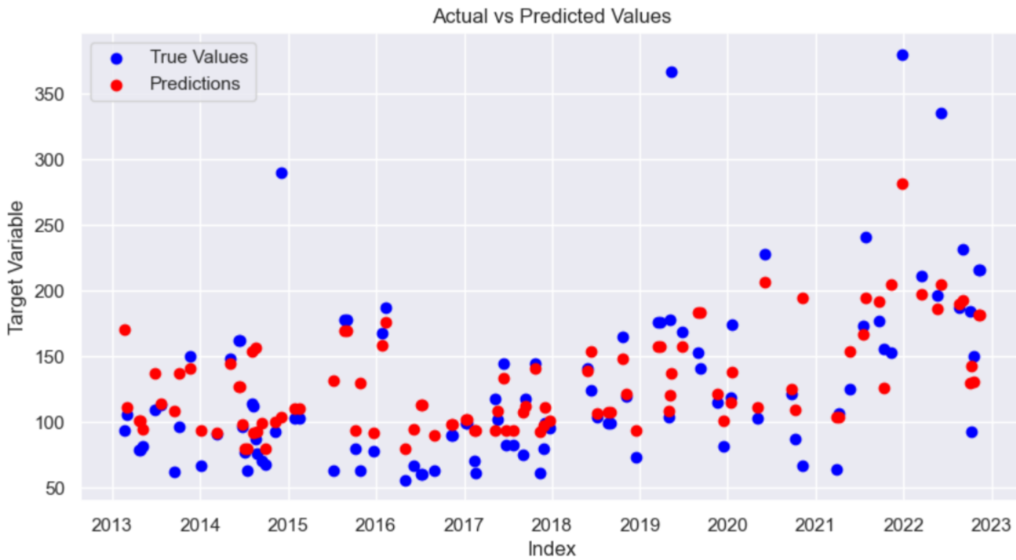
	Resistor			Capacitor			Diode			Amplifier			Interface			Transistor			Power Supply		
	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE
Tree Model	0.81	0.48	0.23	0.78	0.07	0.25	0.73	-0.03	0.37	0.74	0.36	0.21	0.79	0.33	0.27	0.80	0.13	0.28	0.77	0.28	0.30
RF Model	0.87	0.53	0.23	0.75	0.32	0.21	0.86	0.41	0.32	0.82	0.44	0.20	0.89	0.52	0.24	0.85	0.42	0.24	0.88	0.48	0.27
XGB Model	0.89	0.60	0.20	0.73	0.40	0.18	0.87	0.41	0.33	0.84	0.50	0.21	0.90	0.54	0.23	0.87	0.50	0.25	0.89	0.55	0.25

Note: the tree models demonstrate improved performance, with the XGB model outperforming both the regression Tree and the Random Forest Models.

The visualization graphs for of the test and prediction data for transistor and interface using the XGB model are provided in Figure 7 and Figure 8. Both graphs affirm the model`s predictive capabilities. The graphs show that while neither model is perfect, both provide a solid foundation for estimating component behavior. It can be confirmed that the model achieves a modest accuracy (0.5 and 0.54 respectively) which can capture the general trends. But its performance is limited, particularly when there is a high level of variability.

Figure 7

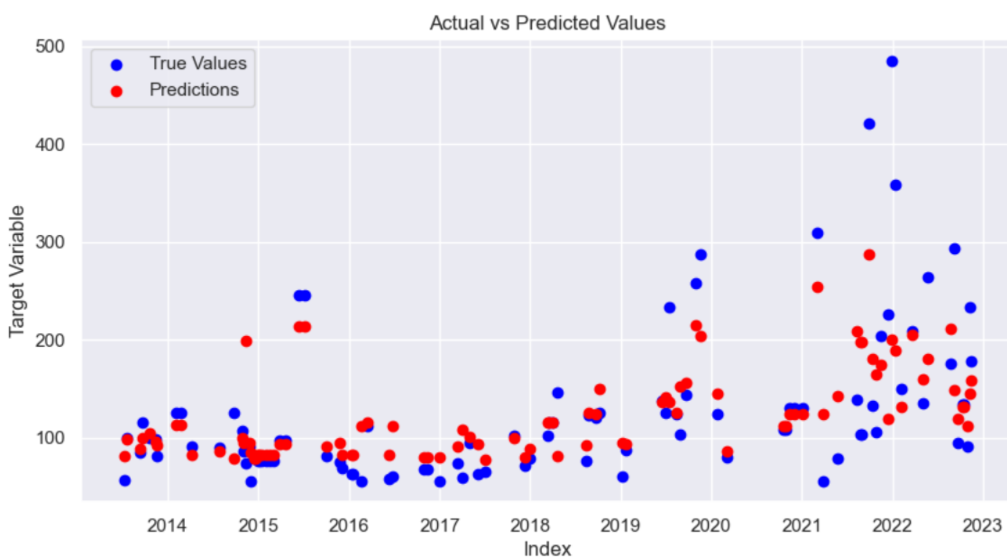
Transistor Average Delivery Lead-time Prediction by Using XGB Model



Note: plot between test value and prediction value, suggest XGB has certain level of capability in predicting delivery lead-time, but it appears to have limitation when dealing with extreme values.

