THE ROLE OF DATA, MACHINE LEARNING, AND SUPPLY CHAIN INTERDEPENDENCIES IN IMPLEMENTING CONNECTED PACKAGING SOLUTIONS FOR ENHANCED SUPPLY CHAIN EFFICIENCY

by

Vijay Babu Palaniappan, MBA

DISSERTATION

Presented to the Swiss School of Business and Management, Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

JUNE, 2025

THE ROLE OF DATA, MACHINE LEARNING, AND SUPPLY CHAIN INTERDEPENDENCIES IN IMPLEMENTING CONNECTED PACKAGING SOLUTIONS FOR ENHANCED SUPPLY CHAIN EFFICIENCY

by

Vijay Babu Palaniappan

Supervised by

Monika Singh

APPROVED BY

Apostolos Dasilas
Dissertation chair

RECEIVED/APPROVED BY:

Renee Goldstein Osmic

Admissions Director

Dedication

This thesis is humbly dedicated to my family and friends, whose unwavering support and encouragement have been instrumental throughout my academic journey.

Acknowledgements

I would like to express my sincere gratitude to all those who have supported me throughout the course of this research.

First and foremost, I am deeply indebted to my mentor, **Ms. Monika Singh**, for her invaluable guidance, encouragement, and consistent support. Her insightful feedback and expert knowledge have played a crucial role in shaping the direction and quality of this thesis.

I would also like to thank the faculty and staff of **UpGrad** and the **Swiss School of Business and Management (SSBM)** for providing a supportive academic environment and the resources necessary to complete this work.

Special thanks are due to my colleagues and peers, whose collaboration and encouragement have been greatly appreciated throughout this journey.

Lastly, I extend my heartfelt gratitude to my family and friends for their unwavering encouragement, patience, and belief in me during every step of this endeavour.

ABSTRACT

THE ROLE OF DATA, MACHINE LEARNING, AND SUPPLY CHAIN INTERDEPENDENCIES IN IMPLEMENTING CONNECTED PACKAGING SOLUTIONS FOR ENHANCED SUPPLY CHAIN EFFICIENCY

Vijay Babu Palaniappan
2025

Dissertation Chair: <Chair's Name>
Co-Chair: <If applicable. Co-Chair's Name>

The complex SCM method combines several elements, which include demand forecasting alongside inventory management and supplier selection, and risk mitigation procedures. Artificial Intelligence integration with Supply Chain Management shows the potential to boost operational effectiveness while upgrading decision systems and minimization of operational safety threats. The study aims to investigate and apply AI methods to SCM for efficiency improvement, decision-making aids, and operational risk mitigation, using novel computational tools such as predictive and optimization models. This study proposes an integrated methodology for demand forecasting, late delivery risk prediction, and delivery status classification to enhance decision-making in SCM. The "DataCo Smart Supply Chain for Big Data Analysis"(DataCo Smart Supply Chain for Big Data Analysis Dataset, no date) dataset is used, which is sourced from Kaggle. The key preprocessing steps included missing data handling, datetime transformation, feature engineering, and variable encoding. Time series forecasting using Prophet was applied to weekly aggregated order data, while classification tasks utilized AdaBoost, Cat Boost, and MLP models. Feature importance analysis and SMOTE were employed to enhance model accuracy and address data imbalance. Model performance was evaluated using MAE, MSE, RMSE, accuracy, recall, precision, and F1 score. The proposed machine learning models demonstrated high effectiveness in optimizing key supply chain tasks. For delivery status prediction, advanced models such as AdaBoost, Cat Boost, and MLP achieved exceptional AUC scores, with Cat Boost and MLP reaching 100%. In late delivery risk prediction, these models consistently delivered strong performance, achieving AUC values above 97%. For demand forecasting, the Prophet model achieved the lowest error rates with an MAE of 0.1391 and an RMSE of 0.1612. These findings demonstrate that deep learning and ensemble methods significantly improve SCM decision-making in terms of accuracy, efficiency, and predictive power.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	Full Form
SCM	Supply Chain Management
SSCM	Sustainable Supply Chain Management
GSCM	Green Supply Chain Management
MLP	Multi-Layer Perceptron
RFID	Radio Frequency Identification
IOSP	Intelligent Or Smart Packaging
ANN	Artificial Neural Networks
ERP	Enterprise Resource Planning
ECR	Effcient Consumer Response
CRM	Customer Relationship Management
TMS	Transportation Management Systems
WMS	Warehouse Management Systems
TQM	Total Quality Management
RTLS	Real-Time Locating System
AI	Artificial Intelligence
FL	Fuzzy Logic
ABS	Urgent-Based System
MAS	Multi-Agent System
GIS	Geographic Information Systems
NFC	Near Field Communication
IoT	Internet Of Things
RL	Reinforcement Learning
ML	Machine Learning

SVM	Support Vector Machine
CART	Classification Or Regression Tree
DGI	Decrease Of Gini Impurity
DL	Deep Learning
CNN	Convolutional Neural Networks
DNN	Deep Neural Network
DBN	Deep Belief Networks
RNN	Recurrent Neural Network
EDA	Exploratory Data Analysis
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
XAI	Explainable AI

CHAPTER I:

INTRODUCTION

1.1 Introduction

In today's supply chains, the packaging business is indispensable, acting as a link between producers and buyers while guaranteeing the security, excellence, and longevity of products (Kozik, 2020). The need for cutting-edge technology that can op00timize design, guarantee quality, reduce waste, and improve operational efficiency is rising in tandem with the need for responsible and eco-friendly packaging solutions (Muhammed et al., 2025). There are two types of packaging technologies used in industries: traditional and modern or smart packaging technologies. Plastic and flexible packaging supplanted metal, paperboard, and glass as a result of the development of new materials such as ethylene vinyl alcohol polymers, polyester, and polypropylene. Radio frequency identification (RFID), intelligent or smart packaging (IOSP), and active packaging (e.g., oxygen scavengers, moisture absorbers, and antimicrobials) are among the many technical breakthroughs in packaging that have taken place in the last few decades. Product safety, quality, and longevity were all improved by these innovations (J. W. Han et al., 2018).

The goal of SCM is to maximize customer satisfaction, overall performance, and procurement efficiency by strategically coordinating an organization's business processes. SCM is essential to any company's operations, and even little data mistakes may cause big problems for customers (Radivojević et al., 2022). For example, presenting products as sold out during in-store transactions or at checkout, or advertising inaccurate inventory levels online, may have a significant influence on consumer satisfaction. Customer frustration and bad word of mouth may result in lost revenue and unfavorable social media evaluations when such mistakes occur. Consequently, it is critical to keep supply chain operations accurate and up-to-date after preserving a favorable reputation for the organization and guarantee a great client experience (Tirkolaee et al., 2021).

The selection of the most appropriate suppliers for the acquisition of raw materials is an essential part of SCM. Making smart supplier selections may boost operational performance, save procurement costs, build lasting partnerships, and satisfy customers more. Production is another important part of SCM. It entails deciding what items to sell based on market demand and the company's capabilities (Yun et al., 2020).

AI methods are the most effective means of addressing this issue with large datasets. Among the many branches of AI, machine learning (ML) is particularly well-liked for its ability to automatically detect and extract relationships between variables in massive datasets. ML algorithms are able to uncover previously unseen patterns in data, provide novel insights, and guide researchers to the most appropriate locations. ML methods train computers to automatically process massive amounts of data with greater efficiency (Dey, 2016). Traditional methods often fail when faced with enormous amounts of data that make pattern or information extraction and interpretation unfeasible. The manufacturing, quality control, and logistics sectors of supply chain management have all seen revolutionary shifts as a result of data analytics advancements. These developments allow for more efficiency and capacities to improve SCM applications via the integration of AI, ML, IoT technologies, and ANN (MacCarthy & Ivanov, 2022).

1.2 Evolution of Supply Chain Management:

The term "supply chain" refers to " an arrangement of at least three groups or persons" that are " directly engaged in the financial, informational, and/or product flows from a source to a consumer, both upstream and downstream." All parties involved in the supply chain, whether directly or indirectly, strive to satisfy a customer's request. Manufacturers, suppliers, carriers, storage facilities, merchants, and ultimately the clients themselves are examples of parties involved. The fact that a single company may participate in many supply chains highlights the interconnected character of these systems (Awasthi, 2024).

Supply networks have been influenced by technological advancements in the last forty years. The process of digitization, which involves transforming data into digital form, improves efficiency by creating a robust digital thread that connects and mirrors real supply networks. The supply chain's data, computer infrastructure, and software are being impacted by the shift towards cloud-based solutions (MacCarthy & Ivanov, 2022). An assortment of technologies may be put to use in supply chains, including sensors for monitoring, inexpensive cloud computing that boosts processing capacity, and real-time data for monitoring shipments and inventories. The overarching objective of SCM is to maximize customer satisfaction via the combined use of operational and strategic skills (Mentzer et al., 2001).

Additionally, as this technology became more and more prevalent in businesses, businesses found numerous significant sub-processes like Enterprise Resource Planning (ERP), Efficient Consumer Response (ECR), Customer Relationship Management (CRM), Transportation Management Systems (TMS), Warehouse Management Systems (WMS), Total Quality Management (TQM), and others. These programs gave them the vital chance to have more efficient and effective operations, which naturally has resulted in significant (Stoychev, 2023).

Supply chain management entered its second golden era in the 1990s. The foundational literature did not reflect any major changes at this time as SCM was still in its early phases of development and definition. Nonetheless, the external variables that have aided in the evolution of SCM over this time frame are of paramount significance.

Firstly, SCM was greatly influenced by the rise of IT and the advent of e-commerce, which allows people to buy products and services online.

The knowledge of SCM has grown significantly in the recent few years, beginning in the early 2000s, leading to fast growth in a short amount of time. This created the potential for several risks and weaknesses inside the SCM. Figure 1.1 illustrates the many connection channels in the supply chain (Mentzer et al., 2001).

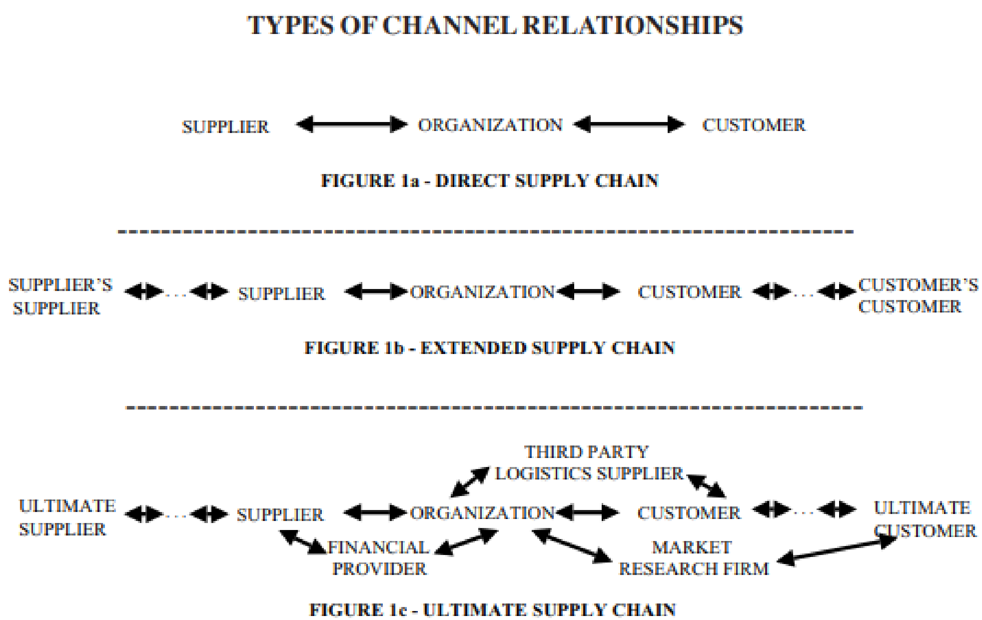

Figure 1.1: Types Of Channel Relationships

This description may be found among the three levels of supply chain complexity: "Ultimate Supply Chain," "Extended Supply Chain," and "Direct Supply Chain." Figure 1.1 a depicts a direct supply chain, which consists of three entities involved in the supply and demand for products and services as well as the transfer of money and data. Figure 1.1b shows that the upstream and downstream flows of commodities, services, money, and data include members of both the immediate supplier's and the immediate customer's supply chains. The last link in the supply chain includes every business that moves money and data from suppliers to consumers, as well as any business that produces, distributes, and uses products and services. Figure 1.1c displays the most complicated final supply chain that might be possible. A financial advisor advising and servicing both companies, a third-party logistics provider facilitating the flow of goods and services between them, and a market research firm supplying information about the final consumer to a company higher up the supply chain are all possible participants in such a scenario.

1.3 Overview Of Current Technology Trends in Supply Chain Management:

There has been a dramatic shift in the previous 20 years regarding the supply chain's usage of digital technologies. Leading supply chain solutions supplier Intermec recently conducted research that ranked the ten technologies with the most influence on supply chain operations. The technologies included in the study were: (1) thorough connection - spanning 802.11 wireless LAN protocols, cellular networks, and Bluetooth; (2) rugged PCs with built-in voice and GPS communication; (3) Speech Recognition; (4) Digital Imaging; (5) Portable Printing; (6) 2D & other bar-coding advances; (7) RFID; (8) RTLS (real-time locating system); (9) remote management; (10) wireless and device security.

It is recommended that companies and logistics service providers keep up with the latest technology advancements. Following global trends in the supply chain sector may help firms accomplish many goals, including cutting costs, improving efficiency, making better use of assets, and obtaining a competitive edge over rivals (Dong et al., 2021). Continue your research on the three primary technologies in this section: blockchain, AI, BDA, and IoT. These developments are classified as important emerging technologies and exponential trends (Hà Huy, 2021).

- **Internet of Thing:** There are several ways in which the IoT could influence the supply chain. These include SCM, increased efficiency and effectiveness, more visibility into logistics, and real-time system control. The author explains the system as a computerized recognition method by using RFID tags or RFID transporters to obtain and store data. It benefits companies as it provides more real-time accurate information of inventory systems than data retrieved by conventional method. The technology reduces labor costs by simplifying the business's processes and preventing errors (Sun, 2012).

- **Big data analytics**: Big Data Analytics have a potential topic which can be implemented by academic and management practitioners. With the availability of not only data generated from traditional equipment such as POD, RFID, and GPS but also information retrieved from digital clickstreams, camera, and surveillance footages. An extensive web of interconnected technologies, according to the author, is facilitating real-time data collection across the supply chain, which in turn increases transparency throughout the logistics chain (Fosso Wamba et al., 2018) points out that companies in the pharmaceutical industry have already been utilizing this technology for detecting demand spikes, stocks, and delivery resources beforehand
- **Artificial Intelligence (AI):** AI in the logistics industry as a system which generates better outputs when problem-solving with higher accuracy, higher speed, and a capability to process more input information. The technology is defined as an emerging competitive advantage for enterprises in the SCM industry, and many companies are switching from remote monitoring to an AI-based system, which gives access to more control, optimization, and functionality. 3 AI techniques which are being used widely in the supply chain industry can be Artificial Neural Networks (ANNs), Fuzzy Logic (FL), and Urgent-based System/ Multi-agent System (ABS/ MAS).

New Developments in Supply Chain Management:

1. **Global Supply Chains:** The difficulty of maintaining an international supplier connection is something that even a small manufacturing business may encounter in today's more globalized global economy. While the concept of globalization has been around for a while, the trend that emerged in the early 2000s caused the globe to become flatter than at any point in history. Unfortunately, the gap in prices between rich and developing nations has been narrowing in recent years. As a result, many businesses are rethinking their supply chain configurations, particularly their cost-driven offshore plans. Businesses of all sizes in the industrial sector have begun to reshore their offshored operations in recent years.
2. **Sustainable Supply Chains:** Sustainable supply chains have gained significant attention over the past decade, driven by growing environmental concerns, regulatory pressures, and increasing consumer demand for eco-friendly practices. This society-wide sustainability movement has pushed many businesses to rethink and redesign their supply chain strategies to minimize their environmental footprint while promoting social and economic responsibility. Environmentally conscious sourcing, packaging, transportation, and waste reduction are some of the techniques that businesses are embracing more and more. These initiatives reflect a broader industry trend where firms are prioritizing circular economy models, promoting reusable and recyclable materials, and implementing carbon footprint reduction programs. Additionally, the integration of technologies like blockchain and IoT in sustainable supply chains enhances transparency and

traceability, enabling businesses to monitor and report on ethical sourcing, fair labor practices, and environmental compliance. (Lu & Swaminathan, 2015)

3. **Humanitarian Supply Chains:** Worldwide, 6,637 natural disasters occurred between 1974 and 2003, affecting over 5.1 billion people and causing $ 1.38 trillion in reported damage (Ergun et al., 2009). The social and financial effects may have been mitigated with forethought, even if these occurrences were difficult to predict. After disasters, quick reaction and relief activities are possible with a well-planned and frequently maintained humanitarian supply network. When compared to conventional supply networks, humanitarian supply chains have unique patterns of demand and supply. Cooperation in real-time between humanitarian organizations, governments, and NGOs is essential for effective HSCs, as is the quick mobilization of resources. Advanced technologies, such as geographic information systems (GIS), blockchain for traceability, and AI-powered predictive analytics, are increasingly being used to improve disaster forecasting, optimize resource allocation, and enhance supply chain visibility.

Applications of Supply Chain Management:

Anticipating and satisfying customer demands, managing the supply and storage of products and raw materials, processing orders, and arranging for delivery are all essential objectives. Companies are using cutting-edge technology like blockchain, AI, and ML to simplify operations and enhance decision-making in their data-driven and more intricate supply chains. Businesses can save costs, increase transparency, and improve overall performance with the use of these technologies in critical areas including demand forecasting and planning, procurement, manufacturing and inventory management, logistics, and transportation. This section delves into the many uses of SCM, primarily focusing on how advancements powered by ML and AI are revolutionizing conventional supply chain operations into more efficient, data-driven ecosystems. This overview delves into the many fields of AI and how it is connected to the supply chain.

- **Demand forecasting and planning:** A number of critical company processes rely on accurate demand forecasts and plans. These processes include marketing, production scheduling, new product creation, inventory management, and research and development. By accurately predicting future demand, businesses may better allocate resources, boost operational efficiency, and increase customer happiness and profitability. In inventory management, reliable demand predictions help minimize holding costs, prevent stockouts, and optimize replenishment strategies. For new product development, demand forecasting identifies emerging market needs, optimizes product launch timing, and reduces market entry risks by predicting potential demand scenarios. By leveraging ML models and predictive analytics, businesses can enhance the accuracy of their forecasts, enabling smarter decision-making, reducing supply chain disruptions, and driving growth (Kumari et al., 2023).
- **Key Management Issues:** Nevertheless, when it comes to novel goods or services without a history, the accuracy of conventional forecasting tools like Box-Jenkins, time series, and moving averages is severely limited. The use of AI algorithms in conjunction with more traditional approaches to improve the precision of demand planning, especially in the era of quickly changing data. Research has examined the efficacy of ANN in a variety of fields, including oil extraction, electricity generation from solar power plants, and future prediction, among many others, for the purposes of demand planning and forecasting. ML in demand forecasting and planning solves a lot of problems in the supply chain,

such the bullwhip effect, erroneous demand information estimates, and the difficulty to use social media data to make true predictions. Next, supply chain planning makes heavy use of ML techniques (Wenzel et al., 2019).

- **Procurement:** Businesses must decide whether to manufacture in-house or outsource depending on several criteria, including production capacity, core competencies, supplier demand, and procurement, which entails locating and acquiring raw materials, supplies, services, and other things required for production or sale. Problems arise when trying to make such judgements while dealing with supplier relationships, budget restrictions, and sourcing techniques. To tackle these issues, scientists are looking at ways to improve supplier management and procurement with artificial intelligence. To evaluate the dependability of suppliers and their sustainability performance, some research has used ANNs, while others have utilized Bayesian learning risk dependency modelling. These methods try to make procurement procedures better and provide businesses more information to work with in the complicated and ever-changing world of supply chains.
- **Production and inventory management:** Managing both production and inventory includes a wide range of tasks, from initial planning to the final delivery of components. An essential part of production management that tries to satisfy service goals while lowering production costs is the efficient scheduling of services and resources. Each level of production management has made heavy use of AI approaches to optimize it (Garvey et al., 2015). For example, ML models have been successful in hierarchical SC scheduling, finding the best batch sizes, and predicting output and demand levels. ML provides solutions for order management, inventory management, and production management by estimating

delivery dates, optimizing reorder points and safety stocks, and identifying variables impacting inventory. Improving inventory management and lessening the impact of the bullwhip effect are two applications of rough set theory.

- **Logistics and Transportation Management:** Additionally, by combining IoT with blockchain, transparency may be provided. At various periods in time, various supply chain participants, including the carrier, may package the data that the vehicle continues to broadcast, which includes its position and other parameters. This allows you to chain the blocks. The data may be stamped, encrypted, and delivered to a member's network node to be stored in that member's blockchain when the car is in range of that member's IoT devices. This manner, members can keep the data that own together secure while also helping one other verify the content of newcomers to the network. One other thing that can be done with AI is to provide personalized commute times (Mallesham, 2022).

1.4 Definition and Concept of packaging

Packaging is an important step in ensuring that food products retain their quality throughout their life cycle, from storage to transportation to consumption. The marketing and distribution processes become more efficient, and the quality is kept from declining. Protection, communication, convenience, and confinement are the four fundamental purposes of packaging. Product packaging serves multiple purposes: first, it shields goods from the elements; second, it conveys information to the buyer through copy, graphics, and logos; third, it caters to the buyer's lifestyle, whether that's saving time (in the case of ready-to-eat and heat-and-eat meals) or making the food easier to handle (in the case of resealable packaging, microwavable, and easy opening); and last, it optimizes logistical efficiency by housing products of varying shapes and sizes (J. H. Han, 2005).

As far back as there has been data tracking sales and customer loyalty, packaging design has been a major factor. Thoughts like selecting colors and materials, as well as whether to use descriptive or non-descriptive trademarks, are crucial because of the limited space and time available. There are two types of packaging, first is traditional packaging system and second is modern packaging system (Biji et al., 2015). The traditional packaging system refers to the conventional methods and materials used for enclosing and safeguarding products throughout the supply chain. It typically includes primary, secondary, and tertiary packaging, using materials such as plastic, glass, paper, metal, and wood (Chitra et al., 2022). Despite its effectiveness in product preservation and marketing, traditional packaging faces limitations in terms of sustainability, traceability, and interactivity. Discuss below in using material for Traditional packaging system:

1. **Glass:** Glass is a popular choice for processing items' packaging because of the inherent moisture and oxygen barriers it provides.
2. **Paper and cardboard:** The production of paper and cardboard involves the utilization of wood, plants, recycled paper, and cardboard waste (11). Paper is the material of choice in the food sector because to its quality, smoothness, and environmental label, as well as the treatment it receives from pulp and paper.
3. **Metal:** Cans, foil wraps, retort pouches, and tin plates are some examples of metal-based packaging items. Other common metals used for food and other packaging purposes include stainless steel, tin-free steel, aluminum, and tin plate.
4. **Plastic**: Many different types of food and other product packaging utilize plastic. Plastics have largely replaced more traditional materials in food packing

containers, which has greatly improved food preservation. The growing demand for food and other products, driven by increasing incomes and population, has led to an increase in the use of plastic packaging.

The next stage of interactive packaging is connected packaging, which allows packaging to actively support a brand. As the packaging is already a kind of brand promotion, this is a brilliant move that allows you to interact with consumers via a location that you control. Connected packaging is great since it can be done via a cell phone, which almost everyone always has on them. Interconnected, perceptive, dynamic, or intellectual. Therefore, you could be familiar with comparable terminology if you've dabbled in this field. Let's clarify the differences (Biji et al., 2015).

- **Connected packaging:** Connected packaging leverages QR codes, NFC tags, RFID chips, and other digital technologies to create an interactive bridge between physical products and digital experiences. By scanning the packaging with their mobile phones, customers can access a wealth of information, including product details, usage instructions, promotional offers, and authenticity verification. Supply chain optimization relies heavily on linked packaging, which allows for real-time monitoring and traceability in addition to customer involvement. It allows manufacturers and retailers to monitor product location, condition, and handling throughout the distribution process, ensuring quality control and reducing counterfeiting risks. Additionally, connected packaging enhances brand transparency by providing customers with insights into sourcing, sustainability practices, and certifications, fostering trust and loyalty. As AI and ML continue to evolve, connected packaging is increasingly used for personalized marketing, offering tailored content and recommendations based on consumer behavior, making it a powerful tool for both customer engagement and supply chain efficiency.

- **Active packaging**: Adding chemicals to packaging is known as "active packaging," and it extends the shelf life of items. The food business makes use of five distinct types of active packaging: (1) scavenging of flavors, UV radiation, moisture, ethylene, carbon dioxide, and oxygen; (2) discharge of ethanol, food preservatives, antioxidants, Sulphur dioxide, or flavors; (3) exclusion of substances from food, such cholesterol or lactose; (4) insulation material temperature regulation, temperature-sensitive packaging, and (5) UV light is used for microbiological and quality control purposes.
- **Intelligent packaging**: Food, drink, and pharmaceutical items are often packaged using intelligent packaging, one kind of smart packaging. Despite its association with the food industry, intelligent packaging has no bearing on the product itself. the capability of communicating the product's packaging conditions without actually handling the goods. The objective is to keep tabs on the product and provide the consumer with updates. The term "intelligent packaging" refers to containers that use sensors to monitor the contents' condition. Indicators, sensors, and RFID tags are the three primary components of intelligent packaging systems (Ghaani et al., 2016). Indicators that plot time against temperature, those that plot color against temperature, those that plot pathogen markers, and those that plot refraction are all examples of such systems.
- **Smart packaging**: The phrase "smart packaging" encompasses a wide range of products whose containers and wrappings serve purposes beyond those of a simple container. It encompasses technologically enhanced packaging solutions that monitor, maintain, or engage with the product or consumer, adding value

throughout the product lifecycle. Smart packaging integrates IoT sensors, RFID tags, NFC chips, and QR codes to provide real-time data on product authenticity, condition, and traceability. In supply chain management, it enables continuous monitoring of temperature, humidity, and handling conditions, ensuring product integrity, especially for perishable goods and pharmaceuticals. On the consumer side, smart packaging enhances user engagement through interactive features, such as personalized content, promotional offers, or product information accessed via smartphone scanning. Furthermore, it plays a key role in anti-counterfeiting by verifying product authenticity through blockchain-backed labels. As technology advances, smart packaging is increasingly being used to reduce waste, improve sustainability, and enhance customer experiences, making it a critical component of modern supply chain and marketing strategies.