Figure 8

Interface Average Delivery Lead-time Prediction by Using XGB Model



Note: plot between test value and prediction value, suggest XGB has certain level of capability in predicting delivery lead-time, but it appears to have limitation when dealing with extreme values.

The permutation importance table (Table 7) for predicting the semiconductor average delivery lead-time highlights the key findings regarding the influence of various exogenous and generated features on prediction models. The market indices, such as the S&P 500 Technology Hardware, S&P 500 Communication Equipment and S&P 500 Healthcare, frequently appear at the top of the importance rankings across different types of semiconductor components. This indicates that the performance of companies tracked by these indices, which include firms like Apple, HP, CISCO and GE Healthcare, might directly influence demand patterns, supply chain pressure, and, consequently, the delivery lead-time. Additionally, the generated features “Dev lag” and “AvgD lag” are consistently ranked within the top three for different types of semiconductor components, suggesting that these lagged deviations and average lead-time features are vital for capturing trends in delivery lead-time. These features likely reflect that the past market conditions are instrumental in forecasting future supply conditions. Moreover, the S&P 500 Oil and Gas Index, which the sponsor company is part of is also added to the prediction model. Its presence in the lower-ranked list suggests that the performance of oil and gas companies does not influence semiconductor delivery time. Lastly, the presence of global indices like NIKKEI and DAX in the lower ranking suggests that while international market conditions influence delivery lead-time, their impact is less significant when compared to the sector-specified indices mentioned above.

Table 7

Permutation Importance of Exogenous Features

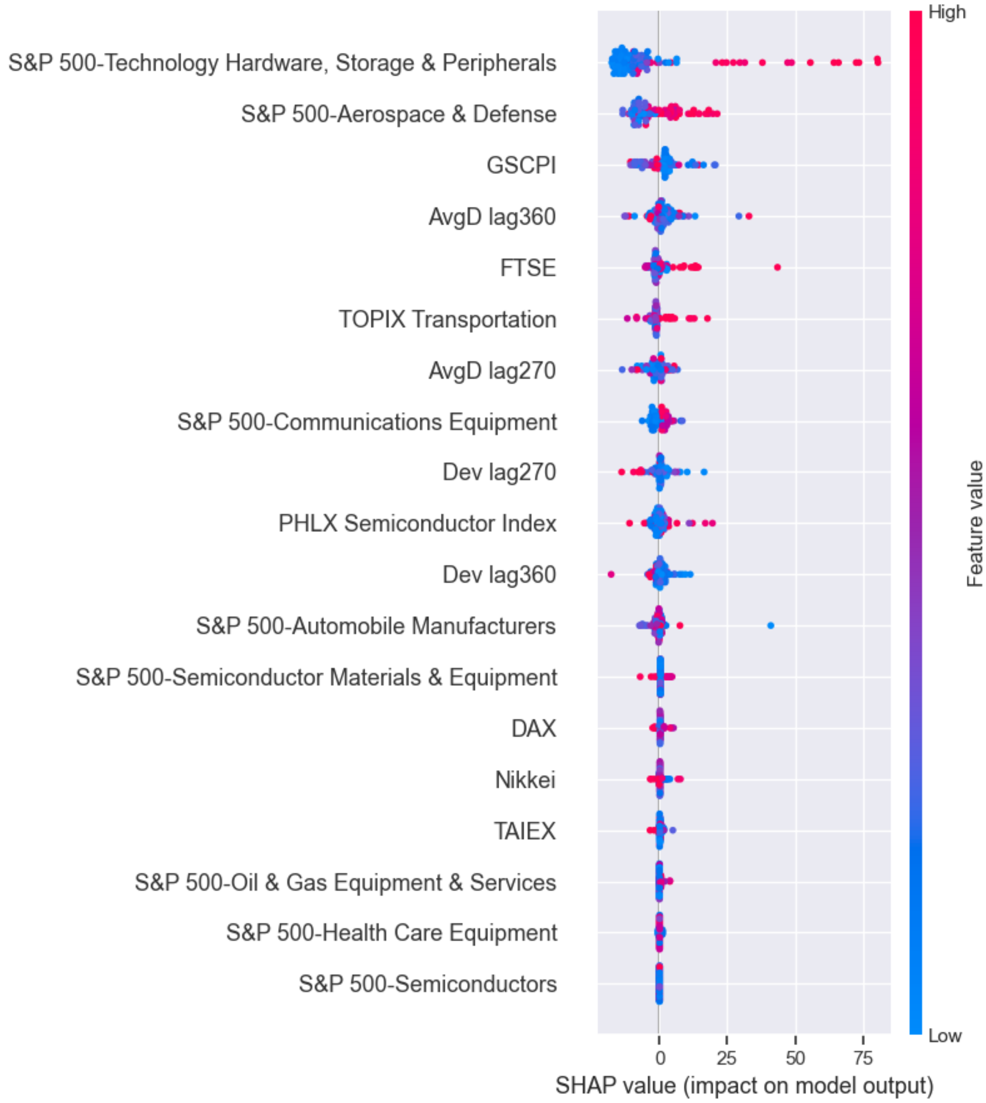
	Resistor	Capacitor	Diode	Amplifier	Interface	Transistor	Power Supply
1	S&P 500 Semicondutor	S&P 500 Health Care	S&P500 Technology Hardware	S&P500 Technology Hardware	S&P 500 Communication Equipment	S&P 500 Technology Hardware	S&P 500 Communication Equipment
2	TOPIX	S&P 500 Aerosapce	Dev lag360	S&P500 Aerosapce	AvgD lag360	AvgD lag360	S&P 500 Technology Hardware
3	AvgD lag360	Dev lag 270	Dev lag270	AvgD lag270	Dev lag270	AvgD lag270	AvgD lag360
17	GSCPI	S&P 500 Semiconductor	S&P 500 Oil&Gas	Nikkei	DAX	S&P 500 Oil&Gas	DAX
18	S&P 500 Health Care	DAX	DAX	S&P 500 Oil&Gas	S&P 500 Oill&Gas	Nikkei	S&P 500 Automobile
19	S&P 500 Semiconductor Equipment	S&P 500 Oil&Gas	TOPIX	S&P 500 Semicondutor	S&P 500 Health care	S&P 500 Health Care	S&P 500 Semicondutor Equipment

Note: The S&P 500 indices for technology, healthcare, aerospace, and communication, which include companies with substantial semiconductor consumption such as Apple, GE Health, Boeing, and Cisco, are key features for predicting delivery lead- time. The lag features which representing historical data also have a relatively high impact on predicting delivery lead-time. In contrast, the S&P 500 Semiconductor and Semiconductor equipment indices, despite being directly related to semiconductor companies, do not demonstrate a marked impact on prediction of lead-time. Moreover, S&P 500 Oil & Gas which includes the sponsor company exhibits a minimal effect on lead-time prediction. The global economic indicators are also not significant influencer for lead-time prediction.

The SHAP value (as shown in Figure 9) illustrates the impact of exogenous features, especially sector-specified indices on the average delivery lead-time for semiconductor components. Taking the example of an amplifier, it is evident that the higher value of indices like S&P 500 Technology Hardware and S&P 500 Aerospace correlates with an increase in delivery lead-time. This suggests that strong sector performance may lead to stretched delivery lead-time due to the increased demand.

Figure 9

SHAP Value of Exogenous Features (amplifier)



Note: The SHAP values explain how each feature impacts the semiconductor delivery lead-time prediction. The higher value of S&P 500 technology index tends to increase the predicted delivery lead-time. In contrast, the lower value of S&P 500 does not correspondingly decrease the predicted delivery lead-time, likely due to the influence of procurement negotiation by the sponsor company.

Mark (2017) discussed Apple\`s procurement strategy, noting that Apple often secures suppliers\` production capacities in advance to guarantee the availability of essential parts. This approach is driven by Apple\`s sales target, which guides the scaling of production and material procurement, including making advance payments when necessary. This practice underscores how demand from major semiconductor consumers directly influences demand patterns, innovation cycles, and, consequently, the pressures on the semiconductor supply chain. This confirms that the sector performance and major consumer procurement strategies are critical factors in influencing delivery lead-time for semiconductor components in the market

The results of Regression Tree, RF, and XGB models when predicting standard deviation of delivery lead-time are provided in Table 8. Similarly, when predicting the average delivery lead-time, the XGB model outperforms both the Tree and RF models. The result highlights that XGB has higher performance in handling both types of predictions.

Table 8

Performance of Tree Models When Predicting STD deviation of Lead-time for Seven Types of Electronic Components

	Resistor		Capacitor		Diode		Amplifier		Interface		Transistor		Power Supply	
	Train R Square	Test R Square	Train R Square	Test R Square	Train R Square	Test R Square	Train R Square	Test R Square	Train R Square	Test R Square	Train R Square	Test R Square	Train R Square	Test R Square
Tree Model	0.74	0.14	0.70	0.07	0.67	0.22	0.69	0.33	0.77	0.22	0.77	0.35	0.89	0.16
RF Model	0.72	0.39	0.63	0.19	0.82	0.39	0.75	0.50	0.81	0.46	0.76	0.51	0.86	0.31
XGB Model	0.73	0.36	0.83	0.19	0.79	0.39	0.90	0.58	0.91	0.50	0.88	0.62	0.92	0.40

Note: In predicting STD deviation of lead time, the XGB model either outperforms or matches the performance of both the regression Tree and the Random Forest Models.