Businesses are under growing pressure to implement eco-friendly policies and procedures as the sustainability conversation has spread throughout society. As an integral aspect of the international Supply Chain, the packaging sector is one area that receives special attention. The destructive effects on the environment caused by conventional packaging techniques, which use non-renewable resources and produce a lot of trash, have recently drawn attention to these issues. Therefore, there is a growing chorus of voices demanding that businesses include sustainable packaging practices into their broader CSR and environmental governance programs (Gupta, 2024). Sustainability in packaging has far-reaching consequences for the whole value chain, making it an increasingly important consideration in supply chain contexts. The manufacture, distribution, and consumption of commodities are all interdependent processes, and packaging is an essential part of this network since it allows for the protection of products during transit, branding, and identification. There is a gap between corporate imperatives and environmental stewardship since traditional packaging strategies sometimes put short-term efficiency and cost concerns ahead of long-term environmental sustainability (Asim et al., 2022).

Key smart technologies for connected packaging:

SCM is seeing a sea change with the advent of connected packaging, which improves efficiency, transparency, and customer interaction. Offering real-time data tracking, product authentication, and improved operational efficiency. Unlike traditional packaging, which primarily serves as a protective and marketing tool, connected packaging integrates smart technologies like QR codes, RFID tags, NFC chips, and IoT sensors.

- **QR codes and Barcodes**: Products are better protected and of higher quality with the help of automated identification devices like these, which also make it easier to share data across the packaging and supply chain. These labels, also known as data carriers, aren't supposed to tell you anything about the product's state; their job is to automate processes, make them easier to track, and prevent theft or counterfeiting. It often sticks to the exterior of a product's box instead than the actual item. QR codes and barcode labels are the most used methods. Retailers and businesses often use QR codes and barcodes for inventory checks and stock monitoring because of their affordability and user-friendliness. In order to represent data with 8 to 12 digits, a barcode uses an arrangement of parallel gaps and bars. An optical barcode scanner reads the encoded data and transmits it to a database for storage and processing (Ozcan, 2020).
- **Radio-frequency identification (RFID) tags:** RFID is a technology that uses wireless sensors to auto-identify items and gather data (Tajima, 2007). Tags and readers are the basis of RFID. The majority of RFID tags have some kind of

identifier that, when scanned, allows the reader to access related data stored in a database and take appropriate action. The product's packaging isn't the only thing that typically has them attached; the outside layer of any given box, shipment, pallet, etc. RFID labels are the most commonly used approach. A data carrier tag may be as sophisticated as an RFID tag. There are primarily three parts to an RFID system: a) a tiny antenna attached to a microchip that makes up a tag, b) a device that sends out radio waves and receives replies from a tag, and c) a web server or local area network that the middleware uses to communicate with the RFID (Ozcan, 2020).

- **Near field Communication (NFC) chips:** Various wireless technologies are being used more and more to obtain different kinds of data. Near-Field Communication (NFC) is one such example. Allowing for product verification and consumer contact via short-range wireless communication (Gegeckienė et al., 2022).
- **Internet of Things (IoT) sensors:** A combination of smart product packaging and the IoT has transformed brand-consumer interactions. Disruptive technologies that change how items are exhibited, marketed, and experienced. Transport and storage conditions may be tracked with the use of embedded sensors that measure things like humidity and temperature (Soundarraj et al., 2023).

IoT packaging includes features such as temperature-sensitive medicinal indicators, logistics monitoring facilitated by RFID, and consumer involvement and product verification QR codes. Businesses may boost productivity, save costs, and satisfy customers more with the aid of these technologies. CPOs use smart packaging to accelerate supply chain change. These developments guarantee compliance, improve visibility, and cut down on inefficiencies. Additionally, by reducing waste and maximizing material use, smart packaging supports environmental objectives. Smart packaging is an investment that may pay off for businesses in the long run by cutting costs and giving them an advantage in the supply chain efficiency race.

1.5 Transforming Supply Chains with Smart Packaging Technologies.

Smart packaging is revolutionizing supply chain efficiency by using sophisticated technologies like RFID, IoT, blockchain, NFC tags, and QR codes. These technologies enable real-time tracking, improved inventory management, and greater transparency throughout the supply chain. By providing continuous monitoring of shipment conditions, product location, and handling, smart packaging enhances visibility and traceability, reducing stockouts, spoilage, and theft. Additionally, it drives cost savings by optimizing warehouse operations, lowering return rates, and reducing warranty claims.

- **Real-Time Tracking and Transparency:** Technology like blockchain, RFID, and the IoT allow for smart packaging to monitor items in real-time and increase visibility all the way through the supply chain. These technologies provide continuous monitoring of shipments, product conditions, and location data, ensuring improved visibility and traceability.
- **Improved Inventory Management**: Smart packaging automates inventory tracking, reducing stockouts and overstocking. IoT-enabled systems enhance accuracy and streamline warehouse operations. Report a 40% reduction in stockouts, while highlighting improved warehouse efficiency through RFID-based tracking.
- **Cost Reduction:** Smart packaging reduces spoilage, theft, and operational costs. Report a 28% reduction in spoilage through temperature-sensitive packaging. Additionally, highlight that condition-monitoring sensors lower return rates and warranty claims.

- **Sustainability Gains:** Smart packaging supports sustainability by reducing waste and promoting reusability. Found that reusable RFID tags cut single-use packaging waste by 30%. In emphasize the role of biodegradable smart labels in eco-friendly packaging (Morashti et al., 2022).
- **Enhanced Customer Experience:** Interactive smart packaging, such as NFC tags and QR codes, boosts customer engagement and trust. It highlights that NFC tags enhance authenticity verification while demonstrating that QR codes improve customer experience with real-time updates (Karpavice et al., 2023).

Applications of SCM in Connected Packaging:

Connected packaging enhances SCM by leveraging IoT sensors and real-time data collection to improve product monitoring, traceability, and inventory management. It ensures product quality and safety through continuous condition tracking, especially for perishable goods, while enhancing transparency by providing detailed traceability data. Businesses may save money and make customers happier by optimizing inventory levels, reducing stockouts, and streamlining transportation via the integration of linked packaging with supply chain management systems. Additionally, it supports sustainability efforts by reducing packaging waste and promoting eco-friendly practices.

- **Real-Time Monitoring and Data Collection:** IoT sensors included in connected packaging allow for constant tracking of product parameters like humidity and temperature all the way through the supply chain. This real-time data collection ensures product quality and safety, particularly for perishable goods (Chen et al., 2020).
- **Enhanced Traceability:** Smart packaging solutions provide detailed tracking information, allowing stakeholders to trace products from origin to destination.

This increased transparency helps in mitigating issues related to counterfeiting and ensures compliance with regulatory standards (Armstrong et al., 2018).

- **Optimized Inventory Management:** By integrating connected packaging data into SCM systems, companies can achieve more accurate inventory levels, reducing overstock situations and stockouts. This leads to better demand forecasting and replenishment strategies.
- **Improved Logistics Efficiency:** Connected packaging facilitates better coordination among supply chain partners, leading to optimized transportation routes and reduced lead times. This efficiency contributes to cost savings and improved customer satisfaction.
- **Sustainability Initiatives:** Implementing smart packaging solutions aids in monitoring and managing packaging waste, promoting sustainable practices within the supply chain. This may improve a brand's reputation and is in line with social objectives (Deborah & Sugihartanto, 2024).

1.6 Artificial Intelligence in Supply Chain Management

ML, cognitive analytics, and NLP are all parts of AI. These technologies aim to imitate human cognitive capabilities. Applying AI to supply chain management allows for the analysis of massive datasets, the optimization of decision-making procedures, and the prediction of future trends. Algorithms powered by AI may simplify logistical processes, increase the accuracy of forecasts, and better manage inventories. AI allows businesses to better understand their supply chain processes, spot inefficiencies, and boost performance using data-driven solutions (Yashan et al., 2024). The AI techniques are described below:

Machine learning

The next part defines "Machine Learning" and discusses the many kinds of ML techniques. As a result, there is more consensus on the approaches employed by the various ML application domains in SCM-connected packaging. Additionally, difficulties are handled.

ML is a branch of AI that enables computational entities (such as algorithms, software, or systems) to acquire new knowledge and skills automatically, without human intervention. Data or observations are usually utilized to build a computer model in ML. The model then analyzes various patterns of data, together with actual and expected results, to enhance the technology's functioning. The algorithm-based ML models are very good at sifting through large datasets for patterns, outliers and predicting insights. The strong qualities of this solution make it an ideal choice for addressing critical issues in the supply chain industry. We can identify four primary groups of algorithms, each of which has a different type of outcome and how much data is used during the learning or training phase (Mohamed-Iliasse et al., 2022). Figure 1.2 illustrates the many learning paradigms that define these families:

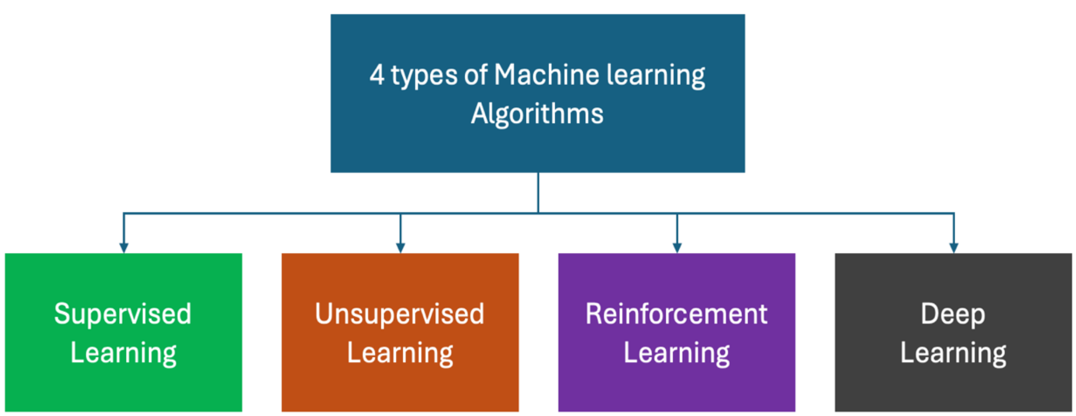

Figure 1.2: Types of machine learning algorithms

1. **Supervised learning**: The term "Supervised Learning" is used to describe a system that gives input data and predicted output data in the context of ML and AI. Classification labels are applied to both the input and output data after setting up a foundation for learning from further data processing. Thus, as seen in figure 1.3, this learning procedure relies on comparing the expected and computed results; learning is characterized as determining the mistake and adjusting the mistake to get the target result. Applied supervised learning finds utility in many contexts across the supply chain, including the automated, real-time location monitoring of items in packaging, status or surveillance zones, and others. The development of patient-specific detectors for the rapid identification of the start of epileptic seizures, which allows for the avoidance of bodily harm and even death, is another well-known use of supervised learning (Pugliese et al., 2021). A supervised ML model is a parameterized function fpfp which can show the input data $\vec{x} \in X^d$ to the output data $y \in Y$. Typically, a vector of characteristics is used to define the input data. For a classification, X^d is a vector space with dimensions d and 𝕐 is the set of classes.

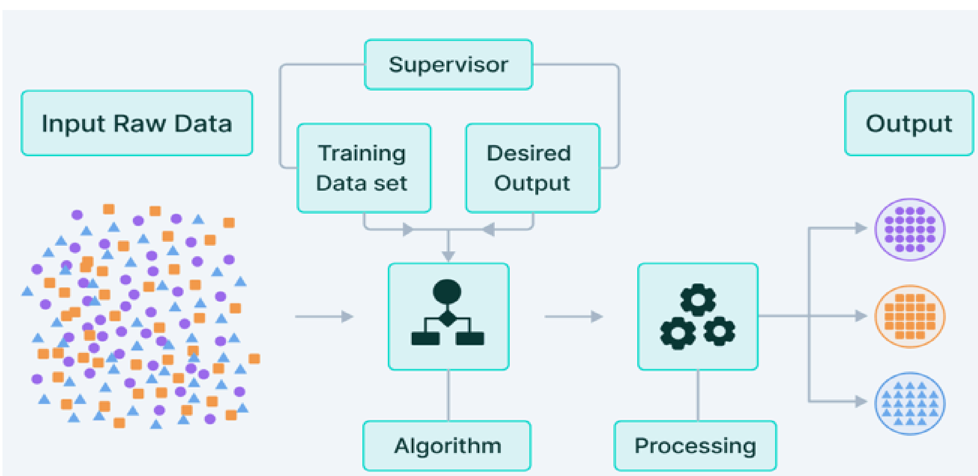

Figure 1.3: Structure of supervised learning

2. **Unsupervised Learning:** Figure 1.4 depicts an unsupervised learning scenario, as contrast to supervised learning, in which examples are annotated. In this case, the system must function using unannotated data. Similar to how animals and people can automatically create categories to match with input data, it must be able to identify that form as a circle or a square. Clustering, also known as segmentation, is the most prevalent unsupervised learning task. In this kind of learning, the goal is to group data into different types of categories, such as forms in photos. Anomaly detection is highly anticipated for applications such as predictive maintenance, cybersecurity, early illness diagnosis, etc. (Liu et al., 2021). Typically, the goal of the algorithm is to ensure that the data within each group is as similar as possible while also creating as many unique groups as possible. Various algorithms are used to classify the data based on factors such as density or density gradient in different contexts. The excessive or unusual nature of the numbers or of a pattern in the data is what is sought in the anomaly detection instance. A parameterized function $g\theta$ that takes in data $\vec{x} \in X^d$ but does not provide any outputs is known as an unsupervised ML model. The features vector serves as the model's input data (Zaadnoordijk et al., 2022). In addition, X^d is a *d*-dimensional vector space.

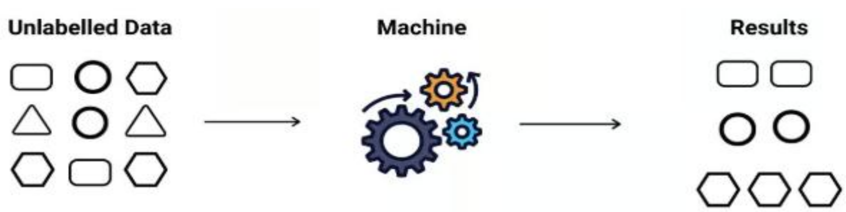

Figure 1.4: Structure of Unsupervised learning

3. **Semi-supervised learning:** Data that is both labelled and unlabeled may be used to train algorithms in semi-supervised learning models. It creates predictions or classifications using a smaller collection of labelled samples in conjunction with a

bigger pool of unlabeled data, as shown in Figure 1.5. Labelling data may be a time-consuming or costly ordeal, but these approaches make the most of constrained resources. In most cases, the accuracy of semi-supervised learning models is higher than that of fully supervised ones. Complex tasks that need a large quantity of labelled data may be too much for it to handle during training, and it is very dependent on the accuracy of the initial labels, which might create bias. In mathematics, semi-supervised learning methods boil down to the following ideas. The collection of training samples in semi-supervised models is described as $L= \{(Xi, Yi) \mid Xi\in Rd, Yi \in \Omega, i=1\ldots, L\}$. A d-dimensional feature space is what makes up each sample in these models, where Xi is the input sample, Yi is the class label Xi, and $\Omega= \{T1\ldots\ldots\ldots TK\}$ presents the target classes. Furthermore, the collection of unlabeled data is denoted as U = $\{X*j|\, j=1\ldots\ldots\ldots, u\}$. There are many types of semi-supervised models, including regression, clustering, and dimensionality reduction.

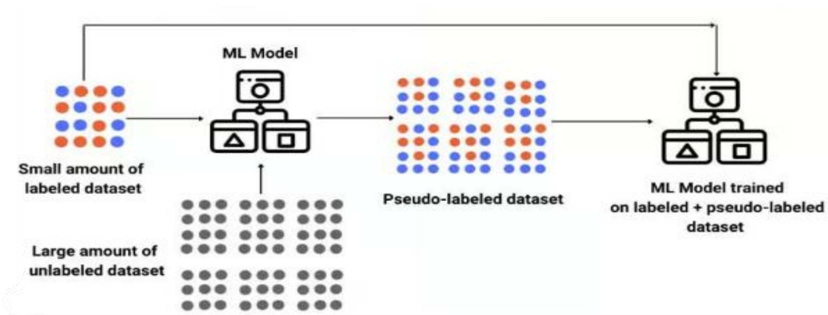

Figure 1.5: Structure of semi–supervised learning

4. **Reinforcement Learning:** Reinforcement learning (RL) models are another kind of machine learning (ML) that uses feedback from an agent's own actions and experiences to teach it to learn in an interactive setting. Reinforcement learning teaches the models how to behave in a way that maximizes rewards in a given environment. Through interactions with the environment and their reactions, these

ideal activities may be carried out. The learner must independently investigate the order of their activities in the absence of a supervisor to maximize the rewards (Polydoros & Nalpantidis, 2017) as shown in figure 1.6. Robot control and game playing are two examples of applications that might benefit from reinforcement learning's capacity to deal with complex and dynamic situations. It also performs very well when input is either delayed or in short supply. Typically, an agent will try to simulate the MDP to learn decision-making issues progressively. At a timestep *t*, an agent gets a notice about a setting $P_t \in P$, where *P* denotes a space of states, and Equation (1.1) states that the agent may choose a behaviour $B_t \in R(P_t)$ depending on the response of the environment.

$$D_t = \sum_{k=0}^{\infty} \quad Y^k R_{t+k+1} \qquad 0 \leq Y \leq 1 \qquad (1.1)$$

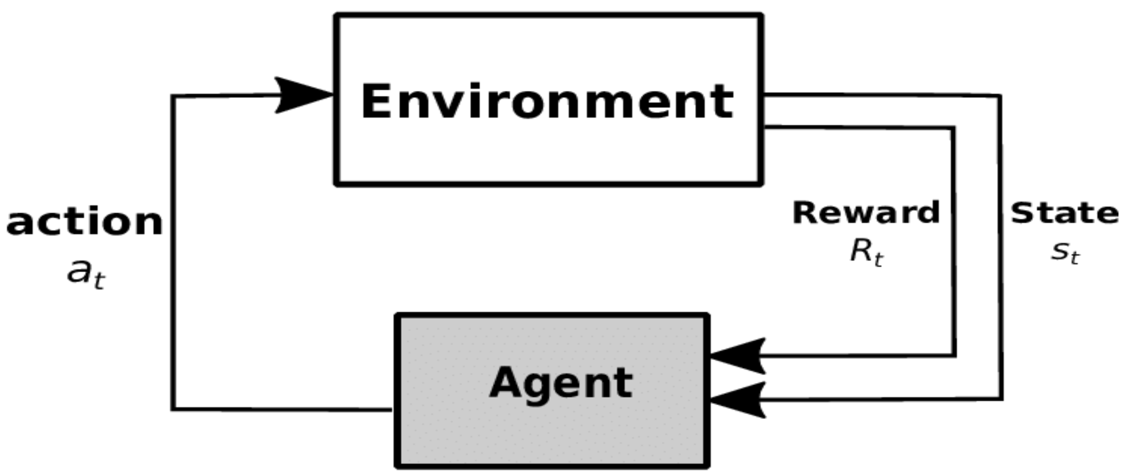

Figure 1.6: Structure of Reinforcement learning

Types of Machine Learning models:

1. **Support Vector Machine:** The SVM is a popular ML technique that is well-known for its efficiency in tasks including outlier identification, regression, and classification. SVM, which is based on statistical learning theory, was first developed for binary classification issues. Nevertheless, it has since been

extended to manage multi-class classification with techniques such as one-versus-all and one-versus-one. SVM finds the optimum hyperplane to maximally partition data points from different classes. Support vectors, or data points closest to the hyperplane, are used to establish the border; this is particularly helpful for complex and high-dimensional datasets. SVMs may handle increasingly complex problems by transforming non-linearly separable data into spaces with extra dimensions using kernel functions, which can be linear, polynomial, or radial basis functions. The SVM method finds the best decision border that divides two classes with the biggest margin, as shown in Figure 1.7, a simplified visual depiction of a binary classification issue addressed by the algorithm. The reliability of SVMs makes them a go-to for many image recognition, bioinformatics, and financial forecasting tasks, particularly when dealing with datasets of moderate size that have distinct categories (Guido et al., 2024).

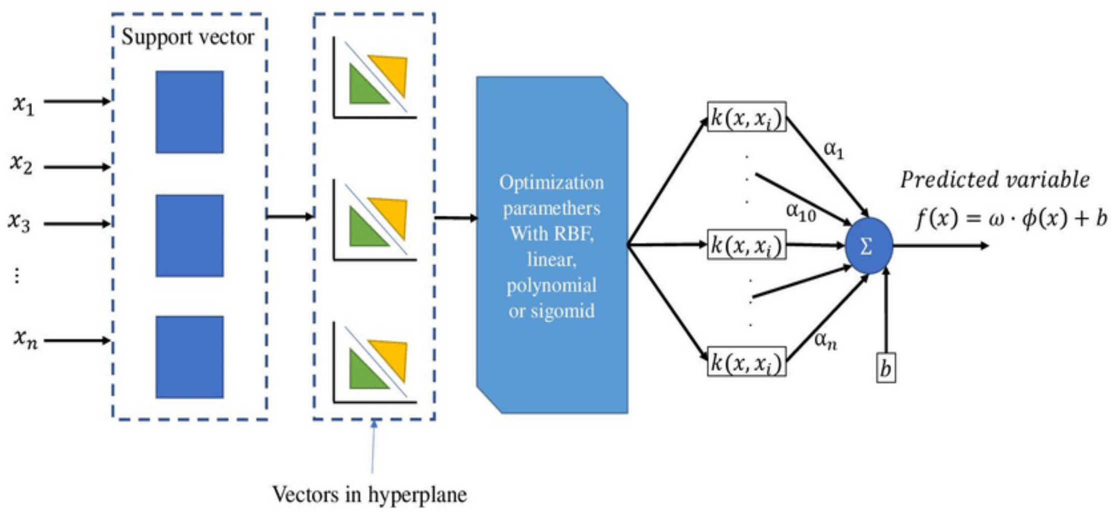

Figure 1.7: General Architecture of SVM

2. **Decision tree:** DT are among the most successful techniques in numerous disciplines, including as pattern recognition, image processing, and ML. The DT model is a sequential one, and it efficiently and unified incorporates a collection

of basic tests, where each test compares a numerical feature to a threshold value. The mathematical weights in the neural network of links between nodes are much more difficult to create than the conceptual principles. The primary use of DT is in the realm of classification. The DT classification approach is also widely used in the Data Mining domain (Jijo & Abdulazeez, 2021). Nodes and branches are characteristic features of all trees. Considering that each subset specifies a potential node value, the network may be seen as expressing attributes within a category that has to be classified. Decision trees have found various application domains due to their easy analysis and accuracy on many types of input. A DT example is shown in Figure 1.8.

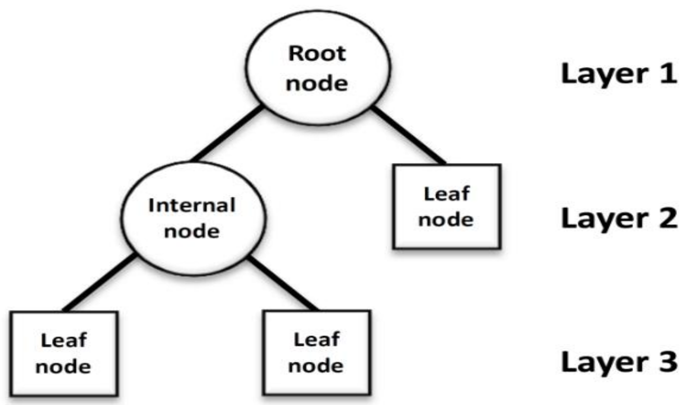

Figure 1.8: Structure of decision tree

3. **Random Forest:** The RF method employs a network of multiple decision trees for classification and regression. For future observations, it uses a collection of trained trees that have been internally verified to forecast how the predictors will respond. Figure 1.9 shows the overall operation of the RF algorithm. Every tree in the original RF approach is a regular Classification or Regression Tree (CART). The so-called Decrease of Gini Impurity (DGI) is used as a splitting

criterion, and the splitting predictor is chosen from a randomly selected group of predictors, which changes at each split.

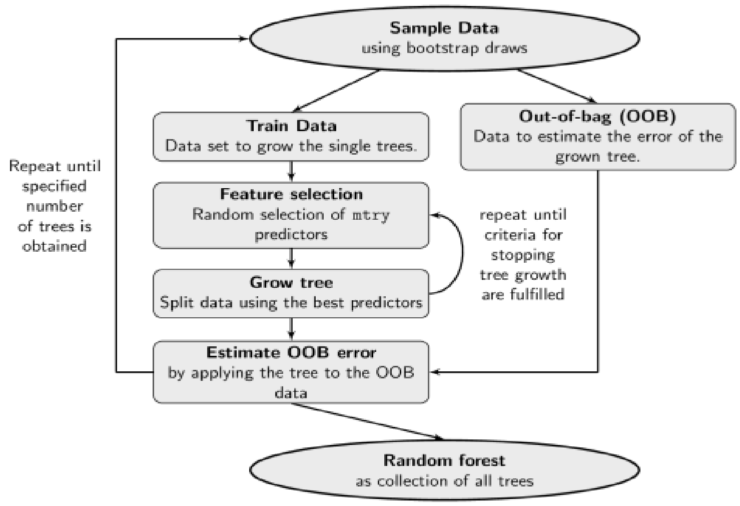

Figure 1.9: Random Forest Algorithm

4. **Linear Regression:** A linear relationship exists among the independent variable X and the dependent variable Y, as per the linear regression statistical process. Figure 1.10 demonstrates the objective of finding an optimum linear function, which is to say, a collection of weights that will allow the function to predict the dependent variable's value as precisely as feasible (K. Qu, 2024). Equation (1.2) is the formal expression of the linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \qquad (1.2)$$

the independent variable X1 through Xn, the intercept β0, the regression coefficients β1 through βn, and the error term ε

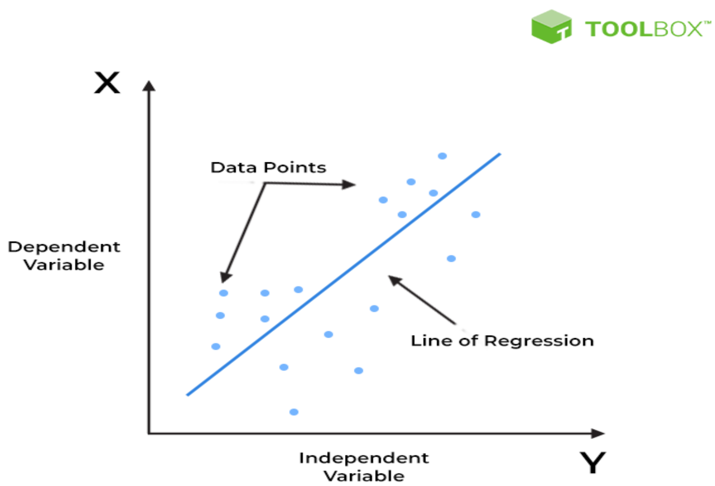

Figure 1.10: Sample of Linear regression

5. **K-Mean Clustering**: To address clustering issues, K-means clustering groups data points into K separate clusters; it is an unsupervised learning approach that has seen extensive application. At the outset, the method chooses K centroids at random, one for each cluster. The placement of these centroids is crucial, as different initial positions can lead to varying results shows figure 1.11. To enhance accuracy, the centroids should ideally be placed far apart to maximize the separation of distinct groups. Clusters are formed as a result of the algorithm iteratively assigning data points to the centroid that is closest to them. By averaging the points in each cluster, the centroids are recalculated after all points have been allocated. Once convergence is achieved, either the centroids stop moving noticeably or the total clustering variance is minimized. Then the procedure ends (Mahesh, 2020).