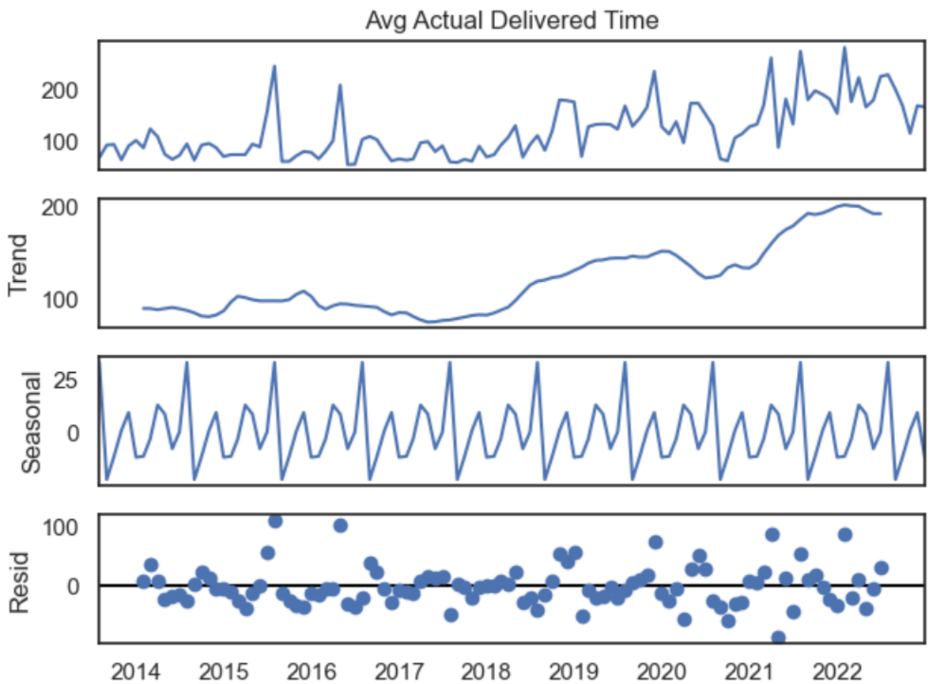
4.3 Study with Time Series Models

A time series model for the average delivery lead-time study was developed. First, we conducted a seasonal decomposition on the dataset to identify underlying patterns. The analysis result (Figure 10) confirmed the presence of a trend component and demonstrated a weekly seasonality. For predictive modeling, the dataset was divided into the training and test dataset, with October 1, 2021, as the cut-off date. The data prior to this date constitute the training dataset.

The study employed Decision Tree, RF, and XGB models to forecast the average delivery lead-time based on the training data. The Decision Tree model was particularly ineffective, resulting in a constant prediction across the timeline and essentially producing a straight line without variability. Both the RF and XGB models also performed poorly on the test dataset. Taking interface component as an example, the models output negative R squared value of -0.58 and -0.63, respectively. These values suggest that the models were unable to capture the underlying pattern effectively. In fact, they performed worse than a simple baseline model that would predict the mean of the target variable. Additionally, the MAPE metrics, standing at 0.54 for the RF model and 0.43 for the XGB model, further underscore the models` limited predictive accuracy. V

Figure 10

Seasonal Decomposition Study on Interface Component



Note. Seasonal index shows the weekly average, calculated on a monthly basis.

Visual analysis (Figure 11 and Figure 12) of the results revealed a significant discrepancy between the test data and prediction data, highlighting a substantial gap and misalignment when using time series model for average delivery lead-time prediction.

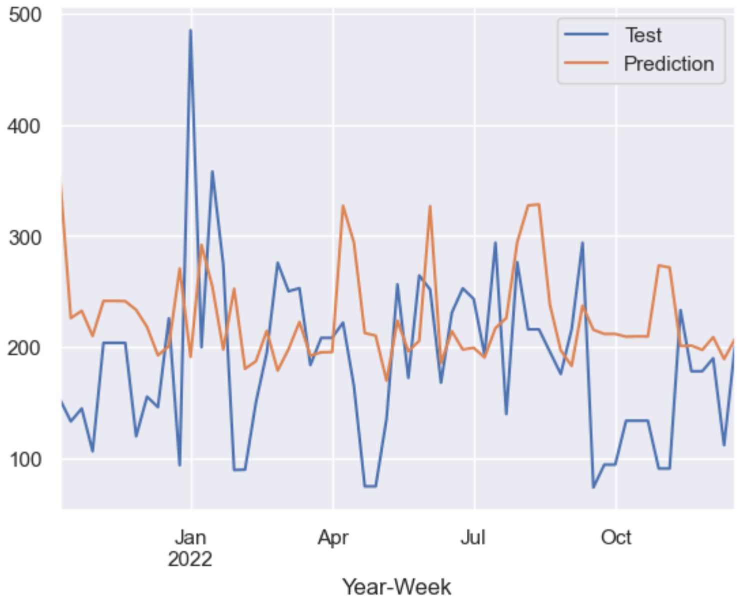
The analysis confirms the presence of both trend and seasonal patterns in the data, as seen in the decomposition graphs. However, the seasonality, which repeats within each year, does not function for future prediction on a year-over-year basis with the current modeling approach.