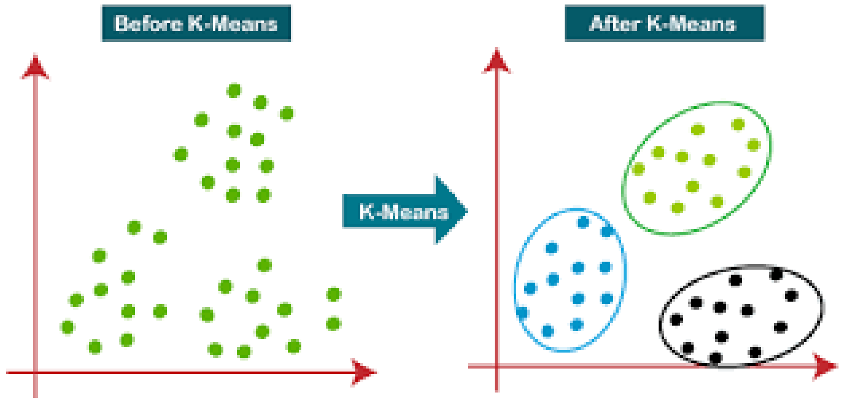

Figure 1.11: Sample of K–Mean Clustering

Deep learning

Deep learning (DL) has revolutionized AI by outperforming traditional methods in a variety of contexts and showing remarkable resilience when faced with large datasets and complicated calculations. A kind of machine learning known as "deep learning" trains ANNs with several layers (deep architectures) to learn data representations by extracting hierarchical features. Data inputs such as images, text, time series, and sensor readings may be automatically processed by deep learning models to produce complex patterns and representations. This contrasts with typical ML techniques, which often need customised feature engineering and domain-specific expertise (Eni et al., 2024). Deep learning uses several types of models in the present time like CNN, ANN, DBN, DNN, and other models to perform in different applications and recognize data. Some models to discuss below shown figure 1.12.

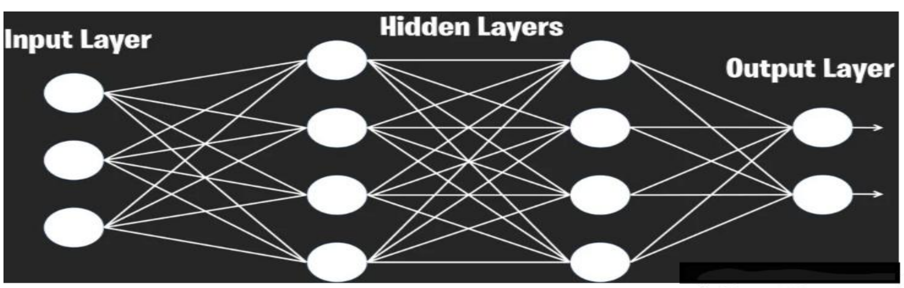

Figure 1.12: Architecture of Deep Learning

Several application disciplines, such as computer vision, NLP, voice recognition, and reinforcement learning, have seen outstanding success with DL algorithms. Convolutional neural networks (CNNs) are particularly well-suited for tasks involving spatial data, such as image recognition and object detection, due to their ability to exploit local spatial correlations through shared weights and pooling operations. In contrast, RNNs keep state information over time steps and capture temporal relationships, making them ideal for processing sequential data like time series and text.

Deep Learning Architecture Types

1. **Artificial Neural Networks:** A biological representation of the brain served as the basis for the lexicon of artificial neural networks. The building blocks of any neural network are the neurons, which are groups of linked cells. Neurons, as seen in Figure 1.13, take impulses from input cells or other neurons, process them in some way, and then send the result to either other neurons or output cells. (Jawad, 2023). The neurons in a neural network are organized into layers, with each layer receiving input from the one below it and sending their output to the one underneath it. For mathematicians, a Multi-Layer Perceptron network is just a function that takes the weighted average of all the functions that correspond to neurons.

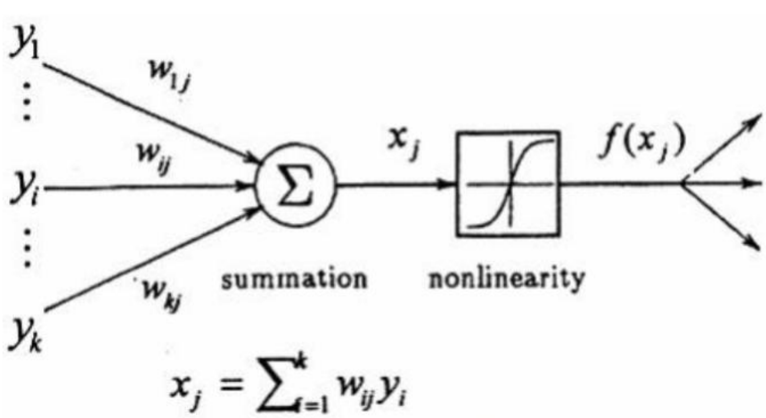

Figure 1.13: Structure of ANN

2. **Convolutional Neural Networks (CNNs):** Popular among discriminative deep learning architectures, the CNN learns from input alone, without the assistance of humans required for feature extraction. Multiple convolution and pooling layers are seen in Figure 1.14, which is a CNN example. Traditional ANNs, like regularized MLP networks, benefit from CNN's improved architecture. The level of model complexity and the ideal parameters for creating meaningful output are both considered by each layer in CNNs (Lecun et al., 1998). Furthermore, CNN uses a 'dropout' to prevent the problem of overfitting that may occur in traditional networks. The versatility of CNNs to process a wide variety of 2D shapes makes them ideal for use in many fields, including visual identification, medical image analysis, image segmentation, NLP, and many more. The capacity to automatically identify key elements from the input without human involvement is what gives it its superiority over a conventional network. The visual geometry group (VGG), ResNet, Xception, Inception, Alex Net, and other CNN variations are available in the field. Their learning capabilities determine their suitability for different application areas.

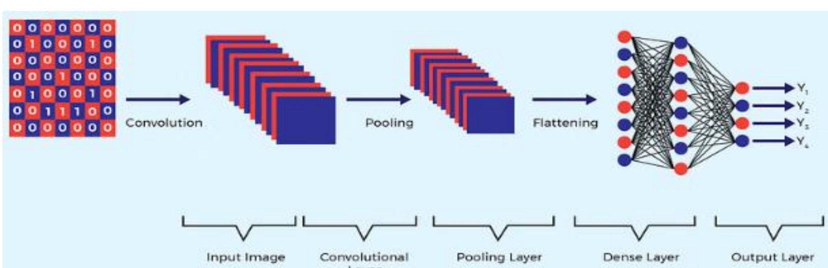

Figure 1.14: Structure of CNN model

3. **Deep Belief Networks:** DBNs are multi-layer generative graphical models that are built by sequentially stacking multiple unsupervised networks, such as RBMs or AEs. Each layer takes its input from the previous layer. This setup is

seen in figure 1.15 (Hinton, 2009). A DBN may so be partitioned into (i) AE-DBN, also called stacked AE, and (ii) "stacked RBM" refers to RBM-DBN, where the autoencoders that make up AE-DBN and the restricted Boltzmann machines that make up RBM-DBN have already been covered. The eventual objective is to create a contrastive divergence-based faster-unsupervised training method for each sub-network. A hierarchical representation of the provided data may be captured by DBN due to its deep structure. Using unlabeled data to train unsupervised feed-forward neural networks and then fine-tuning them using labelled input is the main notion underlying DBN.

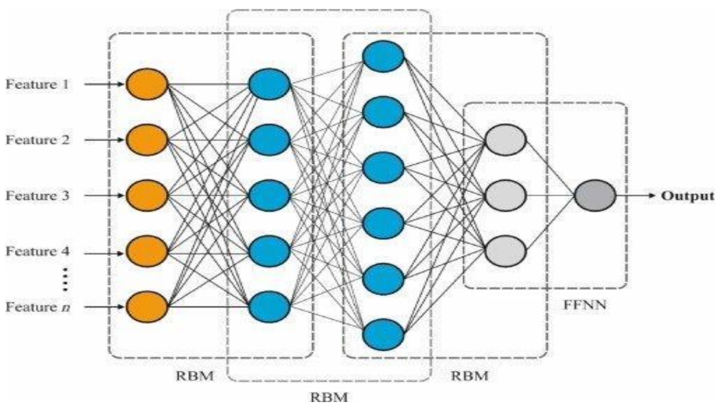

Figure 1.15: Architecture of Deep Belief Network

4. **Deep Neural Network:** DNNs are composed of ANN nodes organized in various hierarchical levels (figure 1.16). Output, hidden, and input layers make up this architecture. Much to sequential models, this one has a defined number of input and output layers and a fixed number of nodes in each of those two levels. Nodes in the input layers are proportional to the characteristics in the input, and in the output layer, the number of classes is directly proportional to the output layer's utility as a classifier. As said, artificial neurons are the fundamental building components. Let's review the composition and operation of a neuron in the human brain to understand its definition (Sewak et al., 2020).

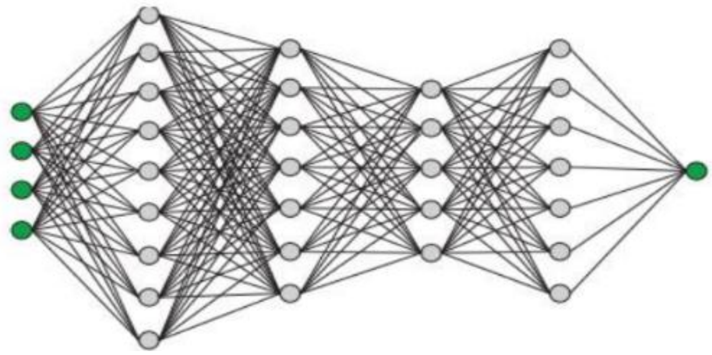

Figure 1.16: Deep neural network

5. **Recurrent Neural Network:** Another popular sort of neural network that employs sequential or time-series data is the RNN, whose input and output are shown in Figure 1.17. Comparable to feedforward and CNN, recurrent networks learn from training input. What sets them apart is their "memory," which lets them alter input and output depending on data from previous inputs (Mienye et al., 2024). In contrast to the typical DNN's assumption of total independence between inputs and outputs, RNNs' outputs are reliant on the components that came before them in the sequence. Typical recurrent networks struggle with learning long data sequences because of the decreasing gradients issue. Here we'll take a look at three popular RNN variants—LSTM, BiLSTM, and GRU—that mitigate the aforementioned issues and excel in a broad range of real-world application domains.

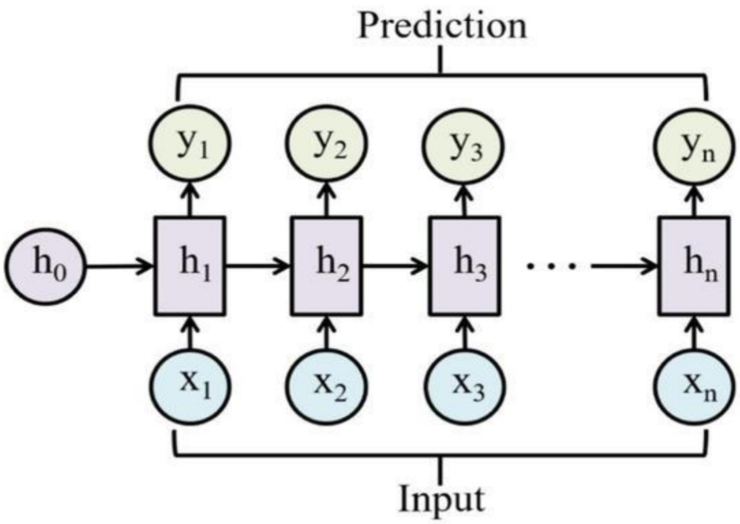

Figure 1.17: Basic RNN architecture.

1.7 Research Problem

In today's intricate and ever-changing supply chain environment, operational responsiveness and efficiency are hampered by ineffective monitoring methods, fragmented data interchange, and a lack of real-time visibility. Traditional packaging solutions fail to provide actionable insights into product movement, condition, and authenticity, resulting in increased costs, delays, and quality issues. The absence of connected packaging solutions, which integrate IoT, data analytics, and machine learning, prevents companies from optimizing their supply chain processes. Improving traceability, minimizing hazards, and allowing data-driven decision-making are all ways in which linked packaging technologies might improve supply chain efficiency, which is the goal of this study.

1.8 Purpose of Research

The study aims to investigate and apply AI methods to SCM for efficiency improvement, decision-making aids, and operational risk mitigation, using novel computational tools such as predictive and optimization models.

Objectives

- To prepare and analyze supply chain data to identify key variables influencing operations.
- To develop models to optimize demand forecasting, inventory management, and operational workflows.
- To propose explainable AI solutions for tasks requiring interpretability, such as risk management and supplier selection.
- To assess various machine learning models to determine their performance in solving critical SCM challenges.
- To deliver strategies based on AI-driven results to address specific supply chain inefficiencies and improve overall performance.

1.9 Significance of the Study

This study is important because it tackles the increasing need for more efficient supply chains by combining data, ML, and linked packaging solutions. By exploring how real-time data insights from smart packaging can optimize supply chain processes, the research provides valuable contributions to improving traceability, reducing operational risks, and enhancing product quality monitoring. The findings offer practical implications for businesses seeking to streamline logistics, reduce waste, and strengthen customer satisfaction through data-driven decision-making. For businesses looking to improve their agility and resilience, the research is especially pertinent since it shows how machine learning can optimize inventory management and forecast supply chain interruptions. In the end, this study is a great resource for experts in the business, lawmakers, and software developers who are trying to figure out how to make supply chains smarter and more efficient by using linked packaging.

1.10 Research Questions

The following Research Questions are as follows:

- What types of data are critical for implementing connected packaging solutions to enhance supply chain efficiency?
- How can machine learning optimize connected packaging for better demand forecasting and inventory management?
- What challenges arise in integrating data from connected packaging systems across interdependent supply chain stakeholders?
- How do interdependencies between supply chain partners influence the adoption of connected packaging solutions?
- What role does connect packaging play in improving real-time supply chain visibility and decision-making?

CHAPTER II:

REVIEW OF LITERATURE

2.1 Background

SCM operates in today's highly dynamic global economy in three keyways: data-driven operations and increased customer focus and enhanced complexity. Traditional supply networks encounter multiple issues because they lack clear visibility, goods tracking becomes inefficient and inter-connected communication systems delay deliveries. Connected packaging represents an innovative answer to supply chain problems through its integration of digital markers such as QR codes along with NFC and RFID tags together with IoT sensors in product packages (Keskin, 2022). Modern connected packaging technologies transform products into intelligent devices through continuous data flow across manufacturing and beyond until the point of user adoption for enhancing agile supply chain operations.

The implementation of connected packaging produces vast data amounts that ML and DL techniques can utilize for analysis. Deep learning technology together with machine learning examines data from sensors and enterprise systems that enables businesses to predict market changes while optimizing transport processes and preventing inventory shortages and maintaining proper storage conditions for vital items such as pharmaceuticals and food (Khedr & S, 2024).

Supply chains consist of connected elements that produce mutual performance effects between components. Any interruption in supply chain activities between procurement and transportation will eventually affect succeeding stages in the network. Connected packaging allows supply chain actors including manufacturers and logistics providers and retailers and consumers to be digitally connected through a synchronized data-rich environment. Better decision-making capabilities and sustainability gains become possible through improved traceability and decreased waste amounts (Kamble et al., 2020). Connected packaging offers stakeholders real-time alerts about expired goods and damaged items which leads to both supply loss reduction and better regulatory adherence (Yang et al., 2023).

Several obstacles exist for implementing connected packaging with ML/DL technology despite its benefits because it demands expertise and needs system integration while being expensive to implement and affecting data protection. Researchers have identified a clear void in understanding how intelligent data analytics with connected packaging functions should be deployed to deliver optimized performance across multiple industrial and geographic segments. Research studies currently fail to explore the complete effect on end-to-end supply chain efficiency because they examine individual uses and examine technology feasibility rather than holistic impacts (Ghaani et al., 2016).

The study aims to fill this gap by investigating the collaborative operation of supply chain structure with data and Machine Learning / Deep Learning techniques to deploy advanced connected packaging systems. This research aims to discover effective methods which produce superior supply chain results, together with greater responsiveness and sustainability performance.

2.2 Related Work

This section offers a synopsis of the research on linked packaging in relation to ML/DL and supply chain management. It shows how technology has improved the function of tracking, observing, and control. Data along with interdependencies become the central elements to enhance supply chain performance levels.

Supply Chain Management in connecting packaging using machine learning

In this research Anantharaman prakash, C.B. Senthil kumar (2025) discussed how with the help of AI SCM becomes smarter by automating various decision-supporting tools, enhancing the predictive capabilities, and by dynamically optimizing. The purpose of the study was to gather information from 150 supply chain specialists from various industries on the adoption, difficulties, and advantages of AI in the supply chain. These results indicate that AI-active business decisions were significantly more accurate with 35% as opposed to the previous 25%, reduced stock costs by 28%, and specifics concerning the suppliers increased by 42%.

In this research Shamsuddoha et al. (2025) investigates the ways in which RPA, AI, and ML impact supply chain operations in response to vulnerabilities and threats. The article centers on the ways in which AI and other pertinent technologies improve forecasting in order to accelerate logistics, boost warehouse efficiency, and encourage quick decision-making based on real demand. These include logistical applications of AI, demand planning, and management, as well as dealing with threats with the use of AI in the supply chain found in this study by way of concept analysis. This system addresses security weaknesses and operational excellence together with resilient solution requirements. The report also explores how AI could reduce risk management processes and predict disruptions as two ways that enhance supply chain resiliency. The results of this study on intelligent automation implementations in practice state the best practices and common issues. The findings offer a guide for pulling into focus organizations that wish to leverage AI for improved performance, increased flexibility, and optimal action grounded on data.

This study Mohammed et al. (2025) examines how AI and ML are transforming demand forecasting by discussing how the technologies may help to improve inventory control, decision makers and uncertainty. Lags and gaps are frequent outcomes of conventional approaches to production planning because they do not consider alterations in the environment. Using the primarily market data together with the second, historical and third, external variables help AI generate better projections in its predictive models. The advances in automotive reinforcement learning, DL and NLP enhance the proficiency in forecasting skills and enable the company to make anticipative changes to its supply chain plans. Furthermore, utilizing current demand to adjust production processes with the help of artificial intelligence, enhances supply chain flexibility, reduces production costs, and minimizes losses. Data quality is one of the obstacles that the article delves into.

This study Pasupuleti et al. (2024) integrates innovative and effective suggestions of modern machine learning for the enhanced organization of logistics and stock. Applying ML models to historical data from a global retail corporation (sales, inventory levels, order fulfilment rates, operational costs, etc.) and using a variety of ML algorithms (regression, classification, clustering, time series analysis, etc.) resulted in significant improvements in key operational areas, such as a 15% improvement in demand forecasting accuracy, a 10% reduction in overstock and stockouts, and a 95% prediction accuracy for order fulfilment timelines. The technique used for the client segmentation also included delivery preference, which enabled appropriate tailored service offerings, and it identified shipments that are in danger.

This study Samuels (2024) shows an overview of how AI is impacting SCM, ranging from Industry 4.0 to Industry 6.0 in relation to volatility, operation, people and how it addresses sustainable development, according to the PRISMA guidelines. This is due to the fact that AI greatly enhances SCM's capacity via better demand forecasting, inventory management, and decision-making. According to the context of Industry 5.0, the cooperation of humans and AI leads to improved relations that affect the sophistication in personalization and problem-solving domains.

This study, Sruthy (2024) makes speculative recommendations for bolstering the supply networks' resilience via the use of AI. The framework of the smart supply chain also includes factors such as risk management, operational effectiveness, and real-time monitoring as key elements of supply chain resilience that is underpinned by an application of AI technologies. It also stresses the need of having a good rapport with supply chain partners via the use of amicable communication and AI-based data-sharing technologies. It also offers an all-embracing blueprint regarding how to apply AI to SCM, which pays particular attention to how it can enhance sustainability, operations, and robustness; in this way, it assists organizations in managing the complexities of the current supply chains effectively.

This research Farooq & Yee Yen (2024) presents a thorough evaluation of scholarly works that investigate the substantial influence of AI on enhancing the robustness and longevity of supply networks. Analyze and evaluate major topics, difficulties, and developments related to AI applications in various supply chain settings using a data-driven method on the Web of Science platform. This review draws on 28 separate research that were published between 2020 and 2023 to form its conclusions. Topics covered in these studies include openness, last-mile delivery, multiagent systems, generative AI, and the significant impacts of AI on MSMEs.

This study Falkner et al. (2024) describes a case study of the deployment of ML in a business process that involves analyzing data and making predictions on demand for many goods and then deploying the results in an ordering system. A heuristic is used to regulate and manage stock levels, frameworks, and languages; the predictions are made using machine learning models based on real-world data from an Austrian shop. It can make the best use out of the current framework if enhanced communication exists between programming languages.

This research Submitted et al. (2024) delves into a potential of modern analytics to improve supply chain resilience and efficiency, including analyzing late deliveries and demand forecasts. This study seeks to improve supply chain performance by reviewing the literature and conducting research on potential applications of analytical approaches such as predictive modelling and machine learning. This research offers useful insights for companies aiming to construct supply chains that are both efficient and robust in the face of ever-changing business environments.

This research Pandey et al. (2024) discusses the ways that ML AI can help SCM. It gives the practical, social, and theoretical implications, explaining the theoretical approaches, method and significant findings in this area. Randomized articles sourced from academic journals, conference proceedings as well as industry reports were reviewed to obtain research papers and publications as part of the systematic literature review for the study. The use of ML and AI in demand forecasting, inventory management, logistics, and risk assessment is shown by these instances. DL, NNs, GA, and SVM are just a few of the AI and ML technologies that have been identified and used in this study to address supply chain issues.

This study Jackson et al. (2024) enables the scholars and practitioners to gain fundamental understanding of how and why the AI and GAI work in the context of Supply Chain Operations Management. The outlined tactic is a more systematic approach to identifying possible applications of AI and GAI in SCOM in enhancing decision making, optimizing processes, managing investments, and skill enhancement. It could help managers determine whether some of the operations they are involved in could be tweaked with the use of AI and GAI in resilience, throughput, accuracy, and effectiveness. The study emphasizes that AI and GAI are very adaptable and might potentially transform SCOM practices, inventions, and research in the future.

This research Verma (2024) draws on research conducted by McKinsey in 2023 to examine the use of AI in SCM with respect to problems like as demand forecasting, real-time stock control, and dynamic supply chain optimization. Digital twin and ML have helped in experiencing vast improvements in logistics cost, inventory management, and services in delivery. In order to successfully address supply chain complexity, the report stresses the strategic relevance of coordinating AI deployments with company objectives.

This research C. Qu & Kim (2024) to survey the literature on AI integrated technologies for SSCM in order to guide the course of future studies. By integrating AI into supply chain operations, the findings demonstrate that sustainability is mainly addressed, with an emphasis on economic and environmental concerns. Concerning social concerns such as working conditions and fair dealing, however, there remains a technical void. Researchers and practitioners may use this dynamic AI in SSCM framework to better understand how to incorporate AI into SSCM technologies and how to optimize supply chain models moving forward.

The study Mugoni et al. (2024) selected from several databases are 140 scholarly articles published in journals between 2012 and 2022. ScienceDirect, ABI/INFORMs (ProQuest), SCOPUS, and DOAJ (Directory of Open Access Journals) were the primary databases used. The findings provide strong evidence that SSCMPs have a beneficial effect on environmental performance. Sustainable SCM and environmental sustainability have attracted a considerable amount of research, but operational performance and social sustainability have received comparatively little attention. The mediating determinants of social and operational performance should be investigated in future research. Unpacking the link among SSCMPs and environmental performance requires additional industry-oriented research to use mixed techniques and mediating and moderating factors.

This research Shebeshe & Sharma (2024) seeks to investigate how SSCM affects industrial organizations’ performance and competitive advantage in Ethiopia. In order to accomplish the goals, data from 221 different manufacturing sectors in Ethiopia were collected and analyzed. This study makes use of descriptive and causal research methodologies, which are part of a quantitative approach. A total of 221 manufacturing industry managers and supervisors participated in the survey by filling out an online questionnaire. Furthermore, Smart-PLS 4.0 included structural equation modelling for data processing. The results show that SSCM significantly improves company efficiency and competitive advantage. Organizations achieve better outcomes through competitive advantage according to statistical data. In addition, the connection between SSCM and OP is indirectly affected by competitive advantage. Strong implementation of SSCM can potentially lead to improvements in both OP and competitive advantage according to the findings.

The research Rojek et al. (2024) considers the benefits and possibilities provided by the IIoT (Industrial Internet of Things) and ML and suggests priority development paths while outlining important trends in this field of research and development. A new era of supply chain efficiency, connection, and speed has dawned with the advent of 6G technology. Sustainable and safe operations are supported by this technology, which also increases visibility, automation, and cooperation. Businesses who improved supply chain design and planning and operations created a robust and flexible supply chain ecosystem.

This research H. Wang et al. (2023) contributes to the understanding of a ML framework for automatically enhancing the security of supply chain. This framework can help to identify fraud, predetermine when equipment needs to be maintained, and expect backorders when materials are out of stock. Machine failure prediction attains 93.4% accuracy, material backorder prediction 89.3% accuracy, and fraud detection 88% accuracy employing sampling techniques, across datasets of varying sizes. All the models were further improved through hyperparameters tuning; for instance, the two supervised algorithms, XGBoost and LightGBM achieved 100% of precision degree.

This study Hendriksen (2023) suggests a new theoretical idea known as the AII, which is two-pronged and addresses the spread of AI implementation via supply chain links and the use of AI in decision-making. AI Integration (AII) considers human capabilities to make meaning as the context which shapes both the integration process and disruptive factors of AI within supply chains. Multiple AI implementations have generated numerous theoretical and practical disruptions for organizations according to the presented work. Furthermore, it also raises some implications on interdisciplinary efforts as well as sociotechnical perspectives that are associated with integration of AI into SCM theory and practice.