4.4 Result of DNN model

A DNN model was developed for the average delivery lead-time study. The model is structured with three dense layers, each containing 32 neurons and using ReLU as an activation function. Adam is employed as the optimizer and Mean Absolute Error (MAE) is used as the loss function. Additionally, an R_squared function was created as one of the metrics and MAPE was used as the other metric. The model is trained by using training data with a batch size of 64 for up to 200 epochs.

Figure 11

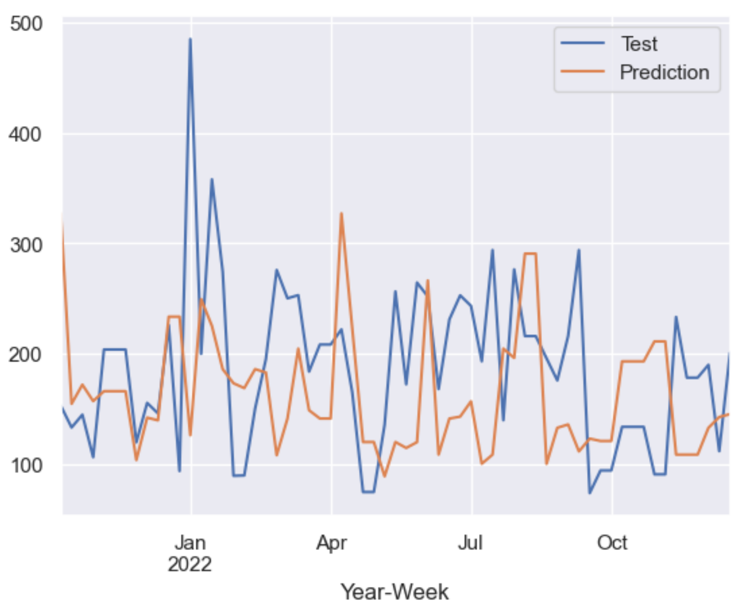
Time Series RF model, Interface Component Test and Prediction Data Comparison



Note: limitation of time series Random Forest model for interface component delivery lead-time prediction. MAPE:0.54

Figure 12

Time Series XGB Model, Interface Component Test and Prediction Data Comparison



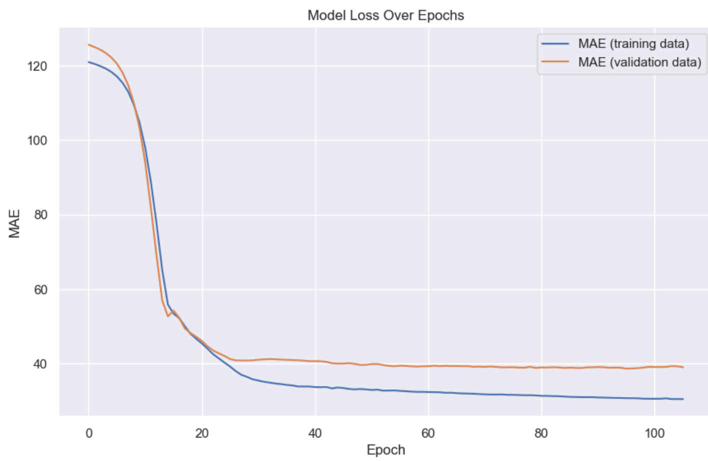
Note: limitation of time series XGB model for interface component delivery lead-time prediction. MAPE:0.43

To avoid overfitting, an early stopping mechanism was implemented. Training stops if there is no improvement in MAE for 10 consecutive epochs with a minimum delta of 0.1.

The loss curve and visualization of test and prediction data for the interface components are provided in Figures 13 and 14, respectively. The loss curve indicates that the DNN model implemented early stopping at epoch 110 to prevent overfitting. The final MAE of the validation set was 0.4. The model`s MAPE was 0.31, which is inferior to the performance of the previously demonstrated XGB model. The visualization graph reveals that the DNN model struggles to accurately predict extreme values in the test data.

Figure 13

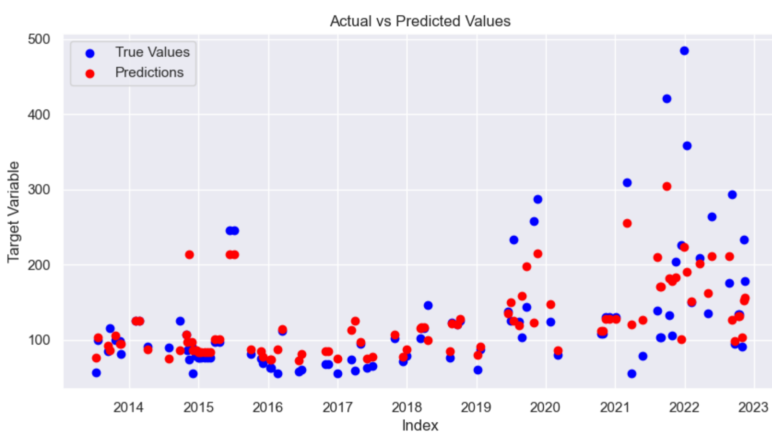
Training and Test Loss from Interface DNN Model



Note: both training and testing loss function stabilize after approximately 20 epochs, and early stopping is triggered at 105 epoch.