This research A. Ahmed et al. (2023) describes the functions, uses, advantages, disadvantages, and potential of AI in the supply chain industry as a whole. The study cites supply chain visibility, demand optimization and planning, Inventory Management, logistics, and risk management as some of the advantages of AI technology. Moreover, it elaborates on how the integration of AI impacts the decision-making and collaboration processes within the supply chain. Some of the difficulties and ethical issues with using AI in the supply chain are also discussed in this article. This study reviews the existing situation and possible future applications of AI in SCM in an effort to add to the knowledge of academics and professionals in the field.

This semi-systematic Rolf et al. (2023) investigates the present level of excellence in SCM reinforcement learning and suggests a category system. Based on supply chain drivers, algorithms, data sources, and industrial sectors, the framework categorizes academic studies. Some crucial insights were uncovered by the reviewed process. As a starting point, the conventional Q-learning algorithm remains the gold standard. Secondly, RLS is most often used in inventory management because of its central role in supply chain synchronization. Finally, most tackle artificial data-driven SCM issues that are similar to toy problems.

In this study Lei et al. (2023) The risk prevention model for corporate business finance is implemented using Python. The experimental finding shows an effectiveness of the CGOA-SVM-SMA technique suggested in. The results show that the model's prediction and decision-making abilities are superior to those of competing models; this could greatly assist supply chain businesses in mitigating financial risks. AI is pivotal in today's fast-paced society and drives the advancement of financial technology. The supply chain relies heavily on machine learning (ML) algorithms, a subfield of AI research, for both day-to-day operations and future development.

In this research Jianying et al. (2021) analysis revealed that the fresh grape supply chain had a low overall risk, but that there were discrepancies in the risk at each link. The link between the links in the chain posed the greatest threat (R0), while economic, social, and cooperation risks were high. The second quadrant contained the most dangerous events, which had a small probability but did high damage. The optimized model outperformed a single BPNN in evaluating hazards in the sustainable grape supply chain, according to the results of the comparison. The results from PSO-BP justified its use due to improved statistical performance combined with reduced evaluation errors. In addition, the results showed which risks were most detrimental to the overall sustainability of the grape supply chain. This research does more than just theoretically improve the method of risk assessment in supply chains; it also provides actionable suggestions for making the fresh grape supply chain more resilient, bringing operational stability, and strengthening risk prevention.

This research Tirkolaee et al. (2021) aims to recognize the uses of ML in SCM, which is among the most popular AI approaches, as an AI method. This study provides a theoretical framework and lists the practical applications of ML in areas such as supplier segmentation and selection, supply chain risk prediction, demand and sales estimation, manufacturing, Inventory Management, transportation, distribution, Sustainable Development (SD), and the Circular Economy (CE). It then discusses the study's implications for future research and offers managerial insights into the study's limitations and challenges.

In this study Radmanesh et al. (2023), to investigate the possible effects of the suggested blockchain-based platform on the sustainability of the supply chain. The significant problem of cloud-based production-distribution systems and their impact on sustainability in both centralized and decentralized countries. The solution's findings show that the decentralized state is much better than the centralized state, and that this improvement improves the supply chain's and the suggested model's sustainability by taking an axiomatic design algorithm into account. Under optimistic circumstances, the distributed model reduces solution time by roughly 87% and under pessimistic ones, it reduces system cost by up to 45%. Less energy usage is a direct outcome of these upgrades, which also increase economic and environmental sustainability.

This study Sharma & Dash (2022) focuses on filling the present knowledge vacuum in smart source chain management related to artificial intelligence. The study also looks for ways that SSCM makes use of AI and Microsoft 365 to make it more successful, as well as for ways that AI and Microsoft Dynamics 365 may help with supply chain management studies and practices. The research primarily focusses on the applications of AI in smart SCM, which includes smart manufacturing, smart retailing, Facebook's AI implementation, intelligent delivery management, and Facebook's AI. Also covered in the study are the following aspects of Microsoft 365 as they pertain to smart supply chain management: SCM, the advantages of using MS Energetic Forces 365 for SCM, the reasons why smart SM should use MS 365, and the features of MS 365 in smart SCM. Perceptions are offered by this study via systematic investigation and synthesis. Finally, the study suggests ways AI may enhance an intelligent supply chain.

In this study J. Li et al. (2022) studied the many ways in which supply chain businesses might use the Internet of things to manage their operations, including RFID, EDI, and ERP. They conclude that the manufacturing business is the best fit for this simulation by looking at how widely used the Internet of things is across various sectors. An employ of the IoT helps to expedite the development of business supply chains. Consequently, businesses' bottom lines are positively impacted by the optimized route supply chain's horizontal integration, which also lends theoretical heft to the IoT service supply chain.

In this research Malviya et al. (2022) the different ML algorithms are examined and contrasted in relation to determining which components are in short supply, taking into account past sales and predictions. This considers backorder forecasts using different ML algorithms and gives plainly understandable and doable backorder decision cases, Prior to the occurrence of any shortages or backorders in the supply chain, industries must determine which items are most likely to be short in order to fulfil a high-performance goal and improve the company's overall performance.

This research Mallesham (2022) investigates the ways in which SCM is being transformed by an employ of new technologies. Demand forecasting and inventory management were two areas that were examined more thoroughly. Finally, they provide some last thoughts after highlighting the most recent developments and difficulties associated with AI and ML in SCM. Managing the movement of products, services, data, and money across a network of interconnected businesses is known as SCM. ML and AI are becoming more popular tools for businesses to use in an increasingly complicated market. Data-driven decision-making is the goal of ML, a subfield of AI. Since 2015, the number of firms claiming to use ML has increased by over 300%, and the total value of the AI industry is estimated to be approximately $2 trillion. The supply chain sector uses ML technology to discover deliveries' best routes while identifying delays in advance and detecting product quality issues before distribution.

In this research Kosasih & Brintrup (2022), an automated technique to find probable connections to the buyer utilizing GNN is proposed for the supply chain visibility issue, which is seen as a link prediction problem in the area of ML. The technique delivers superior results on a real automotive network while providing an additional artificial intelligence method to boost supply chain transparency. A widely used ML tool known as Integrated Gradient is integrated to help explain GNN decisions by showing which input attributes affect GNN outputs. The benefits and drawbacks of employing GNN for link prediction are also discussed, along with potential avenues for future research that could enhance supply chain visibility by means of ML.

This study Song et al. (2022) in order to provide a synopsis by classifying these subjects into six categories: (1) GSCM on platforms, (2) data-driven sustainable supply chain development, (3) variables impacting SCM, (4) methods for dealing with disruptions in SCM, and (5) data-driven medical SCM. Concerning the pressing need to address environmental degradation, the most significant findings from this study are sustainable development and the elements that influence supply chain management in a platform economy. Important parts of GSCM that have been discussed by certain writers include distribution channel choice, water scarcity, remanufacturing, cooperation, and lean management. A number of studies have examined various aspects of sustainable development, including methods for predicting waste discharge, evaluating efficiency, sustainability indices, and collaborative logistics networks; others have focused on SCM-influencing factors, such as Green Management, Stakeholder Motivations, Cause Marketing, Online Consumer reviews, Public Concerns, and sustainability indices. Their findings may be used by policymakers and business decision-makers to back GSCM practices. A lot of other obstacles, nevertheless, must be surmounted before implementation can be considered a success.

The study Lis et al. (2020) the primary areas of study within Sustainable Supply Chain Management (SSCM) include environmental management and economics, supply chains and sustainability, decision-making within SSCM, the practice of supply-chain management, social responsibility (SR) concerns, and the environment and management. Recent scientific studies in the area have centered on the following topics: human factors, sustainable supplier selection, production, the circular economy, efficiency, sustainable practices, trade, expenses, ecological footprint, and the textile sector.

This research Park (2021) utilizes supply chain data to identify the organization to which fresh data from unknown sources pertains when ML algorithms detect such data. A variety of ML methods are used, including LR, KNN, DT, SVM, NB, RF, and MLP. The accuracy, confusion matrix, recall, precision, and F1score are some of a metrics utilized to measure the efficacy of machine learning systems for multi-class classification. Companies (tiers) were correctly predicted by LR and MLP in the experiments, but not by NB, DT, RF, SVM, or K-NN.

In this research Wan (2021) suggested combining Random Forest with the XGBoost algorithm to create a hybrid model. This hybrid model would then have an optimized output that would be the average of the two models' outputs. The high amounts of TP and TF, as well as the big confusion matrices for XGBoost and Random Forest, indicate that the hybrid model works well. In addition, I test the hybrid model using the Kaggle-provided Datacom smart supply chain datasets. Compared to the other ML methods, my technique outperforms them in the experiments. In comparison to the LR method (0.49), the SVM algorithm (0.49), and the GNB algorithm (27.9%), my model has a better F1 score.

This research Mageto (2021) research defines big data analytics as following: processing data, analytics, reporting, integration, security, and economics. Transparency, a culture of sustainability, business objectives, and risk management are the components of sustainable SCM. Evidence suggests that BDA improves industrial supply chains' SSCM. Information technology skills gap and cyberattacks are two of the issues preventing BDA deployment. Incorporating the seldom-used Toulmin's argumentation model into management studies, this study bridges the gap between big data analytics and SSCM in industrial supply chains, adding both theoretical and methodological depth to the existing literature on SCM.

In this research Mastos et al. (2021) to evaluate the sustainability-related efficacy of CSCM, a case study with a real-life supply chain is developed and used to execute a waste-to-energy solution according to industry 4.0 standards. According to the results, CSCM may be possible if supply networks were redesigned using Industry 4.0 technology in order to facilitate the circular economy. There are obvious gains when the suggested approach is connected to the six Resolve model aspects of a circular economy: regenerate, share, optimize, loop, virtualize, and exchange. Considerable intangible benefits include enhanced supply chain traceability due to the suggested solution's full visibility and automation, and significant measurable advantages include a 5% and 15% increase in personnel availability, as well as a 15% increase in fleet resource availability.

This research Sardar et al. (2021) suggestions an ML method for SSCM on-demand forecasting. To alleviate the problem of having too much or too little stock on hand, you may use LSTM to predict future demand. Together with the model, there is a way to quantify the impact on the environment. The manufacturer maintains custody of the inventory under a consignment policy, but the retailer receives a predetermined fee in addition to a commission for each product sold. In this setup, the manufacturer sends RFID readers to the store. By using a traditional optimisation method, two mathematical models are resolved. The ML-RFID model outperforms the conventional system based on the findings from both models.

The research Kollia et al. (2021) employs current machine learning techniques to regulate labelling on food retail packaging, energy management during food maintenance, and efficient food manufacturing. Fully CNN, LSTM, RNN, Attention Mechanisms, Auto-Encoders, Latent Variable extraction and clustering, and Domain Adaptation are all used appropriately for this purpose. Three experimental experiments show that these AI approaches can provide state-of-the-art performance all the way across the food supply chain. This field is a relatively new one in the food processing industry, and it pertains to the efficient and safe transportation of food from the farm to the fork.

This research Sankhye & Hu (2020) to develop quality compliance categorization techniques based on ML and to test these models using a case study of a manufacturing line for appliances with many models. Classification and other supervised ML techniques were able to successfully forecast product compliance quality from production-related manufacturing data, with an accuracy of 0.99 and Cohen'sKappa of

0.91 for unit batch compliance quality. The removal of doubt via accurate prediction may have a significant positive impact on any part of a supply chain.

In this research Chen et al. (2020), focused on how cutting-edge smart packaging technologies may cut down on food waste and improve product tracking. Improved product traceability at every stage of the supply chain is possible with the help of integrated food SCM. Evidence suggests that smart packaging may cut down on food waste, and this article covers some of the most popular printing methods used by smart packaging systems (sensor and indication). Nevertheless, they also cover the difficulties of production and implementation, downsides associated with costs, and subsequent stages of the food supply chain. In general, smart packaging technologies have the potential to greatly enhance the food supply's quality and safety.

In this (Seyedan & Mafakheri (2020), this study aims to classify the predictive BDA applications of SVM, Time Series forecasting, clustering, KNN, NNs, Regression Analysis, and SVR in supply chain demand forecasting; to identify research gaps; and to propose a framework for these applications. This review also emphasizes potential future research directions by drawing attention to the paucity of literature on CLSC uses of BDA for demand forecasting.

This study Dumitrascu et al. (2020) is to create a model for evaluating performance that connects particular issues with the most relevant KPIs for every subsystem of SCM. This research shows that data mining may be used to identify the association between various key performance indicators and certain issues that arise in every company's SCM. Finally, the study presents a neural network-based GUI that makes use of the multilayer perceptron AI algorithm to forecast the most reliable KPIs for each chosen issue.

This study Makkar & Solanki (2020) is an endeavor to explore different ways in which SCM might benefit from ML approaches. The article examines successful and real-life case studies of supply chain optimization via machine learning. Supply chains (SC) function as interconnected facilities which unite many organizational entities. To reduce the overall cost of the supply chain, these organizations should cooperate. This entails several organizations to coordinate, collaborate and exchange information with each other. Still, there is a discrepancy between the ideal and actual supply chain network environments.

This study Del Giudice et al. (2020) examines the role of big data in a supply network and the impact of circular economy strategies in a system like this on business. With the use of an online survey and multiple regression analysis, data was gathered from 378 Italian enterprises who have embraced the concepts of the Circular Economy. Based on the findings, improving business performance from a circular economy standpoint requires careful planning of the supply chain, management of relationships, and human resource management. The connection between HRM tactics and business success is under the direction of a big data supply chain. Through applying circular economy principles to big data analysis, it deepens our understanding of digital transformation and the sustainability of supply chains.

In this Nasurudeen Ahamed & Karthikeyan (2020) proposed to adopt Reinforcement Learning Integrated in Heuristic search method (RLIH), which is a meta-heuristic search algorithm that has incorporated reinforcement learning in its search and planning for the supply chain network with the blockchain-based autonomous vehicles. With a decentralized app at its core, Reinforcement Learning Integrated in Heuristic search method (RLIH) outperforms the state-of-the-art heuristic search technique in terms of both data traffic and service time. It is used in SCM to train self-driving vehicles, and it integrates AI and ML algorithms.

In this Abbas et al. (2020) established and implemented an optimized DSCMR that integrates blockchain and ML. The system is divided into two primary components: a system for managing the supply chain of drugs and another for providing customers with prescription recommendations. Building a medicine SCM system that continuously monitors and traces the drug distribution process using Hyperledger fabrics is the first module's focus on the smart pharmaceutical industry's capabilities in this area. In contrast, the ML module makes use of the N-gram and LightGBM models to suggest the best medications to pharmaceutical sector clients. The models have been trained using the open-source ML repository at the University of California, Irvine's drug reviews dataset, which is well recognized and accessible to the public.

This research Tundys (2020) is to demonstrate how it has assessed the idea and extent of the themes from both a theoretical and a practical standpoint. Content analysis approaches and chosen case studies were used to conduct the study. Here are some of the most important findings: a new way of thinking about and using SSCM; an interpretation of the current state, volume, and scope of SC research; research trends and directions, including where the gaps in our knowledge lie, which could inform future studies; and a bibliometric analysis.

This study Baryannis et al. (2019) recent research into AI has focused on ML techniques and their possible application to supply chain risk management. Although there have been several studies examining prediction performance, interpretability has received less attention. This is a major concern since it is essential for supply chain practitioners to comprehend the results and make decisions that minimize or eliminate risks. This research aims to explore the trade-off of prediction performance and interpretability by using a data-driven AI framework to the issue of delivery delay prediction in a real-world, multi-tiered industrial supply chain. Working together, AI and supply chain professionals are the backbone of the system.

This research Florescu et al. (2019) explores the impact of SCM strategies—Supplier Selection, Product Stewardship, and Logistics Management—on the four pillars of SCM in the oil and gas distribution sector: Planning, Execution, Coordination, and Collaboration. This research used multiple regression models to look at 79 oil and gas distribution businesses in Romania and the Republic of Moldova. Overall and for each of the SCM functions studied, SSCM techniques were shown to have a positive and statistically significant effect.

The purpose Kot (2018) article aims to summarize the current state of sustainable development research on the topic of SCM for SMEs, as well as its future directions and the outcomes of relevant experiments. According to the findings, notwithstanding the imbalance mentioned in the literature, every sustainability area was crucial to the SCM strategies of the SMEs under study. The research also identifies the key components in the specific sustainability domains of SCM and SMEs; in this regard, integrating the idea of sustainable SCM into SMEs' operational plans seems to be a crucial role. Additionally, this supply chain encompasses the commercial, environmental, and social facets of sustainable development.

This study Rezaee (2018) aims to address this gap by illuminating the connection between different aspects of sustainability performance, how they work together to generate shared value for all parties involved, and what this means for the sustainability of the supply chain at large. This study presents a framework including sustainability theories, aspects of sustainability performance, the notion of shared value, and best practices in order to investigate the relationship among SCM and corporate sustainability.

using the suggested framework, businesses may include monetary and non-monetary sustainability measures into their supply chain sustainability initiatives from production design all the way through inbound and outbound logistics, purchasing, and production.

In this De Santis et al. (2017) investigates use of ML classifiers to suggest a prediction model for an unequal class issue, where the relative frequency of products going into backorder is low compared to those that don't. Materials backorder is a typical issue in supply chains that affects the efficiency and efficacy of inventory systems; this job makes use of specific measures such AUC-ROC and precision-recall curves, sampling approaches, and ensemble learning.

Supply Chain Management in connected packaging using deep learning

In this research Douaioui et al. (2024) offered a new method for predicting the likelihood of supply chain delivery delays. The clustering phase of the given framework is carried out using hyperparameter optimization and a new metaheuristic named RIME. The framework also incorporates multi-classification approaches. The multi-classification phase makes use of five separate DL models: GAN, CNN-LSTM, EL via bagging, EL via stacking, and EL via boosting. The GAN and CNN-LSTM forms the basis of the three ensemble learning models. This examination highlights the superior accuracy of Ensemble learning stacking by 0.926, highlighting its ability to make exact predictions.

In this study Sayyad et al. (2024), showcased a novel architecture for predicting sales orders that integrates feature extraction using DANN with several ML models. While keeping the data's unique behaviour, the DANN approach makes the data more generalizable. The strategy overcomes difficulties including limited data availability and considerable unpredictability in sales conduct. This pre-trained DANN model is able to extract useful characteristics from unknown goods by using the transfer learning strategy, which involves training the model on the training data. To the contrary, it is utilized as the basis for predictive models using ML methods. The ensemble models which contain DT and RF components also experience hyperparameter optimization.

This research Danach et al. (2024) completes a major need by providing a thorough framework for analyzing the pros, cons, possibilities, limitations, strengths, and risks of using AI in actual supply chains. Crime prediction, predictive modeling, and other fundamental artificial intelligence concepts like machine learning, optimization algorithms have been discussed in this paper. It provides an insight on how these techniques can be applied when addressing complex S.C. problems such as logistics, inventory management, and demand forecasting. In addition, examples of the cases that support all the suggested AI solutions are presented to demonstrate how the OP and accuracy improved greatly. By proposing a conceptual AI model that can keep up with the ever-growing demands of contemporary supply chain logistics, this study adds to the existing body of knowledge and deepens our theoretical and practical comprehension of AI's function in this domain. The findings that are presented below offer academics as well as leaders and managers, some accurate recommendations on how global supply chain networks could be strengthened by innovation.

The research Laldin Ismaeil & Lalla (2024) provides information on using AI in supply chain in different Arab organizations, which makes it rich with case studies to draw from when explaining how the various technologies are applied. Generally, the above examples are positive and negative impacts of AI implementation for supply chain perspectives, and it gives readers a balanced idea about the topic. Finally, there are several obstacles to AI deployment that must be carefully considered, despite the fact that AI has great potential for improving supply chain management. Technology advisory guidance for maximum AI technology utilization exists in the report addressed to Arab businesses. These rules stress the need of creating an adaptable company culture, doing regular risk assessments, and meticulous planning. Companies may improve their supply chain operations and achieve sustainable development by tackling five critical areas and using the full potential of AI.

This research Shankha Shubhra Goswami et al. (2024) provides an analysis of featured research on the potential of using AI in SCM to give more ideas on how such a great innovation has the ability to revolutionize SCM and hence create opportunities. Nevertheless, it also sheds light on the challenges that organizations are likely to face when implementing AI in SCM, including, data quality issues, challenges in privacy and security considerations and need for domain expertise. These are areas that were adequately addressed by the research as it developed a proper framework that would help in overcoming these challenges. Due to the timely approach, and efficient analysis of the governance and ethical issues associated with AI, and a clear depiction of the steps that need to be taken to incorporate AI in SCM from the initial stages, this research work added elegance to the prospects and challenges of SCM through an integration of AI in present day uncertainty.

This study Dalal et al. (2024) proposes the improvement of supply chain sustainability and efficiency while using CNNs with BiLSTM. The proposed method employs CNNs to determine the relations between geographical locations in the supply chain, pattern, and allocate resources efficiently. One of the main advantages of BiLSTM models is the ability to detect temporal dependencies that make demand forecasting and proactive decision making possible. The supply chain dynamics are explained by integrating these theories. The efficiency is increased by adaptive and monitoring features of both CNNs and BiLSTM, which make them capable to quickly change a setting. Fewer lead times decreased stock outs and improved inventory is among the benefits of predictive analytics. Supply chain sustainability, integrated transportation management, low carbon transportation plan, and green-sourcing decision-making intelligence are all part of sustainability. The suggested hybrid model has a specificity of 94.65%, an accuracy of 96.57%, a sensitivity of 95.67%, and an MCC of 0.85%. Evaluation of the outcomes shows that the suggested model considerably raised the degree of precision. Issues in the supply chain are shown from every angle by this study. Two ways that wisdom derived from CNNs and BiLSTMs may make the worldwide supply network greener are increasing operational efficiency and aligning supply chain processes with sustainability goals.

This Eni et al. (2024) research investigates how deep learning algorithms may be used to improve supply chain operations from a management standpoint. The research starts out by going over the complexity of contemporary supply chains, which are defined by globalized activities, linked networks, and changing client needs. The conditions with which these systems are characterized can be highly complex and unpredictable at times which makes it extremely difficult for traditional optimization issues to solve these problems by using standard optimization techniques. Deep learning functions as part of AI and has gained popularity because it extracts new hidden intelligence from extensive datasets. Some of the data types that deep learning computers can analyze using techniques such as convolutional and neural networks include sales forecasts, stock, logistic paths, and market trends. This is a management research paper that seeks to analyze the possibility of adopting Deep Learning algorithms to supply chain decision-making processes. Specifically, the most crucial ones are managing inventories, demand forecasts, routes, and risks. Possible additional factors include the use of DL to forecasting inventory management, logistical process improvement, and early disruption detection.

This research Lee & Rah (2024) discusses the impact of deep learning for the overall supply chain in the retail environment in such areas as demand prediction, inventory management, logistics, and customer satisfaction. This report from over 200 implementation cases in North America, Europe, and Asia provides an insight of operational efficiency of 20-45 AD: Revenue and forecasting accuracy, 25-35 Inventory cost and 15-30 Transport cost savings. This report also covers areas such as implementation issues in the future as well as how it was successfully done in other organizations. This paper aims at giving a clear theoretical background and implication to assist retailers in the active journey for the AI transformation in supply chain management.

This study Khedr & S (2024) finds the ways in which DL and ML have improved SCM in a number of areas, including as manufacturing, transportation, inventory control, demand and sales prediction, and supplier selection. To improve operational efficiency, resolve existing limits, and identify future research prospects, this study presents a thorough evaluation that examines the integration of DL and ML with SCM in detail. Consolidating previous research on SCM enhancement using ML and DL approaches, a thorough literature table provides a concise summary of goals, results, and development areas, as well as quick insights into the changing SCM environment.

This study Ashraf et al. (2024) enhances the supply chain overall by combining a cognitive digital twin structure with a hybrid DL method for issue detection. Combining a one-class SVM technique with a deep autoencoder neural network is what the suggested disruption detection module does. Furthermore, LSTM neural network models are constructed to ascertain the impacted echelon and the anticipated recovery period after the disruption. The suggested method would collect data on disruptions in real-time, which will aid decision-makers and supply chain practitioners in making smart choices to lessen the blow of disruptions. The results show the compromises between false alarms, the sensitivity of the disruption detection model, and the speed with which disruptions are detected. Much of the more recent writing on the subject has avoided this method.

This research Bassiouni et al. (2024) suggests many DL approaches to get the most of DL, especially for determining whether any product will be supplied late in a complicated SC system because of any unanticipated reason. To extract features, four distinct DL architectures are suggested: Simplest, Deep-LSTM, 1D-CNN, and TCN-1DSPCNN models. To determine if the information is delayed or not, six variant classifiers were employed: SoftMax, Random Trees (RT), RF, KNN, ANN, and SVM. These DL models improve the accuracy of detecting supply chain late orders by capturing complex temporal connections in a seamless manner. These suggested DL models are great for identifying late orders in the supply chain because they use hierarchical feature learning to spot subtle correlations and trends.

In this research Rajeev (2023) recommends looking at how packaging affects supply chain management. This proves that SCM places a premium on packaging. As a result of supply chain management's monetary considerations, issues like careless product packing have come to light. The research approach used is descriptive. Due to the lack of previous practical attempts at this kind of association, it was determined that such research would provide valuable, intriguing, and enlightening insights. Insights into the importance of practical approaches to packaging innovation at both the industry and global operations levels are provided by the findings, which may help improve profitability, guarantee efficiency in business methods, and lend credence to future studies on the topic.

This research Jahin et al. (2023) performs an extensive bibliometric analysis in conjunction with a SLR; after painstakingly reviewing 1,717 research, the authors drew important conclusions from 48 papers published between 2014 and 2023. After offers a more comprehensive picture of the changing SCRA environment, the study addresses important research problems and investigates current AI/ML methodology, results, strategies, and future trajectories. Outcomes show that hybrids, XGBoost, Random Forest, and other AI/ML models significantly improve SCRA precision.

This study Sakas et al. (2023) investigates the demand for DL's ability to classify and optimize operations, as well as its ability to provide more accurate information, revenue growth, inventory turn values, and to conduct correlation and linear regression analyses to glean useful insights for researchers. Next, these findings were used to create statistical coefficients, which were then used to build a hybrid model. The FNN method was used to estimate the behavioural analytical metrics of individual agents, and the result was a model that accurately simulated the chosen key performance transportation indices of supply chain firms. Key transportation performance indices, namely those pertaining to transportation expenses, improve by 60% when the amount of website visits to supply chain enterprises increases. On top of that, research shows that the chosen transport performance indicators (TPIs) fall by an average of 87.7 percent when the visibility of supply chain organizations’ websites increases.

In this Oyewola et al. (2022) suggested deep learning algorithms, including LSTM and 1D-CNN, were used to categorize datasets of health prescription supply chain prices. Then, in order to improve the classification model, the appropriate model hyper-parameters of LSTM and 1D-CNN were determined using Bayesian optimization using the tree parzen estimator and all kNN. To forecast the created models' accuracy, repeated five-fold cross-validation is used.

In this study Chong et al. (2022), propose and conduct a comprehensive Deep Reinforcement Learning research on optimizing the garment supply chain, with an emphasis on Soft Actor-Critic. This research compares six different models based on their performance in terms of inventory-to-sales ratio, sell-through rate, and service quality. When compared to other cutting-edge Actor Critic models, Soft Actor-Critic fared better at controlling inventory and meeting customer requests. In addition, the experimental performance of the models is evaluated by computing explicit indicators, In comparison to the S-policy, Twin Delayed Deep Deterministic Policy Gradient (TD3), and Trust Region Policy Optimization models (TRPO), Soft Actor Critic (SAC) achieved a lower inventory sales ratio of 7%, 41.6%, and 42.8%, respectively, demonstrating its superior ability to make sales and profit from inventory stocks.

This study Terrada et al. (2022) use Deep Learning techniques, such as ARIMA and LSTM, to enhance the SC's demand forecasting system's performance by analyzing a company's past transactions. Experiment findings allow for the selection of the most efficient approach that might outperform the tested methods in terms of accuracy. A substantial quantity of data is created throughout the SC). With the use of AI, this data might be analyzed to help each SC actor improve their performance and develop a better understanding of the consumer.

This propose Ali et al. (2021) an innovative method that shows how real-time data may be used to design the product and the supply chain at the same time. Using a three-step cloud-based management system is fundamental to the suggested methodology. The first stage involves creating a set of product families via the use of "AND" and "OR" nodes in a generic bill of materials. Step two involves creating a cloud-based framework to handle echelons-based real-time expenses. Step three involves solving the optimization issue using a metaheuristic approach based on Genetic Algorithm (GA). The model is shown using a numerical example of a power transformer. It optimizes the SCD based on real-time expenses.

Dörr et al. (2020) put out a strategy for fully automating the identification of packaging structures: In a logistics supply chain, a wide range of transported goods must be processed, identified, and verified at various nodes. The algorithm relies on deep learning models, specifically CNNs for image segmentation, computer vision techniques, and heuristic components. From a single image, one or more transport units can be located, and for each of these units, its characteristics, total number of packaging units, and arrangement can be recognized.

This research S. S. Ahmed et al. (2018) aims to find out what GSCM practices a textile sector in Dhaka's Gazipur District has implemented. The evaluation relies on data collected from 200 respondents. After data was analyzed using weighted arithmetic mean and chi-square testing, it was determined that the textile industry of Gazipur district is using several green SCM strategies. Organizations can manage possible delays and distribute resources efficiently and proactively thanks to their parallel processing capabilities, which enables real-time decision support.

2.3 Research Gaps

Based on the reviewed literature, several critical research gaps emerge in the application of AI and DL within SCM. While the current body of research demonstrates significant advancements in predictive analytics, optimization, and automation particularly through techniques such as LSTM, CNNs, DRL, and hybrid models there remains a noticeable lack of focus on model interpretability and explainability. Many studies prioritize performance metrics without considering the usability of AI models by supply chain practitioners, limiting their adoption in real-world scenarios. Furthermore, although behavioural aspects and decision-making dynamics are increasingly relevant, few studies incorporate human-centric or behavioural modelling in their frameworks. Another key gap lies in the limited cross-industry application and generalizability of proposed models. Most research is tailored to specific sectors like healthcare, apparel, or logistics, with minimal exploration of transferability across different contexts or regions. Additionally, the integration of AI in sustainable supply chain practices remains underdeveloped. While green supply chain initiatives are explored through traditional methods, there is a lack of AI-driven frameworks addressing environmental and social sustainability metrics. Lastly, many AI applications rely heavily on cloud infrastructure and rich datasets, which poses scalability and accessibility challenges, especially for SMEs in developing regions. These gaps highlight the need for future research that prioritizes interpretability, behavioural integration, sustainability, scalability, and domain adaptability of AI applications in SCM.

CHAPTER III:

RESEARCH METHODOLOGY

3.1 Proposed methodology

This study applies a comprehensive research methodology to address three key predictive tasks: demand forecasting, late delivery risk prediction, and delivery status prediction using the DataCo Smart Supply Chain dataset.

The forecasting process requires three initial steps followed by relevant variable selection (Date Orders, Order Item Quantity, Order Region) and missing data management and datetime transformation of order dates. All information about time ordering including year month day and weekday was extracted from the order date which led to removal of the original date column because it contained repetitive information. The analysis through Exploratory Data Analysis (EDA) produced graphical results which included histograms with count plots along with line charts along with boxplots for investigating order distribution methods and shipping duration patterns and regional activity. Daily order quantity outliers were found through application of the IQR method so researchers could enhance the data quality. Analysis of ordering patterns required aggregating the data into weekly and monthly quantity groups (YEAR_WEEK, YEAR_MONTH). To prepare the weekly data for Prophet time series forecasting we converted the timestamp columns to Prophet format and named fields as ds and y before performing normalization through Min-Max Scaler. The data received an 80:20 partition between training and testing sections. The model training process with historical data established a Prophet framework that generated predictions for future weeks along with trend-based visualization. Actual vs. predicted data plots were created together with the evaluation of training and testing performance through MAE, MSE and RMSE metrics. The flowchart of demand forecasting appears in Figure 3.1.

The late delivery risk modeling process started with data preprocessing through extensive exploratory data analysis which included handling missing values, irrelevant variable removal, datetime format application, temporal extraction for hour and day of week measurement and daypart categorization. The feature engineering procedure involved converting categorical variables to numerical data points as well as applying one-hot encoding techniques for appropriate columns. The data normalization involved applying Min-Max Scaler while splitting the data into training (70%) and testing (30%) portions. Random Forest analysis determined the ten most influential features which were displayed through a feature importance visual representation. Three classification models namely AdaBoost and Cat Boost and Multi-Layer Perceptron (MLP) were first implemented for late delivery risk prediction. Accuracy, precision, re-call, and f1score computations are the cornerstones of model performance assessment. A flowchart depicting late risk prediction appears in Figure 3.1.

For delivery status classification. The methodology started with data preprocessing steps which included check data structure while handling missing values and discarding irrelevant features and non-informative categorical variables to boost model performance and minimize noise. The target variable 'Delivery Status' received numerical encoding and the date features required datetime processing to obtain shipping duration calculations. The remaining categorical variables went through one-hot encoding while dropping the first category to prevent the issue of multicollinearity. A Random Forest classifier ran to determine the ten most influential attributes out of all predictors for selection in the model. The target variable imbalance was handled through SMOTE (Synthetic Minority Oversampling Technique) methodology. The preprocessed data underwent a split which divided it into training 80% and testing 20% portions. The three classification models are applied, namely AdaBoost, Cat Boost and Multi-Layer Perceptron (MLP) as shown in Figure 3.2 delivery status prediction flowchart.

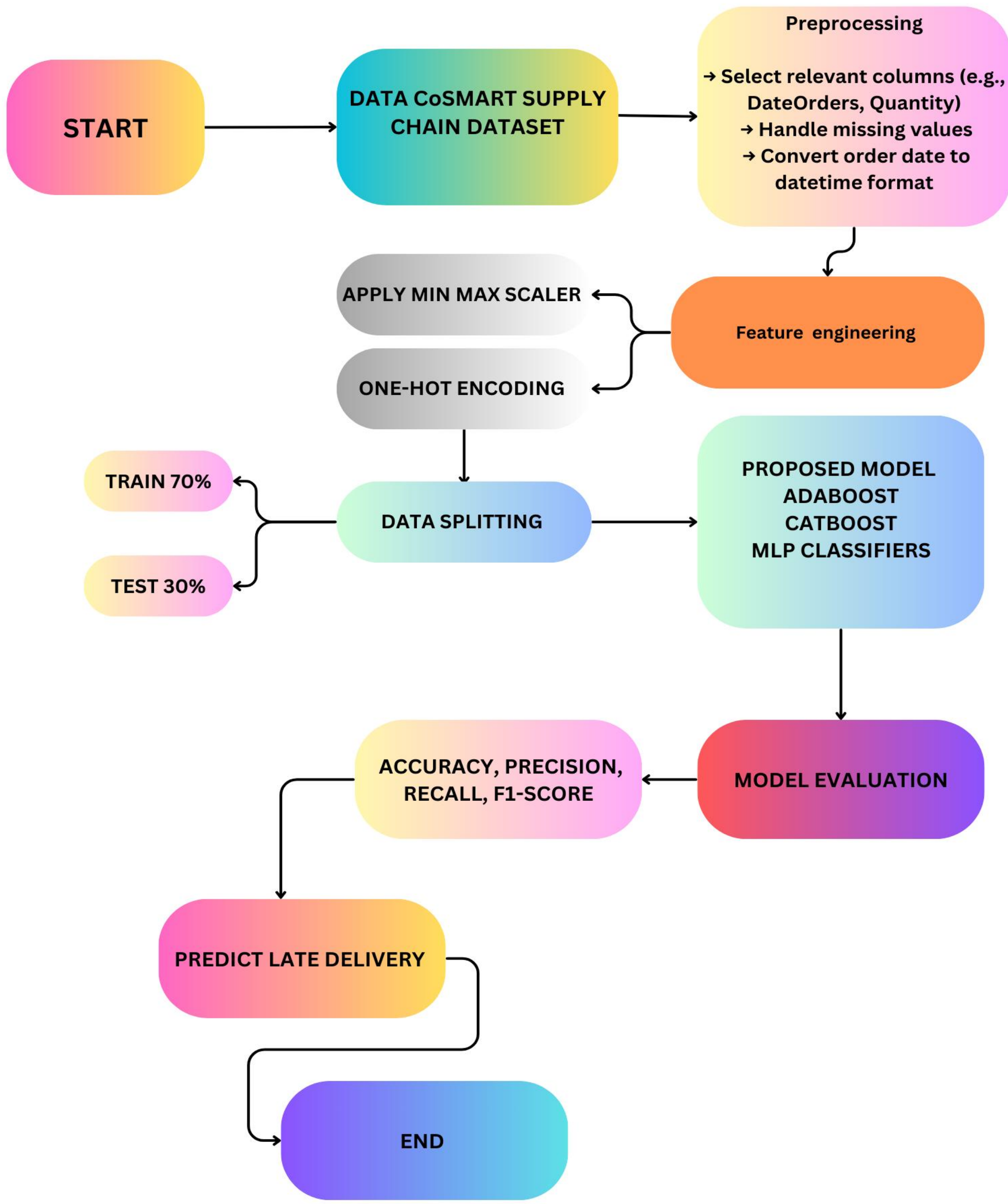

Figure 3.1: Late risk delivery prediction flowchart

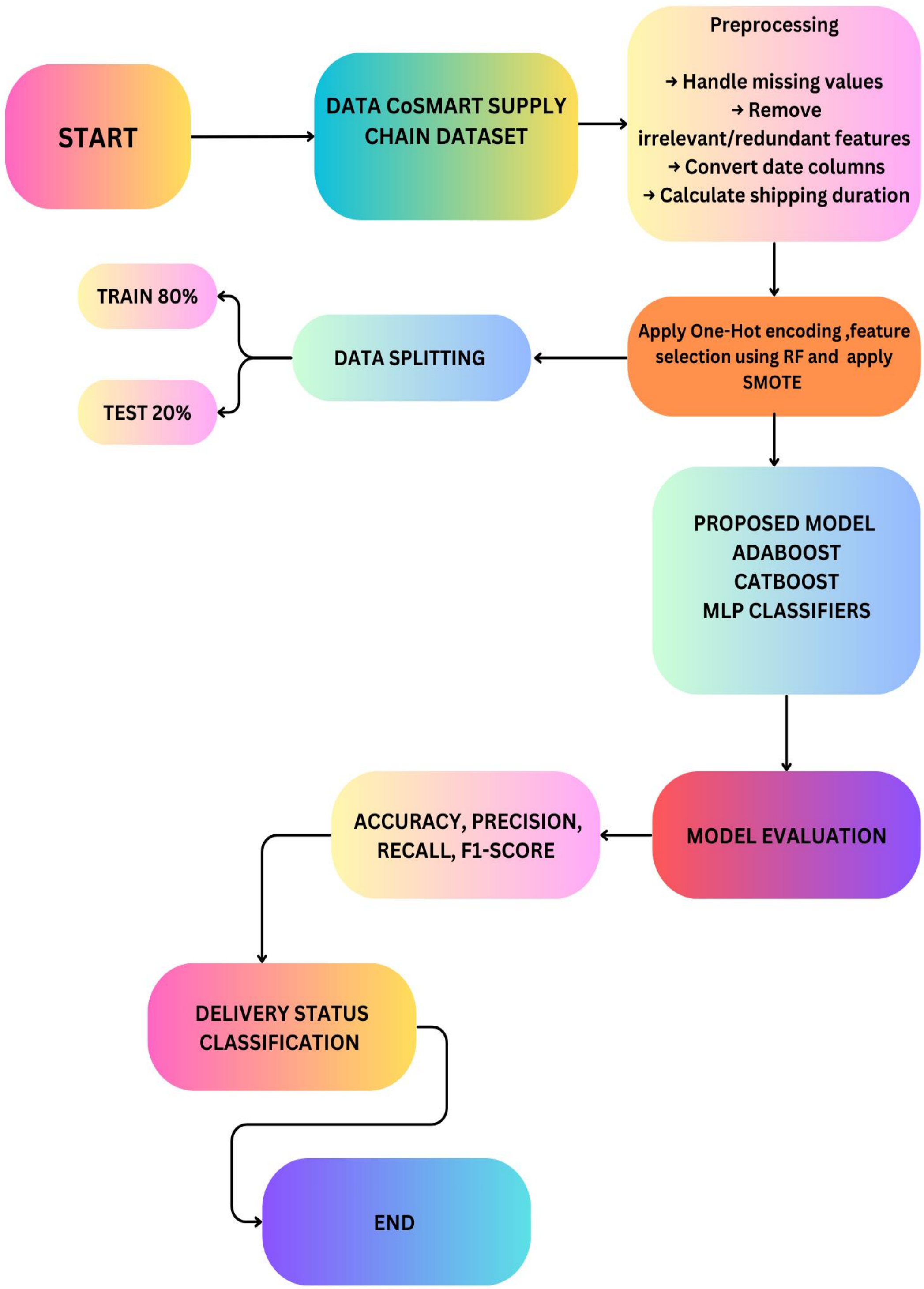

Figure 3.2: Delivery status classification flowchart

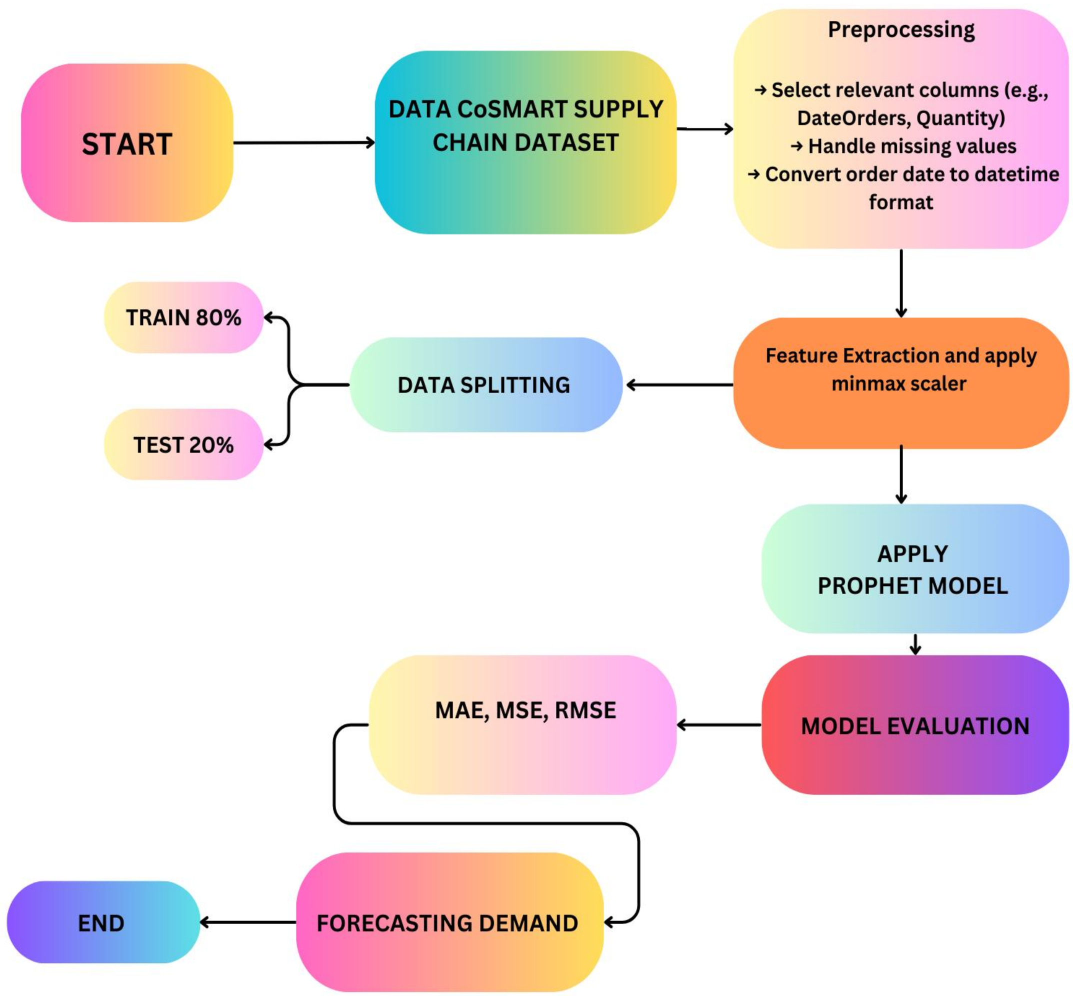

Figure 3.3: Demand forecasting Flowchart

3.2 Data collection

This study uses a primary dataset, DataCo Supply Chain Dataset[1].csv, contains structured data suitable for applying machine learning algorithms. In addition to this, the dataset is complemented by unstructured data in tokenized_access_logs.csv, which includes clickstream data, enabling the correlation of behavioral patterns with transactional records for deeper insights. The dataset includes a huge variety of products

[1] https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis

from many various industries, like clothing, sports, and electronics. To support understanding and feature interpretation, a separate file, DescriptionDataCoSupplyChain.csv, provides detailed metadata describing every variable present in the structured dataset, enhancing the usability of the data for advanced analytical tasks.

3.3 Preprocessing of Demand Forecasting Task

The following preprocessing are performed in demand forecasting:

- **Selection of Relevant Columns:** The dataset underwent a filtering process to maintain only the columns which supported demand forecasting. The selected column variables contained Date Orders, Category Name, Category Id, Order Item Quantity, Order Region, Order Status, ProductName, Product Card Id, Days for shipping (real), and Days for shipment (scheduled). By following this step the meaningful variables remain while the system removes noise which enhances forecasting model efficiency.
- **Checking for Missing Values:** The .is null (). sum () function executed a null check procedure. The used function enabled the identification of empty values across the chosen columns. The early detection of missing data enables the assessment of data quality for deciding between imputation or row removal procedures in subsequent processing steps.
- **Converting Order Date to Datetime Format:** The conversion of Date Orders to datetime format happened through application of pd.to_datetime () function. The conversion allows different time-based actions and helps users extract year, month or day components from dates.
- **Verifying Dataset Structure:** The .info () function was used to inspect the data structure and confirm that the Date Orders column was correctly converted to

datetime. It also provided an overview of each column's data type and non-null count, helping to validate the readiness of the dataset for time-series analysis.

- **Extracting Temporal Features:** The Date Orders column generated several new time counter features. The additional features comprised ORDER_YEAR, ORDER_MONTH, ORDER_DAY as well as ORDER_WEEKDAY which assigned Monday a value of 0 and ORDER_DATE. The features serve to capture time patterns and advance the forecasting model's capability to detect seasonality in addition to trends.
- **Dropping Original Order Date Column:** The Date Orders column received its extracted temporal features until being removed from the dataset. The researchers eliminated this column to minimize redundant variables before modeling, yet it maintained valuable data points for analysis in their separated parts.
- **Outlier Removal Using IQR Method:** A study eliminated anomalous data points in daily order quantities through application of the Interquartile Range (IQR) procedure. The IQR technique removed values which were substantially different from normal patterns so model training stayed accurate, and predictions would improve.
- **Creating Monthly and Weekly Period Features:** A new ordering timeline is made from the ORDER_DATE feature set.:
  - **YEAR_MONTH:** Extracted the year and month (e.g., "2024-01") to facilitate monthly aggregation.
  - **YEAR_WEEK:** Extracted the year and week number (e.g., "2024-W01") to enable weekly-level analysis.
- **Aggregating Order Item Quantity by Month and week:** Order Item Quantity was summed across each period in the data following its division by the

YEAR_MONTH feature. Aggregating the data by month allowed teams to trace seasonal trends and determine busy periods and planning cycles based on historical order quantity measurements. Weekly aggregation performed in a similar manner through the YEAR_WEEK column. Summary of weekly order quantities delivered better insight into order patterns which assisted in making both short-term forecasts and operational choices.

- **Preparing Data for Time Series Forecasting:** The weekly aggregated data needed to be reformatted to fulfill the specifications needed for time series modeling. The script processed the YEAR_WEEK data by converting its date format to timestamp using.dt.to_timestamp () and it also reset the index while renaming the columns to ds (date) and y (target value) for Prophet and other models.
- **Scaling Target Variable Using Min-Max Scaler:** The weekly order quantity data in the y column received normalization through Min-Max Scaler from Scikit-learn. Min-Max Scaler standardized the values while keeping their original distribution within a 0 to 1 range. The magnitude normalization technique enhances both convergence and accuracy levels in models which react to variations in value scales.

3.4 Preprocessing in Late Risk Prediction Task

The following preprocessing are performed in the Late risk prediction:

- **Dataset Information Retrieval:** To begin the analysis, process the dataset underwent loading then I accessed its general information through functions including .info () and. head (). The review of the dataset through these functions delivered an understandable data summary that enabled the detection of any formatting or inconsistency problems.

- **Null Value Check:** A thorough examination of null values occurred with the. is null (). sum () function to identify all missing data points. The evaluation of null values through. is null (). sum () created understanding about data quality problems and identified which values required correction through imputation or removal.
- **Removal of Irrelevant Features:** A bulk of features in the dataset included numerical and categorical variables which failed to serve as useful indicators for predicating late delivery risks. Dropping 16 categorical and 14 numerical features such as ‘Customer Email’, ‘Order City’, ‘Product Name’ and ‘Customer Id’, ‘Order Id’, ‘Latitude’ allowed for dataset simplification leading to enhanced model performance by minimizing dataset noise.
- **Conversion of Date Columns to Datetime Format:** The conversion process for datetime variables utilized pd.to_datetime () function to interpret Date Orders columns as Shipping Date and Order Date. Proper datetime conversion became necessary for both time-based feature calculation and uniform temporal research.
- **Time Difference Calculation:** A new column was generated to measure the hours between shipment dates and order dates through the calculation of diff(hours). Users obtained shipping delay and time-to-ship measurements through this feature.
- **Extracting Day of the Week:** The analysts extracted ship_day_of_week and order_day_of_week features through transformations of shipping and order date information. The order_day_of_week and ship_day_of_week features assigned numerical values that corresponded to day numbers starting from Monday as 0 until reaching Sunday at value 6.

- **Mapping Day Numbers to Names:** The numeric day of the week was converted into readable words such as Monday, Tuesday, etc., for shipping days and order days. The database generated two additional columns named ship_day_of_week_name and order_day_of_week_name.
- **Extracting Order and Shipping Hours:** The hour part of both the order and shipping timestamps was extracted and stored in order_hour and ship_hour. These features revealed temporal order and shipping patterns, which may correlate with delays.
- **Categorizing Hours into Dayparts:** A custom function divided the time spans into six specific dayparts: 'EarlyMorning', 'Morning', 'Noon', 'Eve', 'Night', and 'LateNight'. The function was employed for organizing order_hour and ship_hour into order_daypart and ship_daypart.
- **Numeric Mapping of Dayparts:** The model compatibility required the daypart categories to receive numerical assignments. The data now includes two numeric columns order_daypart_n and ship_daypart_n which serve as inputs for machine learning applications.
- **Dropping Useless Columns:** The removal of Delivery Status along with Customer Segment and datetime strings columns occurred after the completion of feature engineering because these features proved redundant or unhelpful for modeling purposes.
- **One-Hot Encoding for Categorical Variables:** The Type and Shipping Mode categorical features received their transformation through the one-hot encoding method using the pd.get_dummies function before getting cast as integers for model development. Dummy variables needed conversion to integer type since it served as essential components for model training consistency.

- **Normalization Using Min-Max Scaler:** The dataset numerical attributes received normalization by implementing Scikit-learn's Min-Max Scaler. The normalization through Min-Max Scaler produced values ranging from 0 to 1 which improved both convergence rate and balanced the features' relative importance.

3.5 Delivery status prediction Preprocessing phases

The following preprocessing are performed in delivery status prediction:

- **Data Inspection and Null Value Check:** The preprocessing process started with functional checks of the dataset using .info () and describe () to analyze data types, value ranges and distribution characteristics. A null value check was performed to identify missing or incomplete data entries which could influence the modelling process.
- **Removal of Irrelevant and Redundant Features:** The dataset underwent a reduction process where unneeded or repetitive features got eliminated. Amongst the categorical data features Category Name, Customer City, Product Name and Type were erased because it failed to produce predictive value. Model performance received no meaningful benefit from keeping numerical identifiers which primarily functioned as unique identifiers like Category Id, Customer Id, Order Id and Latitude. The removal process decreased the dataset size while reducing meaningless data points.
- **Target Variable Encoding:** A dictionary transformed Delivery Status categorical values into numerical data which allowed the algorithm to process them. The values within the Delivery Status column received numerical encoding through the Dictionary: {'Shipping on time': 0, 'Late delivery': 1, 'Advance shipping': 2, 'Shipping canceled': 3}. The transformation enabled machine learning algorithms

to accept the data. The developed encoding scheme resulted in a new Delivery_Status_Encoded column, which replaced the original Delivery Status column due to the removal of duplication.

- **Date Conversion and Feature Engineering:** The *Date Orders* and *Shipping Date* columns were converted from string format to datetime objects. This allowed for accurate date manipulations and calculations. Using these columns, a new feature called *Shipping Duration* was created by calculating the difference in days between the shipping and order dates. This new feature provided valuable information about delivery timelines.
- **One-Hot Encoding of Categorical Features:** The remaining categorical variables were formatted for use in ML models by transforming them using one-hot encoding. For every feature category, this procedure generated a binary column. To avoid multicollinearity, the first category of each encoded feature was dropped while still preserving the original information.