Figure 14

Interface Average Delivery Lead-time Prediction by Using DNN Model (MAPE:0.31)



Note: DNN model shows limitation on extreme values.

The results of the DNN model when predicting the average delivery lead-time are provided in Table 9. The data confirm that the DNN model underperforms compared to the XGB model across all seven types of components for both the training and test datasets. The MAPE also indicates that the performance of the DNN model is either slightly worse or equal to that of the XGB model. It is important to note that the DNN model`s complexity, including factors such as layer adjustments, unit size, and batch size, can significantly influence the model`s performance. Due to time constraints, this capstone project could not conduct a thorough study on optimizing parameters for the DNN model. Identifying the optimal parameters for the DNN model remains an action item for future research.

Table 9

Performance of DNN Models When Predicting Average Delivery Lead-time for Seven Types of Electronic Components

	Resistor			Capacitor			Diode			Amplifier			Interface			Transistor			Power Supply		
	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE	Train R Square	Test R Square	MAPE
DNN Model	0.41	0.43	0.21	0.51	0.21	0.19	0.15	-0.25	0.36	0.47	0.14	0.23	0.31	0.3	0.31	0.47	0.34	0.29	0.14	0.12	0.26

Note: Given that the R-squared values for both training and testing are low, the DNN model is underfitting the dataset.

4.5 Safety Stock Simulation by Using SimPy

In this section, we used SimPy to build a simulation model to compare the wait time based on component availability using static safety stock (SS) and dynamic safety stock (SS). The static SS is calculated using formula $\sigma_D \times \sqrt{LTq}$ where σ_D is the standard deviation of demand while LTq is the LT quoted by the supplier. The dynamic SS is calculated using Formula 1, where LTp and δ_L^2 are derived from the XGB model we developed in section 4-2.

There are two types of variability that can be compared in the simulation model. One is the demand variability. The simulation model is designed to explore demand variability in a high-mix, low-volume production environment, tailored for the sponsor company. The model`s parameters can be found in Table 10. This model generates sales orders in the fourth week of each month, with an average quantity of four pieces. The variability in demand is reflected through a standard deviation ranging from 1 to 4 for these orders. To effectively manage this variability, both static and dynamic safety stock quantities are adjusted accordingly. Additionally, the model accounts for a default administration time of one week from the order`s generation to its processing.

Table 10

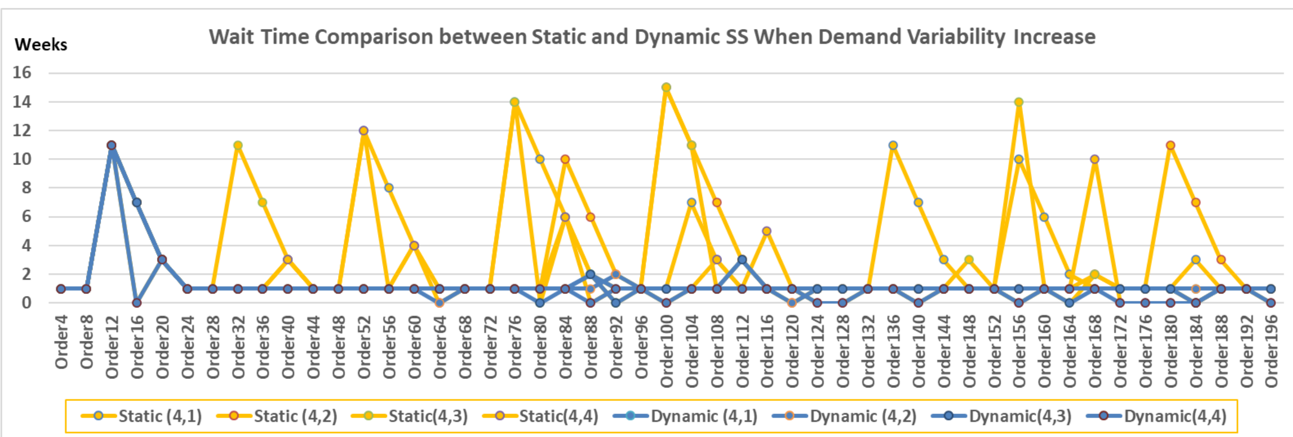
Simulation Parameter for A Demand Variability Increasing Environment