Data balancing using SMOTE in Delivery status prediction

SMOTE (Chawla et al., 2002) is a resampling method utilized to balance datasets with a severely uneven ratio. The plan's objective is to generate more synthetic minority class samples in order to boost the total number of minority class samples. In order to avoid the overfitting issue, the synthetic manufacturing of new samples was distinct from the multiplication approach.

The fundamental premise of Synthetic Minority Oversampling Technique (SMOTE) is to create new minority class data samples by extrapolating from nearby minority class samples (D. C. Li et al., 2022). Consequently, SMOTE enhances the classifier's capacity to generalize by boosting the representation of the minority class in an imbalanced dataset. You may understand the official SMOTE process by looking at this: The first step is to set N to an integer value that represents the desired level of oversampling. This figure may be chosen as it balances the dataset with a 1:1 ratio between the various classes. Iteratively, three primary stages should then be followed. Selecting a subset of the minority group at random is the first step. The second step is to locate its K nearest neighbors, with five being the default. The third step is to create more samples and interpolate using N randomly selected neighbours out of K. Figure 3.4 illustrates a basic understanding of how SMOTE operates.

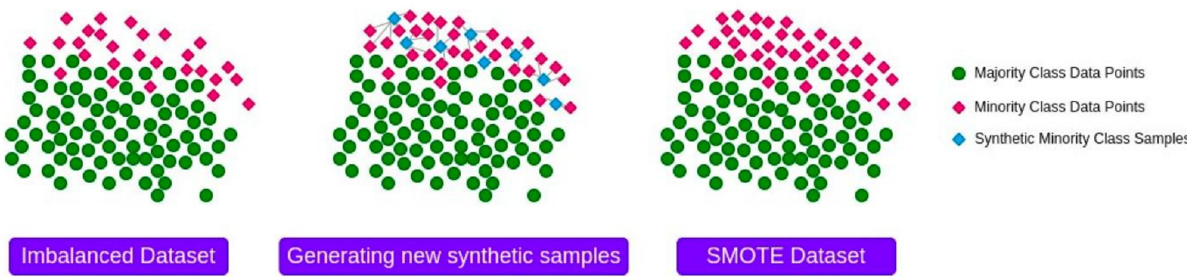

Figure 3.4: SMOTE process

3.6 Feature selection using Random Forest Model in late delivery risk and delivery status task

Feature selection using the RF model is a powerful technique for identifying the most influential variables in predicting outcomes in late delivery risk and delivery status. While training, the random forest method takes into account the ensemble's impurity reduction efforts and uses them to rank the features in order of relevance. This process highlights the features that have the greatest impact on model performance. By selecting the top 10 features with the highest importance scores, the model focuses on the most relevant predictors, which not only improves predictive accuracy but also reduces computational complexity and the risk of overfitting. Furthermore, this targeted approach enhances interpretability by allowing practitioners to concentrate on the variables that most significantly affect delivery-related outcomes.

3.7 Data Splitting

Splitting the datasets into training and testing sets allowed for effective evaluation of model performance in all three predicting tasks. An 80/20 split was applied to the time series data for demand forecasting, with 80% going into training and 20% into testing. In the late delivery risk prediction task, the data is split using a 70-30 ratio to train. For delivery status prediction, the dataset was divided using an 80-20 split, ensuring a balanced and representative training and testing distribution.

3.8 Proposed Prophet Model for Demand Forecasting Task

The proposed framework for product portfolio categorization and sales forecasting relies on a strategy or instrument that can reliably anticipate time series. Despite the availability of other tools and approaches, this research relies on Facebook's Prophet tool due to its capacity to provide somewhat accurate forecasts on a broad scale. The Prophet method is open-source software that was developed by Facebook's Core Data Science team. Time series data may be forecasted with the use of an additive model that accounts for nonlinear trends, yearly, weekly, and daily seasonality, and the effects of holidays. This approach works well with time series that include data from many seasons and are significantly affected by seasonality. Prophet is robust against missing data and shifting trends, and it often performs well when dealing with outliers. Many Facebook applications use Prophet to make reliable predictions since Taylor and Letham's research shows that it works better than all other methods in most situations (Žunić et al., 2020).

3.9 Proposed models for late risk and delivery status prediction

This section presents the proposed models for late risk and delivery status prediction using AdaBoost, Cat Boost, and Multi-Layer Perceptron (MLP). Each model is discussed based on its core architecture, learning approach, and advantages in handling classification tasks. These models are selected for their robustness, accuracy, and ability to handle complex, categorical, and imbalanced data in supply chain prediction scenarios.

Adaboost

An approach called adaptive boosting has been suggested to increase the ensemble's accuracy (Freund & Schapire, 1997). The main idea behind boosting is to train a chain of classifiers, with the hope that each one would give more weight to the tuples that were incorrectly categorized in the previous round. When used together, classifiers in an ensemble boost each other's accuracy, resulting in a very precise set of results. The term "boosting" refers to a generic approach of improving a random learning algorithm. There is no overfitting, and the model technique is simple to understand. It handles both kinds of binary classification issues in the same manner as ML handles problems with many classes. The expansion to the existing regression concerns is also provided by AdaBoost. When contrasted with bagging on noise-free data, the boosting technique seems to be rather effective. The datasets determine the algorithm's usage, which is to combine many classifiers into one advanced model. This leads to the term "successive classifier creation." (Riansyah et al., 2023).

The input used to construct a classifier is the preparation or training set. Repeatedly calling the collection of fundamental learning algorithms also ensures that the weights remain consistent across the preparation set. All weights are initially set equally, but each round expands the weights of examples that are incorrectly categorized in order to force the weak learner to focus on the challenging instances in the preparation data. Thirdly, two structures, boosting by sampling and boosting by weighing, may be used to link this boosting. With the boosting learning method, the basic learning algorithm may directly take a weighted set of training data. The basic learning algorithm now gets the whole preparation set with these techniques. Additionally, the boosting by sampling method draws models by replacing the training set with a probability distribution that is proportional to the weights of the variables. Using the cross-validation method, the iteration may be stopped (Walker & Jiang, 2019)

Cat boost

The term Cat Boost is derived from the words "Categorical" and "Boosting." Popular programming languages like R and Python make use of this open-source ML technique created by Yandex (Hancock & Khoshgoftaar, 2020). The Cat Boost framework is a GBDT that uses symmetric DT as its main learners. It processes class-type information rapidly and fairly, has fewer parameters, and achieves great accuracy. It also supports class variables. In addition, by fixing gradient bias and prediction shift, it reduces the chances of overfitting. The DT's node splitting criterion will be the label means, which are also known as greedy target variable statistics. The equation for this is provided as

$$\hat{x}_k^i = \frac{\sum_{j=1}^{p-1}\left[x_{j,k} = x_{i,k}\right] \cdot Y_i}{\sum_{j=1}^{n}\left[x_{j,k} = x_{i,k}\right]} \qquad \text{…... (3.1)}$$

Feature data frequently contains more information than label data, which is the most obvious problem with this approach. When the data structure and distribution of the training and test datasets are different, it may lead to conditional bias when the features are displayed by the average of the labels. A popular technique for improving Greedy TS is the addition of a priori distribution terms, which lessen the effect of noise and low-frequency category data on a data distribution. This method is expressed by this formula:

$$\widehat{x}_k^i = \frac{\sum_{j=1}^{p-1}\left[x_{\delta_{j,k}} = x_{\delta_{p,k}}\right] \cdot Y_{\delta_j} + a \cdot p}{\sum_{j=1}^{p-1}\left[x_{\delta_{j,k}} = x_{\delta_{p,k}}\right] + a} \qquad \text{… (3.2)}$$

where p is the previously added term and the weight coefficient that is larger than zero is usually denoted by aa. The prior likelihood of positive cases is the preceding phrase in a classification problem. Simultaneously, in order to make the model even more expressive, the algorithm will dynamically merge category features into new features.

The blueberry ecological suitability dataset is mostly organized using categories, taking into consideration the advantages of the Cat Boost approach discussed previously. More data may be trained using this method, which improves the model's expressiveness even more (Gayathri et al., 2022). Figure 3.5 displays the flow diagram of the Cat Boost algorithm.

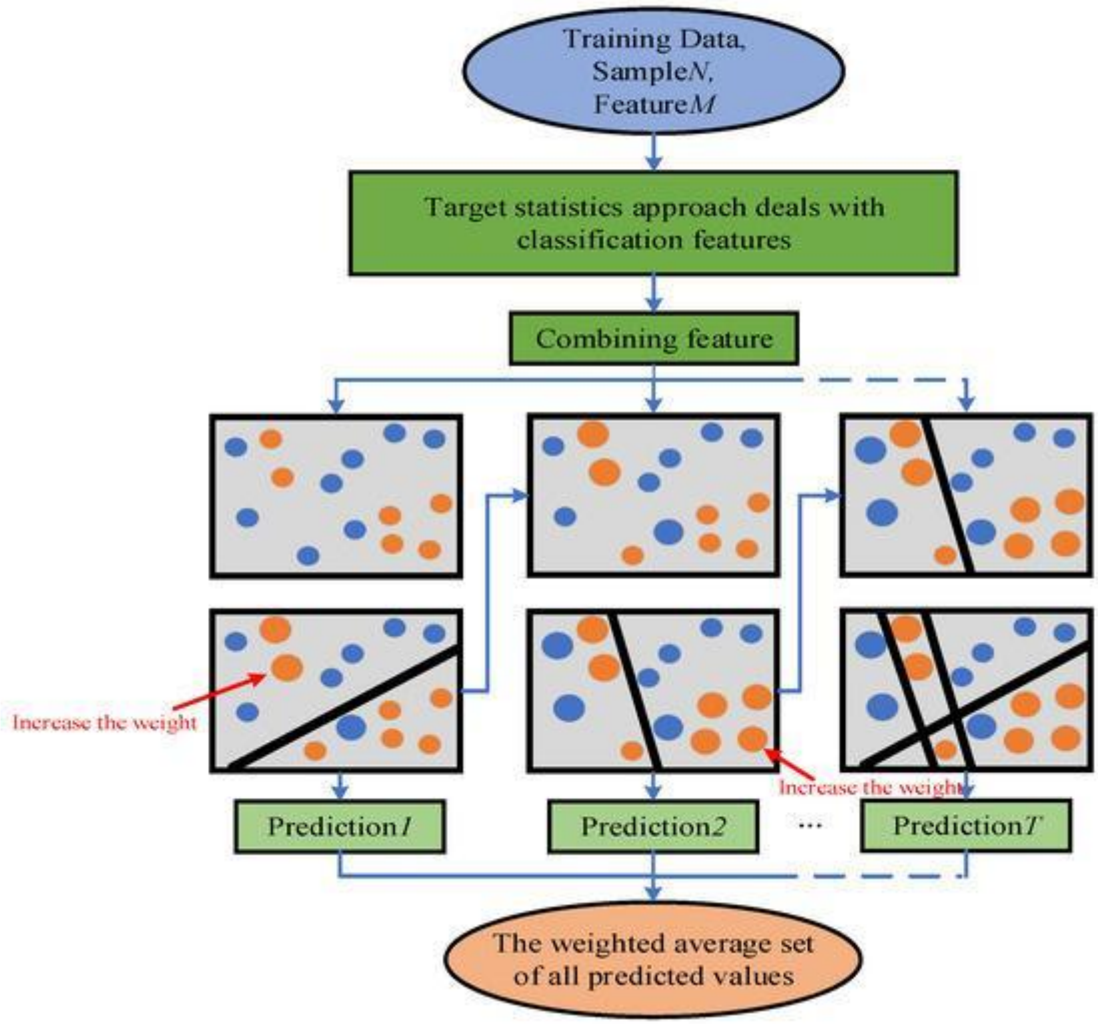

Figure 3.5: Flow chart of Cat Boost algorithm.

Multi-Layer perceptron MLP

Pattern recognition, classification, and prediction are just a few of the many uses for the MLP, an ANN. It can convert data sets from inputs into outputs; this property makes it a Feed-Forward ANN. Data points in an MLP network flow unidirectionally from the input layer to the hidden levels for processing, and then back to the output layer for return, as shown in Figure 3.6. Neurons are taught using the backpropagation learning method, and data flows forward from the input to the output layer. The output layer is responsible for carrying out the necessary tasks, including classification and prediction. A neuron's output is generated by feeding the sum of all the weights multiplied by the outputs from the preceding layer into an activation function. This process continues for every layer except the input layer, where the input layer is not involved.

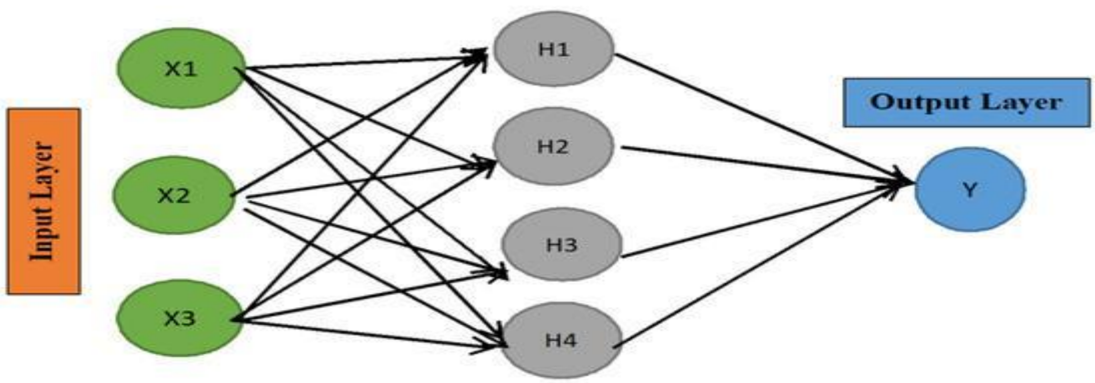

Figure 3.6: Simple Multi-Layer Perceptron ANN. The input layer's nodes are all linked to the concealed layer's nodes. After receiving data from the last hidden layer, the output layer applies the learnt model to provide predictions.

Neurons known as perceptions make up the MLP. Features with weights are inputted into the perceptron as X= (X1, X2, X3), where X= (X1, X2, X3). The following is the formula for computing the weighted sum, which requires numerical input features:

$$S(x) = \sum_{i=1}^{n} w_{ij} X_i. \quad \text{.... (3.3)}$$

where wijj is the weight that links neuron i to neuron j over two successive levels. The activation function fff is used to generate the perceptron's output after receiving the weighted summation as input. The sigmoid/logistic function is used as the activation function inside the buried layer:

$$f(x) = \frac{1}{1 + e^{-x}} \quad \ldots (3.4)$$

We utilize the SoftMax activation function in the output layer since we have binary categorization.

3.10 Proposed algorithm

Input Dataset: DataCo SMART SUPPLY CHAIN dataset

Output Prediction Tasks:

1. **Demand Forecasting** – Forecast future weekly order quantities
2. **Late Delivery Risk Prediction** – Classify orders as high or low risk of late delivery
3. **Delivery Status Prediction** – Classify order status (e.g., Delivered, Shipped, In Transit)

1. Prophet Algorithm – Demand Forecasting

1. Load the dataset and filter relevant forecasting features
2. Clean missing data and convert date to datetime format
3. Extract temporal features: year, month, day, weekday
4. Remove redundant original date column
5. Apply IQR method to detect and remove outliers in daily order quantity
6. Aggregate data by week (YEAR_WEEK) and month (YEAR_MONTH)
7. Rename columns to ds (timestamp) and y (order quantity) for Prophet
8. Normalize target variable using Min-Max Scaler
9. Split data into 80% training and 20% testing
10. Train Prophet model on weekly historical data
11. Forecast future weekly demand

12. Visualize actual vs. predicted demand trends
13. Evaluate model using MAE, MSE, RMSE

AdaBoost, Cat Boost, MLP – Late Delivery Risk Prediction

1. Clean missing and irrelevant values
2. Convert date fields to datetime format
3. Extract features: delivery hour, day of week, and daypart
4. Apply one-hot encoding to categorical variables
5. Normalize features using Min-Max Scaler
6. Split data into 70% training and 30% testing sets
7. Use Random Forest to identify top 10 important predictors
8. Train classification models: AdaBoost, Cat Boost, and MLP
9. Predict late delivery risk
10. Evaluate each model using Accuracy, Precision, Recall, and F1-score
11. Visualize feature importance and prediction outcomes

AdaBoost, Cat Boost, MLP – Delivery Status Prediction

1. Clean dataset: drop missing and irrelevant fields
2. Label encode Delivery Status
3. Convert date columns to datetime and compute shipping duration
4. Apply one-hot encoding to categorical features
5. Drop the first category to avoid multicollinearity
6. Use Random Forest to select top 10 features
7. Apply SMOTE to handle target imbalance
8. Split data into 80% training and 20% testing
9. Train AdaBoost, Cat Boost, and MLP classifiers
10. Predict delivery status

11. Evaluate performance using Accuracy, Precision, Recall, F1-score
12. Visualize model results and class predictions

CHAPTER IV:

RESULTS ANALYSIS

4.1 Experimental Configuration

This study's experimental setup makes use of a 64-bit Windows 10 PC outfitted with an Intel Core i5 processor, an AMD R7 2700x 8-core 3.7GHz CPU, 16 GB of RAM, and 500 GB of solid-state drive space. The research makes use of Jupyter Notebook, Python, and some of its prominent libraries, such as Pandas, NumPy, plotly, matplotlib, seaborn, and NLTK.

Programming Language Python

Python is a powerful and easy-to-learn programming language. The most advanced data structures are available, together with an approach to object-oriented programming that is both simple and effective (Van Rossum & Drake, 2006). Through its interpretive nature and dynamic typing together with beautiful syntax Python succeeds as a scripting language and fast application developer across most platforms. The wide user base adopts Python as a general-purpose programming language with its high-level comprehensive nature (Nosrati, 2011). Code readability stands as a fundamental design principle and the syntax allows developers to express concepts using fewer code lines beyond C language standards. Built-in features support the creation of programs at all scales (Rawat, 2020). The programming language Python supports multiple paradigms such as procedural as well as object-oriented and imperative and functional approaches (McCane, 2009). Python contains an extensive standard library and manages memory automatically and implements dynamic types (Summerfield, 2011). Memory management in Python involves dynamic typing as well as reference counting together with another tool known as a garbage collector that checks for cycles. Python features dynamic name resolution as its essential component since it performs binding to variable and method names while the program executes (Huang, 2023).

Python Libraries

1) **Pandas:** The Python module Pandas includes many data structures and tools for working with structured data collections. These sets exist in multiple scientific domains such as statistics economics along with the social sciences field. These types of data sets may be easily analyzed and manipulated with the help of the library's integrated, user-friendly procedures (McKinney, 2011). The system aims to function as the essential statistical processing foundation for Python. It functions as a powerful expansion of the existing scientific Python stack which duplicates and extends data manipulation functions found in R and different statistical programming languages.
2) **NumPy:** Python's numeric computing functionality is built into the language via the NumPy module. A set of extensions known as Numeric Python extensions (henceforth "NumPy") allows Python programmers to deal with large collections of objects more efficiently by using grid-like structuring (Ascher et al., 2001). An array is a collection of things with one or more dimensions, which may be

between zero and two. A one-dimensional array is analogous to a standard Python sequence, and a two-dimensional array is analogous to a matrix in linear algebra.

3) **Scikit-Learn:** The open-source Python ML software with the most features is scikit-learn. Due to its frequent incorporation into broader applications, including web services, machine learning is best provided using the same programming language as the rest of the application to ensure smooth integration. For ML-related applications, Scikit-learn is becoming more popular as a result of Python's broad applicability (Hao & Ho, 2019).

4) **Matplotlib:** Python, being free and open-source software, has generated over 100,000 toolkits. When it comes to visualizing data in Python, Matplotlib is a popular choice. A variety of 2D charts, as well as some simple 3D charts, may be quickly and readily drawn with Matplotlib (Cao et al., 2021). It has the capability to generate figures of publishing quality in many formats. As an export format, it works with several common ones, such as PDF, SVG, JPG, PNG, BMP, and GIF. A wide variety of plots are within its capabilities, including Line Plots, Scatter Plots, histograms, Bar Charts, Error Charts, Pie Charts, and Box Plots, among many more.

5) **Seaborn:** A free and open-source Python software for making statistical visualizations, Seaborn is available for download and use. An easy way to use matplotlib at a high level, it is compatible with pandas' data structures (Waskom, 2021). Fast data exploration and visualization prototypes with a lot of the stability and flexibility required to generate publishing-quality photos are made possible with the help of the seaborn library's matplotlib interface. It is applicable to a broad variety of domains and may be used to exhibit tabular representations of

various datasets. The ability to automatically "map" dataset variables to graph visual qualities is a crucial aspect of Seaborn. A "semantic" mapping is one that modifies the characteristics so that have significance in relation to the dataset.

Jupyter Notebook

The open-source Jupyter Notebook enables users to create computational workflows through an interactive computing environment which functions in browsers to help users generate and distribute their work documents (Pimentel et al., 2019). I Python served as the foundation for developing this powerful tool which now functions with Python, Julia, R and JavaScript as well as data science tools and scientific programs and learning applications. The system ties together running code execution and rich text plus mathematical equations and visualizations and images and videos as well as interactive widgets that enable researchers and developers to create single structured documents combining code with data and explanations. Jupyter Notebooks serve as virtual lab notebooks which support real-time data analysis and scholarly communication and experimental work for improved reproducibility and team collaboration (J. Wang et al., 2020). These notebooks serve both humans and digital programs so scientists can use them with external programs and digital repository systems. These platforms enable researchers to connect different research objects that include datasets and software and workflows to publications, so it becomes accessible while remaining transparent. The Jupyter Notebooks function as a FAIR (Findable, Accessible, Interoperable, Reusable) service that supports open science through mechanisms for researchers to share and maintain reusable findings (Randles et al., 2017). The interactive computing cells come with built-in features for version control systems along with digital identification and persistence functionality that adds value to academic and research applications. As a key component of modern computational research, Jupyter Notebooks streamline data-driven exploration and digital scholarly communication.

4.2 Performance Measure

Confusion Matrix

Confusion Matrix is an effective technique to measure a performance of a classifier. A confusion matrix serves as an effective tool to display classification issues by showing the different prediction and result combinations through a tabular format (Ramya & Stephan Thangaiah, 2022). The Confusion Matrix pieces are employed to identify four effective aspects, namely, accuracy, precision, recall, and F1score that compiles a table of all depended and real values of a classifier. To make it easier to predict the classes in the given classification it is important to find out the best separation boundary between the classes. These measures when compared highlight the best course of action leading to the best definition of the border. The following formulae compute accuracy, precision, recall, and F1score from this Confusion Matrix.

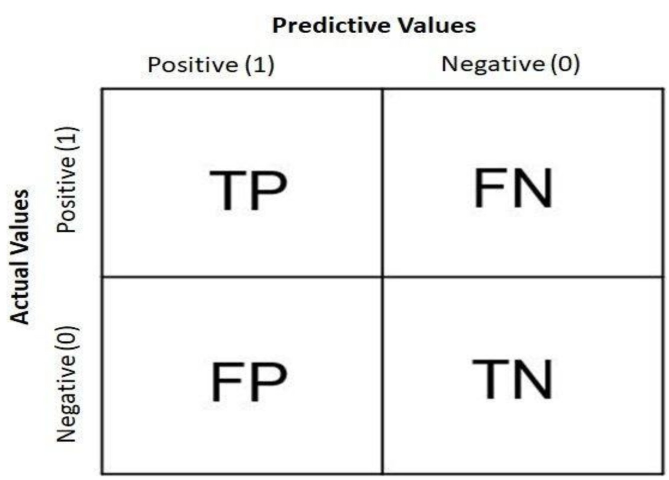

Figure 4.1: Confusion Matrix

- **Accuracy:** A percentage of samples that were accurately forecasted relative to a total number of samples is called accuracy. According to Equation (4.1), the model's accuracy increases as the value rises.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \dots\dots\dots\dots\dots (4.1)$$

- **Precision:** The term 'precision' is used to mean the number of samples correctly classified to the overall number of samples in the test set, using the following formula:

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots\dots\dots\dots\dots\dots\dots (4.2)$$

- **Recall or Sensitivity:** Sensitivity, also referred to as recall or detection rate, refers to the ability of the system to correctly identify conditionally late orders. Here is the equation that defines it:

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots\dots\dots\dots\dots\dots\dots (4.3)$$

- **F1-Score:** F-score is a measure of recall and precision that is given as a weighted harmonic mean, where we assign a weight of beta to precision. A higher F1-score indicates a better model, with a value of 1 indicating that beta is 1, and the formula is given by Equation (4.4):

$$F1 - Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \dots\dots\dots\dots\dots (4.4)$$

- **ROC Curve:** A method for representing, categorizing, and choosing classifiers according to their performance is a ROC graph (Fawcett, 2006). ROC graphs that describe the relationship of the hit rate and false alarm rate of the classifiers are not new to signal detection theory. It is a valuable feature of ROC graphs that enables a competent visualization and categorization of classifier accuracy separated from the class distributions and expenses of both types of errors. This is very important when there is learning that is sensitive to costs or when the distribution of samples is skewed.

4.3 Dataset Description

A dataset called "DataCo Smart Supply Chain for Big Data Analysis" (*Dataco Smart Supply Chain for Big Data Analysis Dataset*, 2023.) was obtained from Kaggle. It contains 180,519 rows and 53 columns that record several parts of supply chain activities, such as manufacturing, provisioning, sales, and commercial distribution. It includes both unstructured (tokenized_access_logs.csv) and structured (DataCoSupplyChainDataset.csv) data, allowing for the correlation of both types of data for knowledge creation. The dataset includes key attributes such as order details (OrderId, Order Date, Order Status, Order Profit, Order Item Quantity, Sales, Order Item Total, and Order Region), customer information (Customer Id, Name, Email, Segment, City, State, Country, and Zip code), product data (Product Id, Product Name, Product Price, Category, Description, and Status), and shipping details (Days for Shipping, Shipping Mode, Delivery Status, and Late Delivery Risk). The dataset covers three main product categories: clothing, sports, and electronic supplies and supports ML algorithms. This dataset is particularly valuable for studying the role of data, ML, and supply chain interdependencies in implementing connected packaging solutions for enhanced supply chain efficiency. Connected packaging methods may be improved with the use of real-time data and predictive analytics, which provide insights into shipment tracking, delivery performance, and client demand forecasts. Further aiding sophisticated analytics and decision-making in SCM is a supplemental file, DescriptionDataCoSupplyChain.csv, which includes thorough explanations of each variable.