Parameter	Value	Remark
Supplier Quoted LT	10 weeks	
Average LT	20 weeks	Derived from XGB model
STD deviation LT	8 weeks	Derived from XGB model
Sale Order Demand	4 pieces	Demand be generated in the 4th week of each month
Demand Deviation	1,2,3,4 pieces	
Static SS	4,6,7,11 pieces	Change aligns with STD deviation of Demand
Dynamic SS	8,11,13,20 pieces	Change aligns with STD deviation of Demand
Simulation Time	200 Weeks	

Figure 15 presents the comparative result between the static SS and dynamic SS as demand variability increases. It is evident that the dynamic SS model, which incorporates the predicted delivery lead-time and the standard deviation of lead-time, significantly reduces the risks associated with supply chain variability in components. Apart from the initial phase, where both the static SS and dynamic SS models exhibit the wait time due to an initial buildup of inventory, the dynamic SS model effectively mitigates component shortages throughout the simulation time steps.

Figure 15

Wait Time Comparison between Static and Dynamic SS When Demand Variability Increase



Note: The wait time due to components shortage. Excluding the initial time step, where the model builds up initial inventory, the dynamic safety stock model outperforms the static safety stock model.

The second simulation (Table 11) focuses on lead-time variability while maintaining a constant demand of four pieces every month (generated on the fourth week of each month). The model evaluates the impact of varying the standard deviation of lead-time, set at two, four, six and eight weeks. The simulation was conducted 500 times to ensure reliability, and the average wait time reported is derived from the mean of these 500 simulation runs.

Table 11

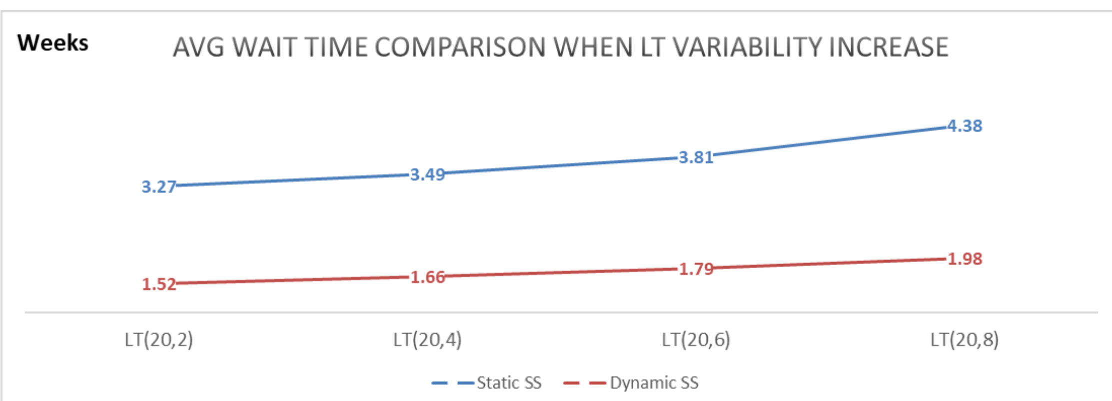
Simulation Parameter for a lead-time Variability Increasing Environment

Parameter	Value	Remark
Supplier Quoted LT	10 weeks	
Average LT	20 weeks	Derived from XGB model
STD deviation LT	2,4,6,8 weeks	
Sale Order Demand	4 pieces	Demand be generated in the 4th week of each month
Static SS	5 pieces	One number as Statis SS
Dynamic SS	8,9,10,11 pieces	Change aligns with STD deviation of delivery lead time
Simulation Time	200 Weeks	

Figure 16 presents the results of a comparison of average wait times between the static SS model and the dynamic SS model when lead-time variability increases. Notably, the dynamic SS model consistently outperforms the static SS model across all levels of lead-time deviation. This trend indicates that the dynamic SS model is more robust against fluctuations in lead-time, suggesting that it is an effective strategy for managing inventory in an environment where lead-time has great variability.

Figure 16

Average Wait Time Comparison between Static and Dynamic SS When Lead-time Variability Increase (interface component). The dynamic SS model outperforms the static SS model across all levels of lead-time deviation.



Note: The average of wait time observed across 500 simulation runs. When utilizing dynamic safety stock approach, the wait time is reduced to 50% compared with static safety stock approach.

4.6 Discussion and Recommendation

Currently, the sponsor company is using supplier-quoted lead-time as MRP parameter for production planning execution. This method proves ineffective for semiconductor commodities due to highly fluctuating supply chain variability. To address this challenge, the implementation of the XGB model has been proposed. This machine learning approach predicts both the delivery lead-time and the standard deviation of lead-time, offering a data-driven strategy to procurement planning.

The XGB model serves as a predictive tool that enhances decision making by predicting delivery lead-time. With this prediction, the procurement team can compare it with the supplier-quoted lead-time. By adopting the greater of the two as the lead-time parameter, procurement team can better safeguard against underestimating the market actual delivery time. This approach not only improves reliability in production planning but also mitigates risks associated with delayed components deliveries.