Exploratory Data Analysis for Delivery Status Prediction Using Proposed Dataset

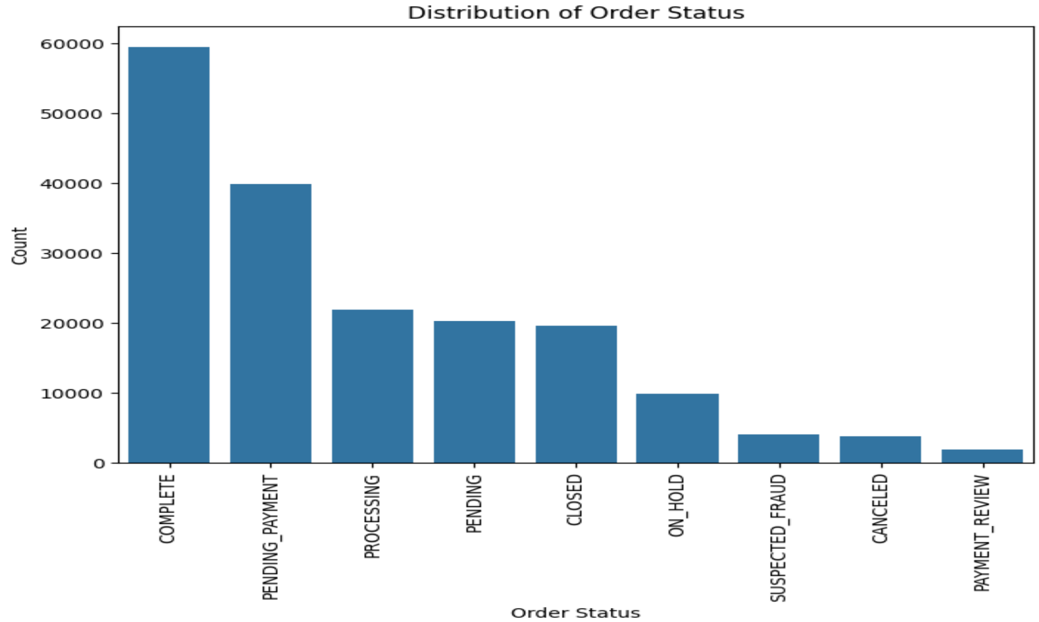

Figure 4.2: Distribution of the Order Status

Using the DataCo SMART SUPPLY Chain dataset, figure 4.2 displayed a count plot for a distribution of order status count. With a count surpassing 50,000, the count plot shows that "COMPLETE" is the most often occurring order status, followed by "PENDING_PAYMENT" with around 40,000 occurrences. While "closed" orders also reach 20,000, "processing" and "pending" statuses both report about 20,000 counts. Lower frequencies are recorded for "ON_HOLD" (around 10,000), "SUSPECTED_FRAUD" and "CANCELLED" (each about 4,000), and finally, "PAYMENT_REVIEW" has the lowest count, dipping below 2,000. This visualization offers understanding of the whole order fulfilment process and points out possible areas of concern, such the number of outstanding payments or orders labelled as questionable for fraud.

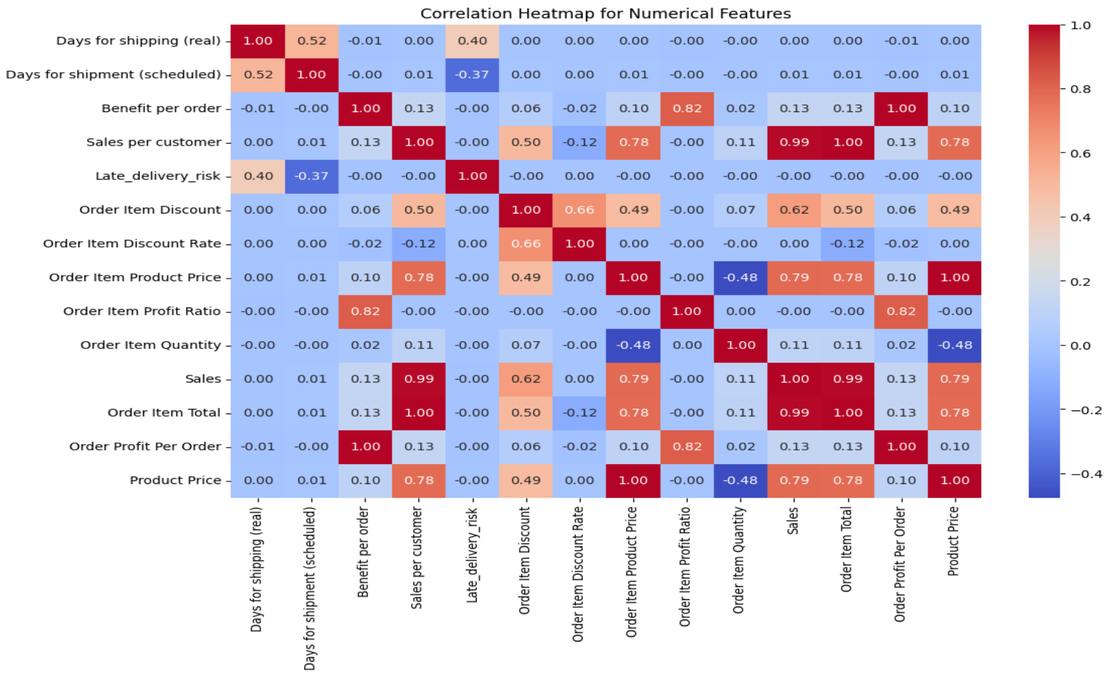

Figure 4.3: Correlation Heatmap for Numeric Features of DataCo SMART SUPPLY CHAIN Dataset

A correlation heatmap depicting the relationships among the numerical attributes of the DataCo SMART SUPPLY CHAIN dataset is shown in Figure 4.3. Heatmaps use a color-coded scale from -1 to 1 to show the direction and intensity of correlations. The "Days for shipment (scheduled)" and "Days for shipping (real)" variables show a significant positive correlation of 0.52, suggesting that the actual and scheduled delivery periods are almost identical. With a strong positive correlation of 0.99 among "Sales per customer" and "Sales," "Order Item Total" and "Sales" both illustrate the influence of total order value on overall sales numbers. A robust positive correlation of 0.79 between "Order Item Product Price" and "Sales" and 0.78 between "Order Item Total" provides more evidence of this. Conversely, "Late Delivery Risk" and "Days for shipping (real)" are significantly correlated with a negative value of -0.40, indicating that delayed deliveries are more probable in cases where shipping takes longer. It is useful for checking for potential multicollinearity issues in models and provides information on interactions of features that can be utilized for corporate decision making.

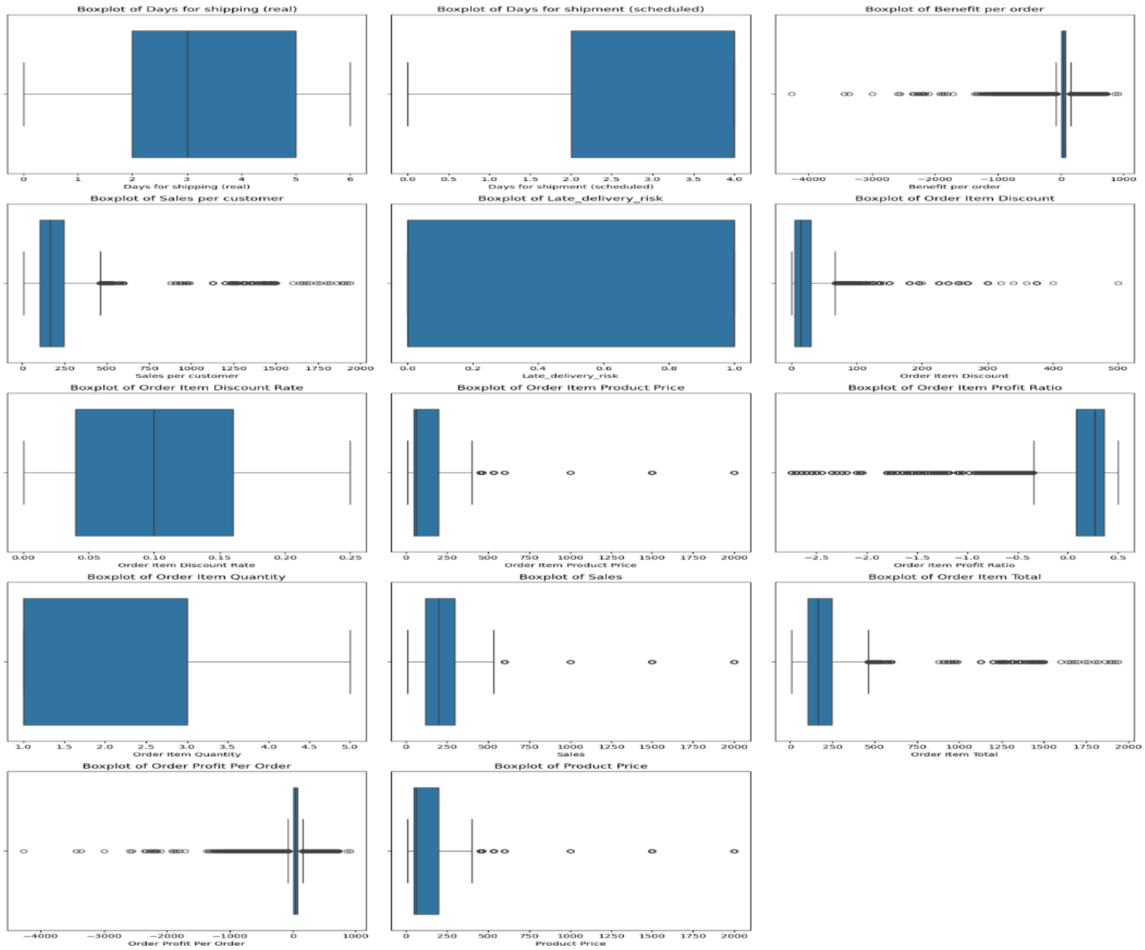

Figure 4.4: Boxplots for Numerical Features of DataCo SMART SUPPLY CHAIN Dataset

Figure 4.4 displays multiple boxplots which show the statistical distributions of different quantitative attributes found in the DataCo SMART SUPPLY CHAIN dataset. These boxplots reveal important information regarding the distribution patterns including spread and center values as well as outlier occurrences for multiple features. The delivery time variations appear in both "Days for Shipping (Real)" and "Days for Shipment (Scheduled)" plots, and the "Benefit per Order" and "Order Profit per Order" plots demonstrate profitability ranges showing distinct outliers. The data indicates two distributions "Sales per Customer" and "Sales" exhibit skewness because it contains numerous extreme values. The Late Delivery Risk variable shows binary or categorical nature by displaying data in a restricted value range. Multiple data points within "Order Item Discount Rate" and "Order Item Profit Ratio" extend beyond the median range to produce varied data distributions. The plots of "Order Item Product Price" and "Order Item Total" together with "Product Price" show the distribution of pricing data where high-value outliers exist. The distribution of "Order Item Quantity" data shows moderate variation although it has only a few outlier observations. The combined set of graphical visualizations lets users see distribution patterns of essential supply chain data and recognize both major trends and abnormal data points.

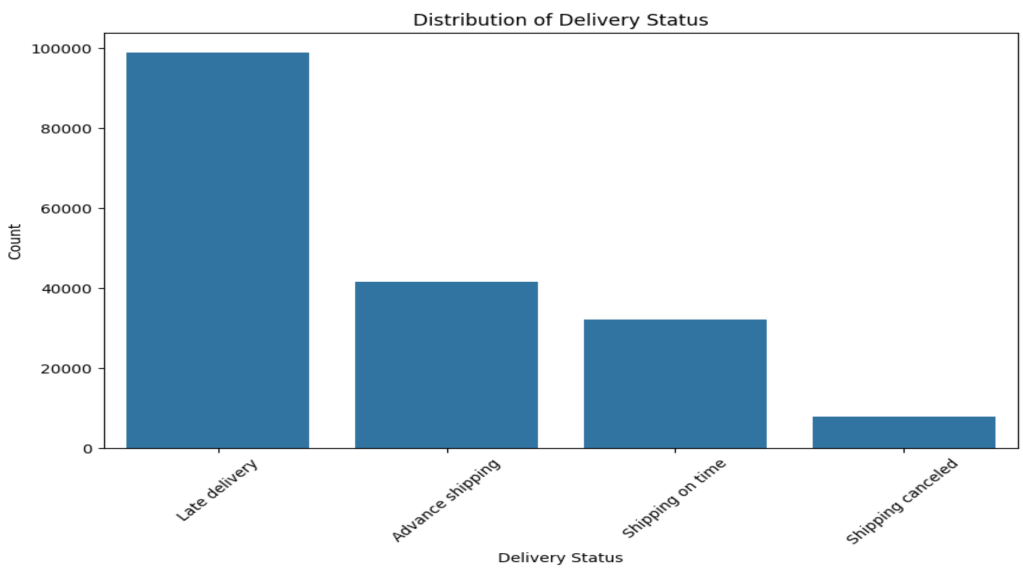

Figure 4.5: Distribution of Delivery Status

Figure 4.5 displayed a Count Plot of the distribution of delivery statuses, revealing that "Late delivery" is the most prevalent status, with a count approaching 100,000. With over 40,000 counts, "Advance shipping" is the second greatest category. For approximately 30,000 deliveries, "Shipping on time" is implicated. The status that occurs the least frequently is "Shipping canceled," which is less than 10,000 counts. This visualization indicates that the business may have a potential area of concern, as the high proportion of late deliveries indicates a need for further investigation into the causes of these delays.

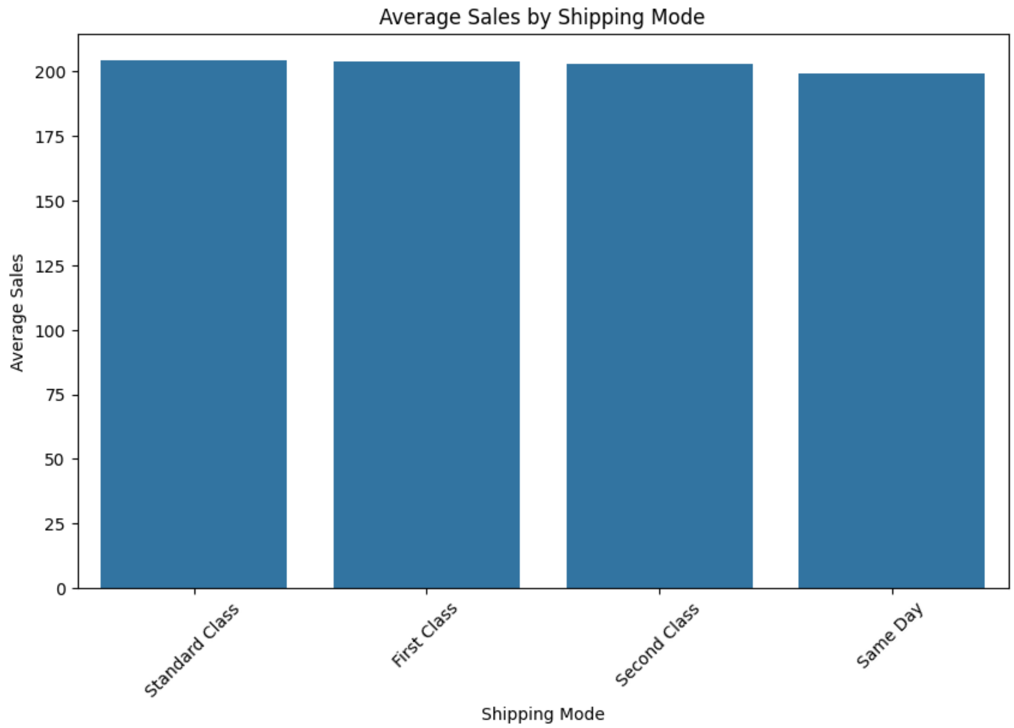

Figure 4.6: Bar Graph of Average Sales by Shipping Mode

Figure 4.6 showcases the average sales categorized by shipping mode within the DataCo SMART SUPPLY CHAIN dataset. The bar graph illustrates that "Standard Class" yields the highest average sales, closely followed by "First Class". "Second Class" demonstrates a slightly lower average sales figure, while "Same Day" shipping exhibits the lowest average sales among the four categories. The specific average sales values for each shipping mode are: Standard Class (203.45), First Class (200.75), Second Class (198.15), and Same Day (196.25). This visualization allows for a direct comparison of sales performance across different shipping options, potentially informing decisions on pricing strategies or service prioritization.

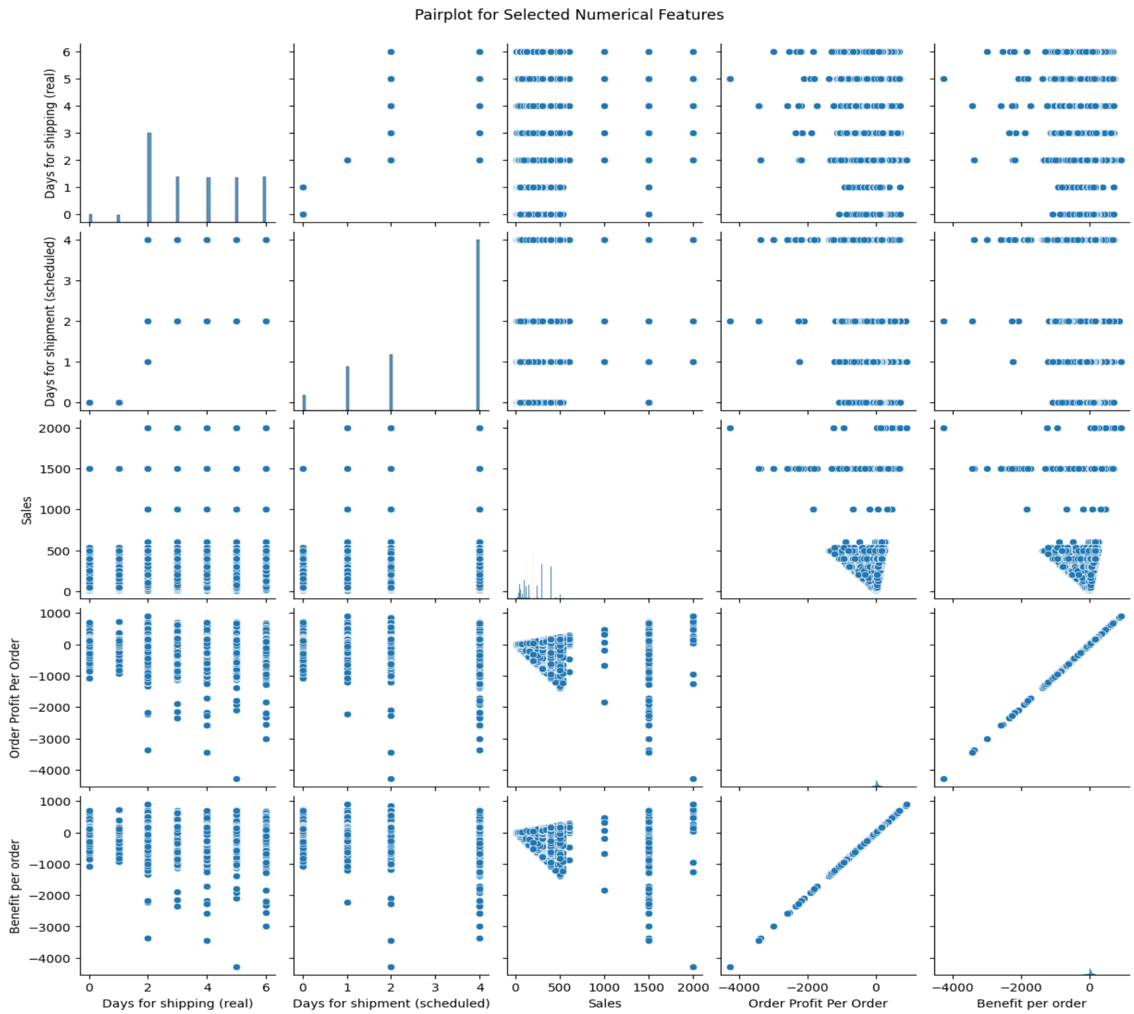

Figure 4.7: Pair Plot for Selected Numerical Features

A Pair Plot is illustrated in Figure 4.7, which illustrates the relationships between specific numerical features in the DataCo SMART SUPPLY CHAIN dataset. The pairwise correlations and distributions of "Days for shipping (real)," "Days for shipment (scheduled)," "Sales," "Order Profit Per Order," and "Benefit per Order" are investigated in this matrix of scatter graphs and histograms. The diagram features histograms representing each feature distribution which is shown in the diagonal section. The relationships between pairings of features are revealed by the scatter plots off the diagonal. It is important to note that the tightly concentrated nodes along a line imply a strong positive linear correlation among "Order Profit Per Order" and "Benefit per Order." Additionally, "Sales" exhibits a positive correlation with both profit metrics, albeit to a lesser extent. The "Days for shipping (real)" and "Days for shipment (scheduled)" exhibit a moderate positive correlation. This visualization facilitates comprehension of the connections between these critical numerical variables and may prove advantageous for predictive modeling feature selection or engineering.

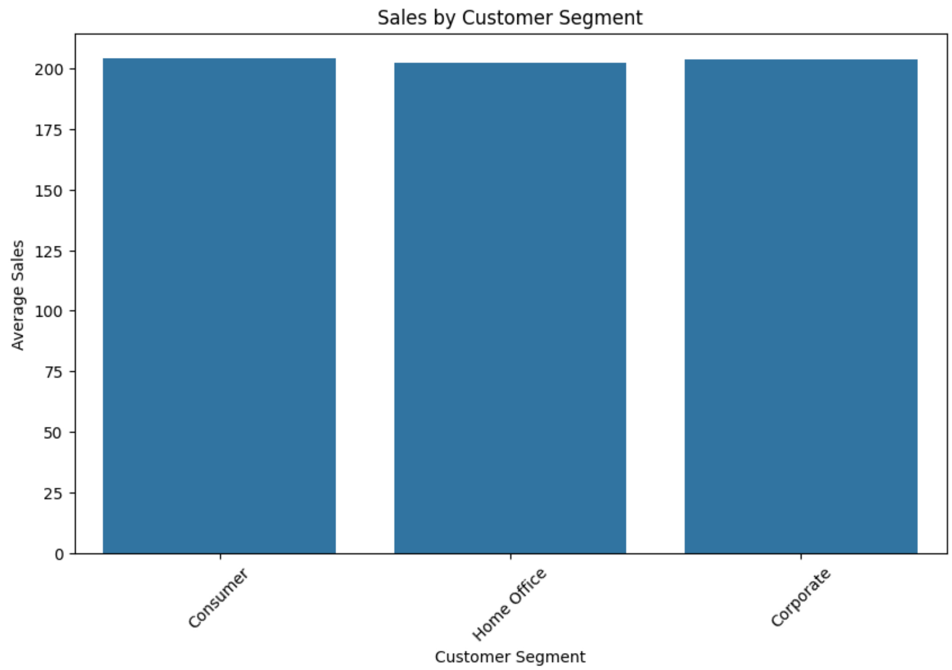

Figure 4.8: Bar graph of Sales by Customer Segment

Figure 4.8 displays the average sales across different customer segments. The "Consumer" category has the greatest average sales, followed closely by the "Corporate" segment, according to the bar graph. The "Home Office" segment demonstrates a slightly lower average sales figure compared to the other two. Specifically, the approximate average sales values for each segment are: Consumer (202.50), Corporate (201.25), and Home Office (198.75). This visualization allows for a comparison of sales performance across different customer groups, potentially informing targeted marketing strategies or sales efforts.

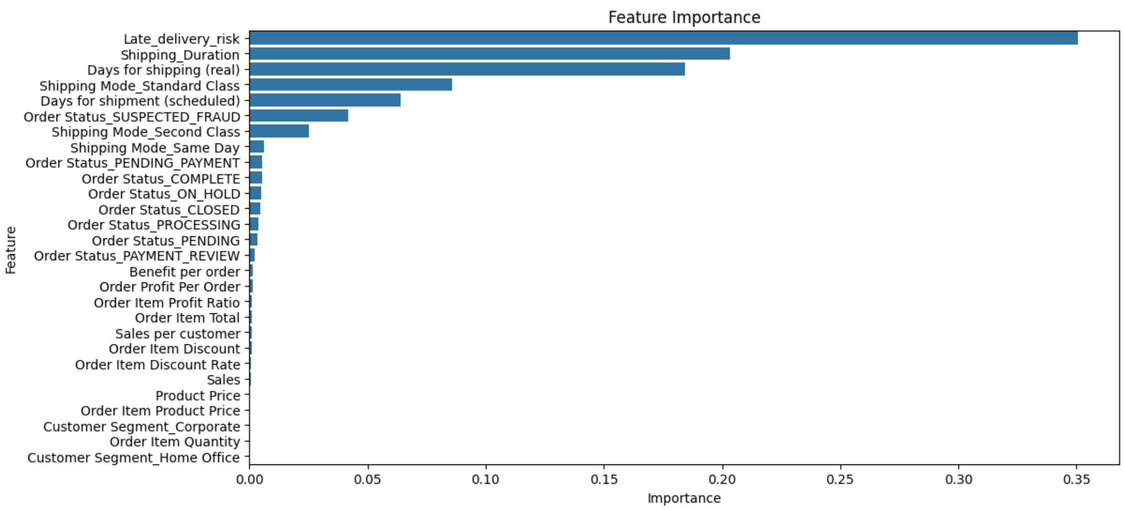

Figure 4.9: Bar Graph of Feature Importance

Figure 4.9 presents a feature importance bar graph derived from a random forest model, highlighting the top contributing factors in a dataset. The "Late_delivery_risk" variable proves to be the top contributing factor according to an information value score higher than 0.35. Theaters "Shipping Duration" along with "Days for shipping (real)" show substantial importance in the dataset according to their values exceeding 0.20 and 0.15 respectively. Two variables remain important to the overall prediction, yet it have an effect size of around 0.10: "Shipping Mode Standard Class" and "Days for shipment (scheduled)." The following set of features including "Order Status_SUSPECTED_FRAUD", "Shipping Mode Second Class", "Shipping Mode Same Day", "Order Status PENDING_PAYMENT", and "Order Status COMPLETE" present diminished yet perceptible importance levels. The visualization shows key predictive variables chosen by the random forest model to help explain which variables influence the target variable the most.