As illustrated in section 4.2, the delivery lead-time predicted by XGB model is less accurate when extreme values occur. To compensate for this, the dynamic SS model, simulated in section 4.5 was proposed to the sponsor company. The SS quantity, which is calculated from dynamic SS policy (Equation 1), can be reviewed and adjusted every six months. The dynamic SS policy should incorporate both the greater of the predicted delivery lead-time or the supplier-quoted lead-time and the standard deviation of lead-time as predicted by the XGB model. This ensures a more resilient approach to managing inventory and mitigating risks in a volatile supply chain environment.

5. Conclusion

The comprehensive analysis conducted in this study underscores the power of machine learning modeling techniques in managing supply chain risks associated with semiconductors when delivery lead-time is highly fluctuating. Through the deployment of machine learning models such as XGB and the implementation of DNN, this project highlighted the challenges in accurately predicting lead-time. Although the DNN models underperformed compared with the XGB models, they highlight the necessity for further optimization and parameter tuning for model improvement. Furthermore, the use of SHAP values and permutation importance provided valuable insights into the key factors which impact lead-time. These analyses revealed the influence of sector-specific indices and lagged features on the prediction of semiconductor delivery lead-times.

The simulation segment of this study utilized SimPy to model the safety stock strategy under varying conditions of demand and lead-time variability. The dynamic safety stock model, which incorporates the predicted delivery lead-time and the standard deviation of lead-time, proved to be more effective in mitigating risks associated with supply chains where high variability exists. This was evident in both scenarios of increasing demand variability and lead-time variability. The wait time due to component shortage could reach 15 weeks. However, this shortage was reduced to 2 weeks when applying the dynamic SS, which incorporate the predicted delivery lead-time through XGB model. Moreover, the average wait time across all time steps saw a reduction of over 70% when dynamic SS was implemented. These findings offer a framework for the sponsor company`s SSP program to enhance their supply chain resilience and responsiveness, potentially leading to reduced stockout and improved customer satisfaction.

References

Alvarez-Chaves, H., Barrero, D.F., Cobos, M., et al. (2021). Patients forecasting in emergency services by using machine learning and exogenous variables. *In: International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 167–180.

Asadzadeh, S. M., Azadeh, A., & Ziaeifar, A. (2011). A neuro-fuzzy regression algorithm for improved prediction of manufacturing lead time with machine breakdowns. *Concurrent Engineering*, 19(4), 269–281.

Bahroun, Z., & Belgacem, N. (2019). Determination of dynamic safety stocks for cyclic production schedules. *Operations Management Research*, 12, 62–93.

Becker, J., et al. (2013). Dynamic Safety-Stock Calculation*. International Scholarly and Scientific Research & Innovation,* 7(10), 2060-2063.

Corti, D., Pozzetti, A., & Zorzini, M. (2006). A capacity-driven approach to establish reliable due dates in a MTO environment. *International Journal of Production Economics*, 104(2), 536–554.

Gordon, V., Proth, J., & Chu, C. (2002). A survey of the state-of-the-art of common due date assignment and scheduling research. *European Journal of Operational Research*, 139(1), 1–25.

Gyulai, D., et al. (2018). Lead time prediction in a flow-shop environment with analytical and machine learning approaches. *In Proceedings of the 16th IFAC Symposium on Information Control Problems in Manufacturing* (pp. 1029-1034).

Hsu, S. Y., & Sha, D. Y. (2007). Due date assignment using artificial neural networks under different shop floor control strategies. *International Journal of Production Research*, 42(9), 1727–1745

Kanet, J. J., Gorman, M. F., & Stößlein, M. (2010). Dynamic planned safety stocks in supply networks. *International Journal of Production Research*, 48(22), 6859-6880.

Lingitz, L., et al. (2018*).* Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. *In Proceedings of the 51st CIRP Conference on Manufacturing Systems* (pp. 1051-1056).

Oliveira, M. B. de, et al. (2021). Lead Time Forecasting with Machine Learning Techniques for a Pharmaceutical Supply Chain. *In Proceedings of the 23rd International Conference on Enterprise Information Systems* - Volume 1 (pp. 634-641).

Sethi, F. (2020). Using Machine Learning Methods to Predict Order Lead Times. *International Journal of Sciences: Basic and Applied Research (IJSBAR),* 54(3), 87-96.

Shen, X., et al. (2022). Data-Driven Prediction of Order Lead Time in Semiconductor Supply Chain. *Operations Research Proceedings* 2022, 645-652

Wang C, Jiang P (2019) Deep neural networks based order completion time prediction by using real-time job shop RFID data. *J Intell Manuf* 30(3):1303–1318

Roos,G.(2017) *Power Semiconductors: Lead Times Continue to Stretch*. EPS News https://epsnews.com/2017/08/16/power-semiconductors-lead-times-continue-stretch/