Exploratory Data Analysis for Late Risk Prediction Using Proposed Dataset

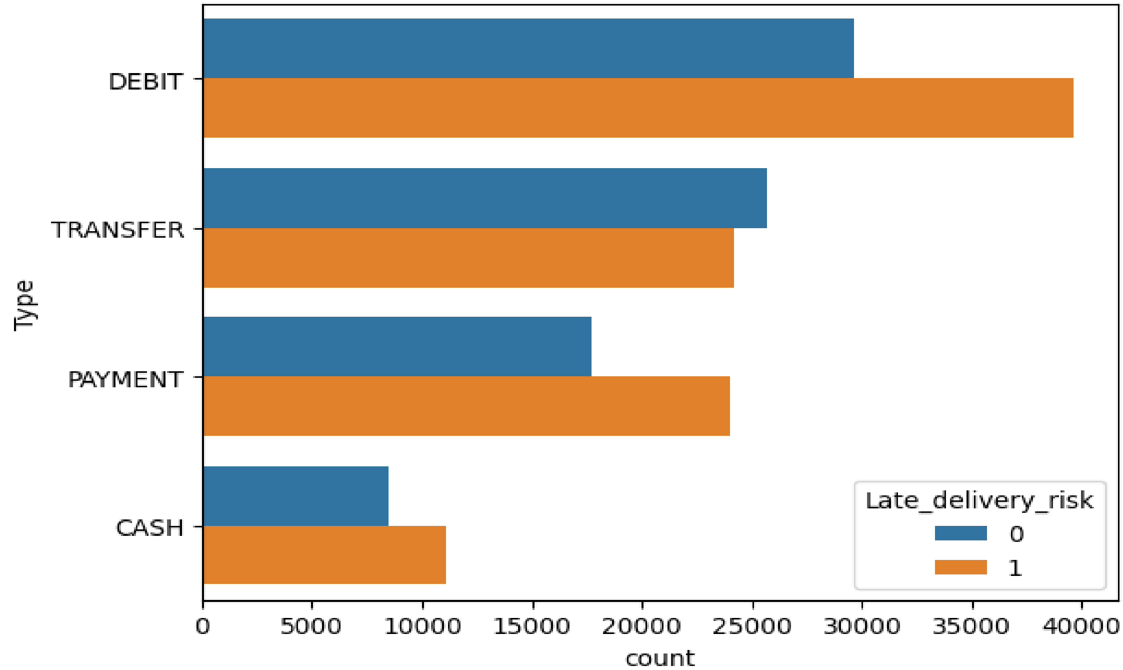

Figure 4.10: Distribution of Late Delivery Risk by Product Type

Figure 4.10 displays a distribution of Late Delivery Risk across various product types, categorized by payment method, with specific count values provided. The stacked horizontal bar chart displays a count of orders for every product type ("DEBIT," "TRANSFER," "PAYMENT," "CASH") segmented by whether it was delivered late (Late Delivery Risk = 1) or not (Late Delivery Risk = 0). For "DEBIT" transactions, there were approximately 26,000 on-time deliveries and 40,000 late deliveries. "TRANSFER" shows roughly 23,000 on-time and 25,000 late. "PAYMENT" has about 18,000 on-time and 23,000 late. "CASH" transactions had approximately 9,000 on-time and 11,000 late deliveries. This visualization emphasizes that "DEBIT" transactions exhibit both the highest overall volume and the largest difference between on-time and late deliveries, with a considerably greater number of late deliveries.

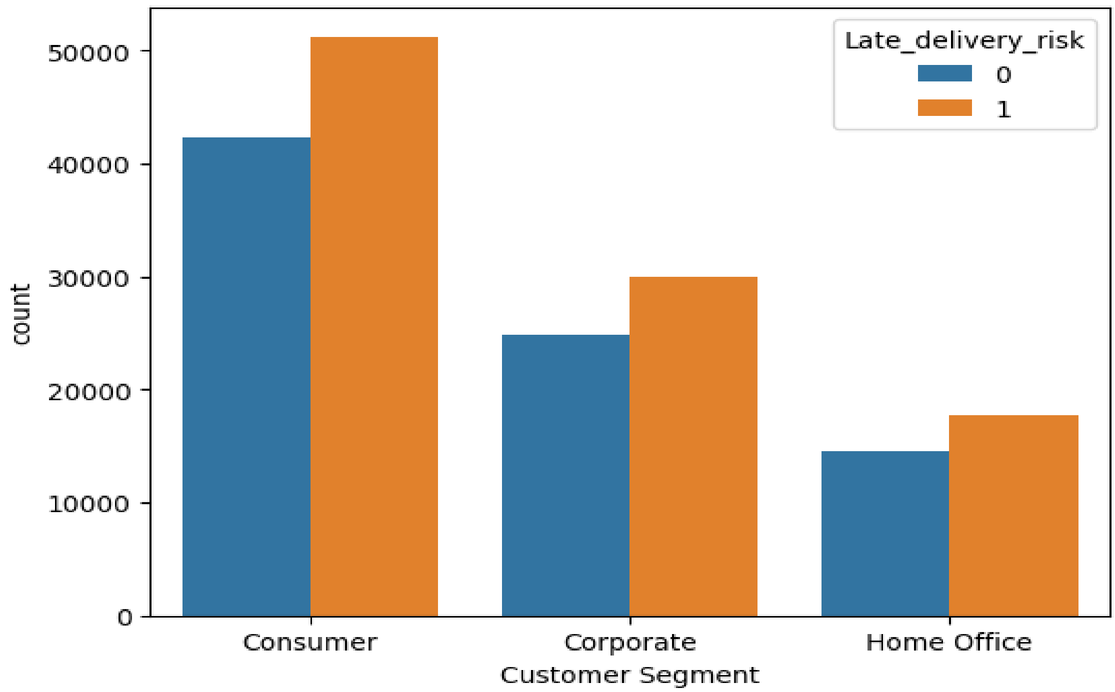

Figure 4.11: Distribution of Late Delivery Risk by Customer Segment

Figure 4.11 presents the distribution of Late Delivery Risk across various customer segments, with specific counts for on-time and late deliveries. The clustered bar chart displays the counts for each segment ("Consumer," "Corporate," and "Home Office") separated by delivery status (Late Delivery Risk = 0 for on-time, Late Delivery Risk = 1 for late). For the "Consumer" segment, there were approximately 42,000 on-time deliveries and 52,000 late deliveries. The "Corporate" segment had roughly 25,000 on-time and 30,000 late. The "Home Office" segment registered about 14,000 on-time and 18,000 late deliveries. This visualization highlights that while the "Consumer" segment has the highest overall volume, all segments demonstrate a higher number of late deliveries compared to on-time, suggesting a systemic issue affecting all customer groups.

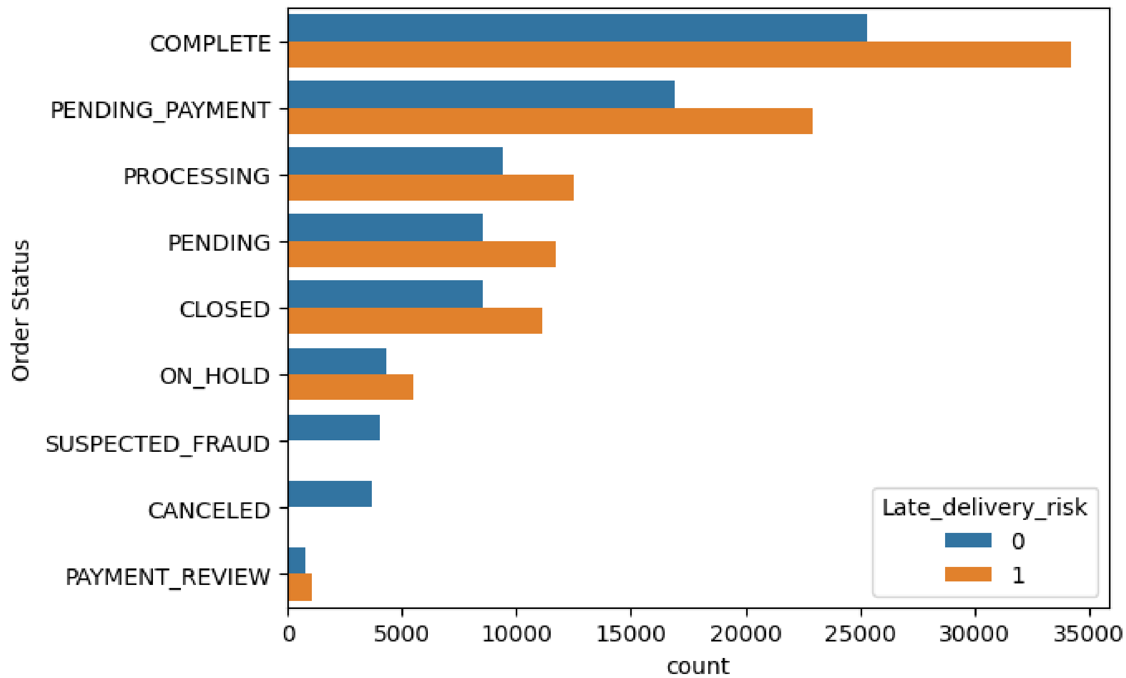

Figure 4.12: Distribution of Late Delivery Risk by Order Status

A distribution of the Late Delivery Risk across various order statuses is shown in Figure 4.11. Whether an order was delivered late (Late Delivery Risk = 1) or not (Late Delivery Risk = 0) is broken down in the stacked horizontal bar chart that displayed the count of orders for every status ("COMPLETE," "PENDING_PAYMENT," "PROCESSING," "PENDING," "CLOSED," "ON_HOLD," "SUSPECTED_FRAUD," "CANCELLED," "PAYMENT review"). With the biggest volume and the clear majority of on-time delivery shown by "COMPLETE" orders. Nevertheless, there is a greater rate of late deliveries than on-time deliveries for a number of statuses, such as "PENDING_PAYMENT," "PROCESSING," "PENDING," and "CLOSED." Orders marked as "ON_HOLD" or "SUSPECTED_FRAUD" also tend to have delivery delays. Although there are less "CANCELLED" and "PAYMENT_REVIEW" orders overall, there is a noticeable pattern of more late deliveries than on-time deliveries for these types of purchases. In order to enhance on-time delivery performance, this visualization highlights possible areas in the order fulfilment process that may need attention based on the order state and the Late Delivery Risk

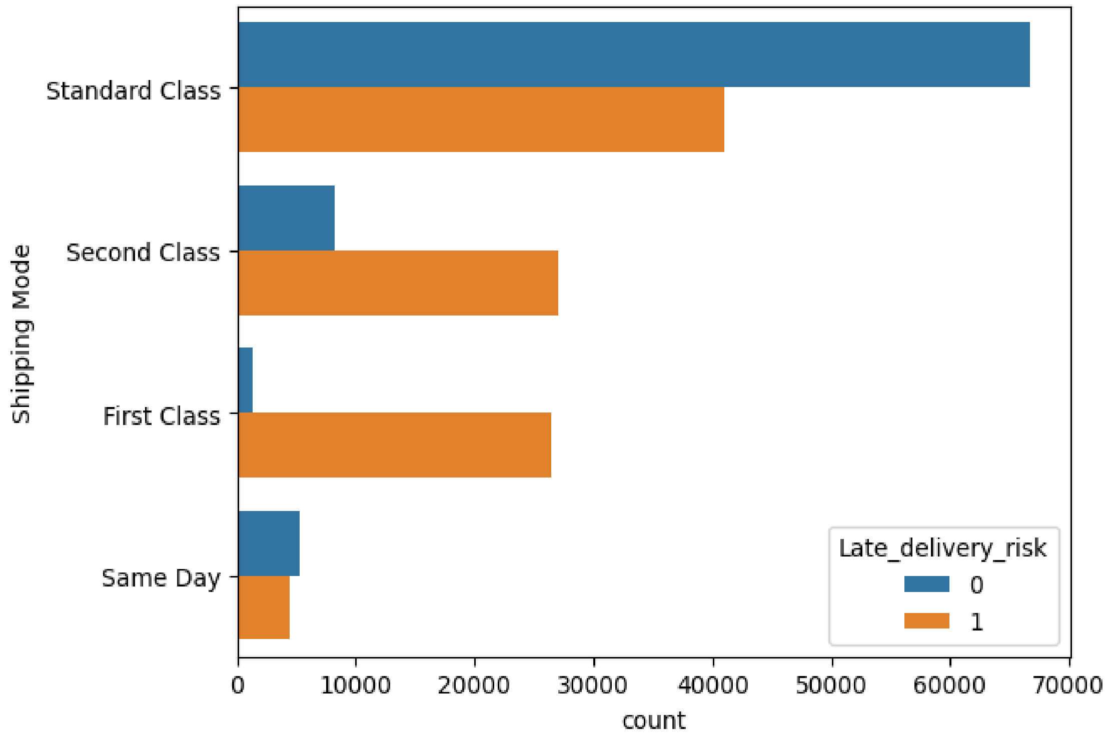

Figure 4.13: Distribution of Late Delivery Risk by Shipping Mode

Figure 4.13 shows how Late Delivery Risk is distributed across various shipment options. The stacked horizontal bar chart shows the order count for shipping modes ("Standard Class", "Second Class", "First Class", "Same Day") which are separated into orders with and without late delivery (Late Delivery Risk = 1/0). Standard Class experiences the biggest order volume with multiple delivery delays but delivers most orders within schedule. The distribution of "Second Class" orders includes an especially significant number of late deliveries while on-time deliveries remain lower. Among the shipping methods "First Class" displays balanced timing patterns because its total number of on-time deliveries somewhat surpasses delivery delays. The shipping service "Same Day" contains the lowest number of orders combined with fewer late deliveries but its delivery delays as a percentage exceed those observed within the "First Class" service. The visual data indicates that delivery method affects risk of delay since both "Standard Class" and "Second Class" labels show higher susceptibility to delivery delays.

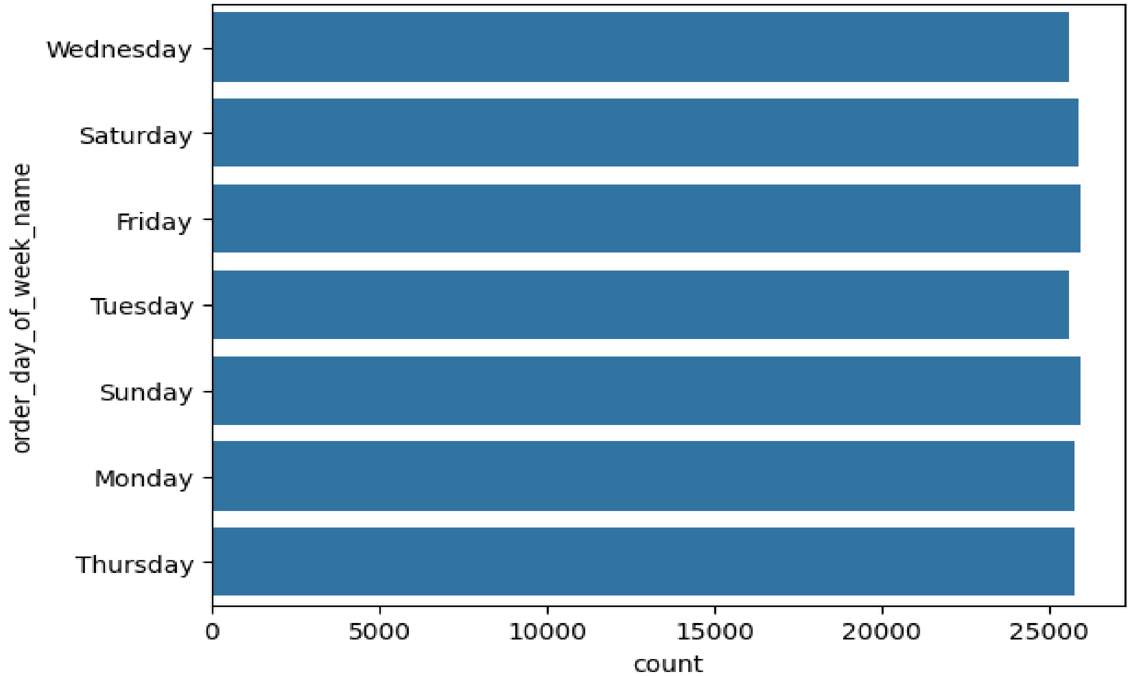

Figure 4.14: Distribution of Orders Across Days of the Week

Figure 4.14 displayed a Count Plot for the "order_day_of_week_name" feature, showing the distribution of orders across the days of the week. The horizontal bars represent the count of orders for each day, with the days arranged in descending order of frequency. "Thursday" exhibits the highest number of orders, closely followed by "Wednesday". "Tuesday" and "Saturday" show similar counts, slightly lower than the top two days. "Friday" and "Monday" also have comparable counts, while "Sunday" registers the lowest order volume. This visualization reveals a potential pattern in order frequency across the week, with mid-weekdays generally having higher order counts compared to weekends.

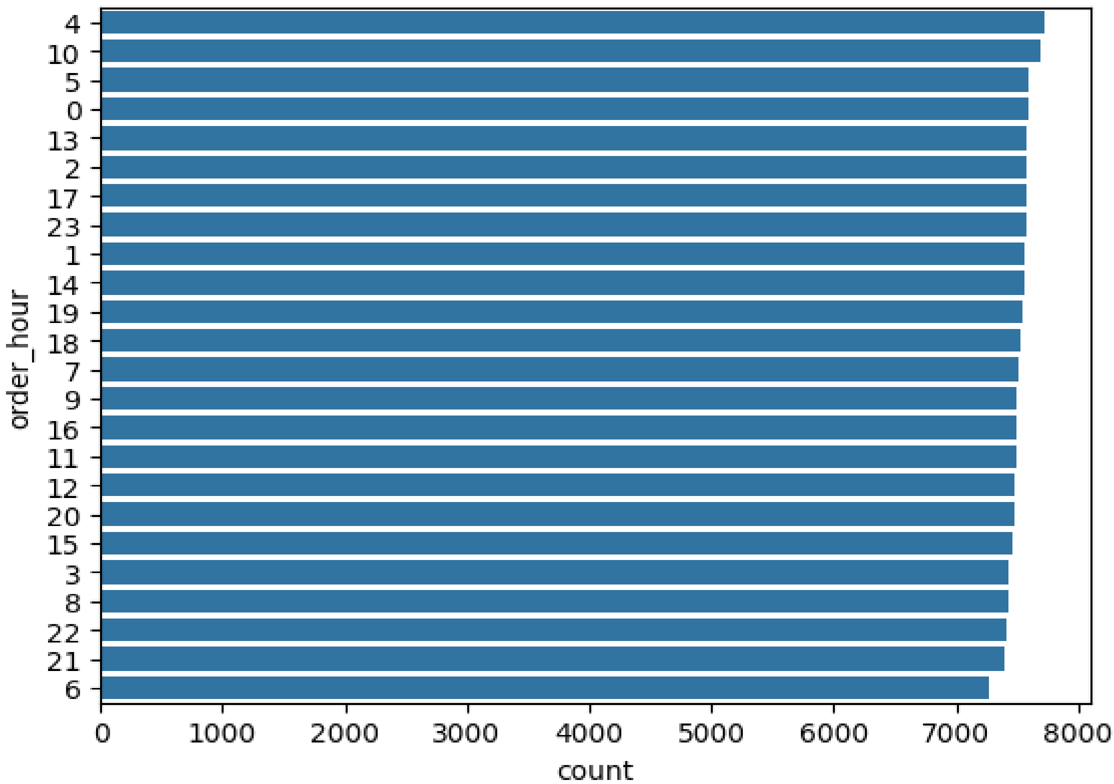

Figure 4.15: Distribution of Order Hours

Figure 4.15 displayed a Count Plot for the distribution of order hours, showing the frequency of orders placed at each hour of the day. The horizontal bars represent the count of orders for each hour, numbered from 0 to 23. The plot reveals a generally consistent volume of orders across most hours, with a slight dip observed at hour 6, which corresponds to 6 AM. The hours 4, 5, 10, 0, 13, 2, 17, and 23 (representing 4 AM, 5 AM, 10 AM, 12 AM, 1 PM, 2 AM, 5 PM, and 11 PM respectively) show the highest order counts, all clustered closely together around the 7500-8000 order mark. This visualization suggests a relatively even distribution of order activity throughout the day, with a minor decrease in the early morning hours.

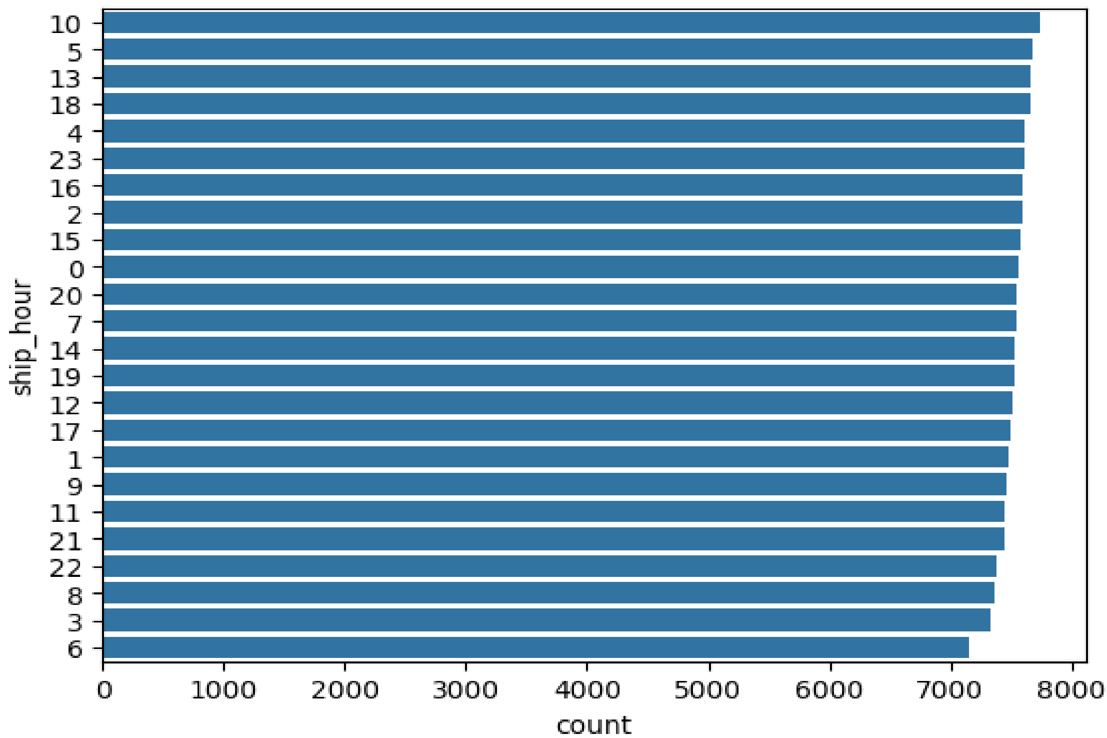

Figure 4.16; Distribution of Shipping Hours

Figure 4.16 displayed a Count Plot for the distribution of shipping hours, showing the frequency of shipments occurring at each hour of the day. The horizontal bars represent the count of shipments for each hour, numbered from 0 to 23. The plot reveals a generally consistent volume of shipments across most hours, with a slight dip observed at hour 6, corresponding to 6 AM. The hours 4, 5, 10, 13, 18, 0, 23, 16, 2, 15, and 20 (representing 4 AM, 5 AM, 10 AM, 1 PM, 6 PM, 12 AM, 11 PM, 4 PM, 2 AM, 3 PM and 8 PM respectively) show the highest shipment counts, all clustered closely together around the 7000-8000 shipment mark. This visualization suggests a relatively even distribution of shipping activity throughout the day, with a minor decrease in the early morning hours.

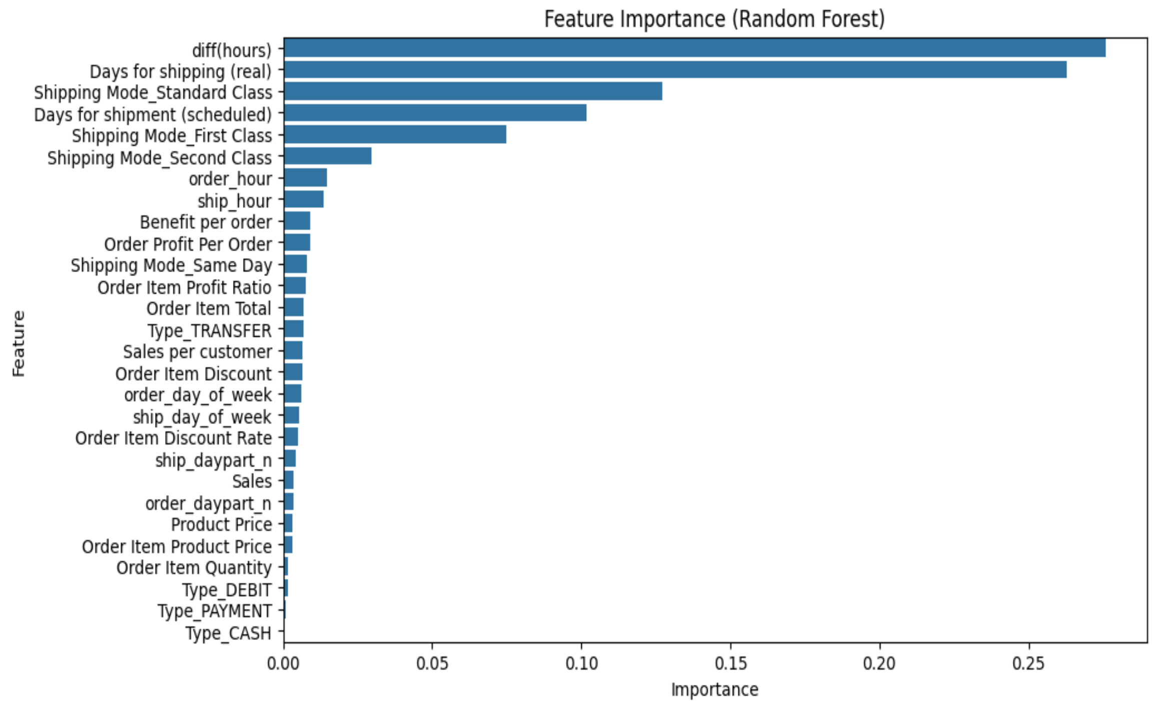

Figure 4.17: Feature Importance Bar Graph

Figure 4.17 presents a feature importance bar graph derived from a random forest model, highlighting the top contributing factors in a dataset. "diff(hours)" emerges as the most influential feature, boasting the highest importance score exceeding 0.25. "Days for shipping (real)" also demonstrates significant importance, registering a value around 0.24. "Shipping Mode Standard Class" and "Days for shipment (scheduled)" contribute moderately, with importance scores above 0.10. The subsequent features, including "Shipping Mode First Class," "Shipping Mode Second Class," "order_hour," and "ship hour," exhibit lower yet still notable importance. This visualization emphasizes the key predictive variables identified by the Random Forest model, providing insights into which factors most strongly influence the target variable.

Exploratory Data Analysis for Demand Forecasting Using Proposed Dataset

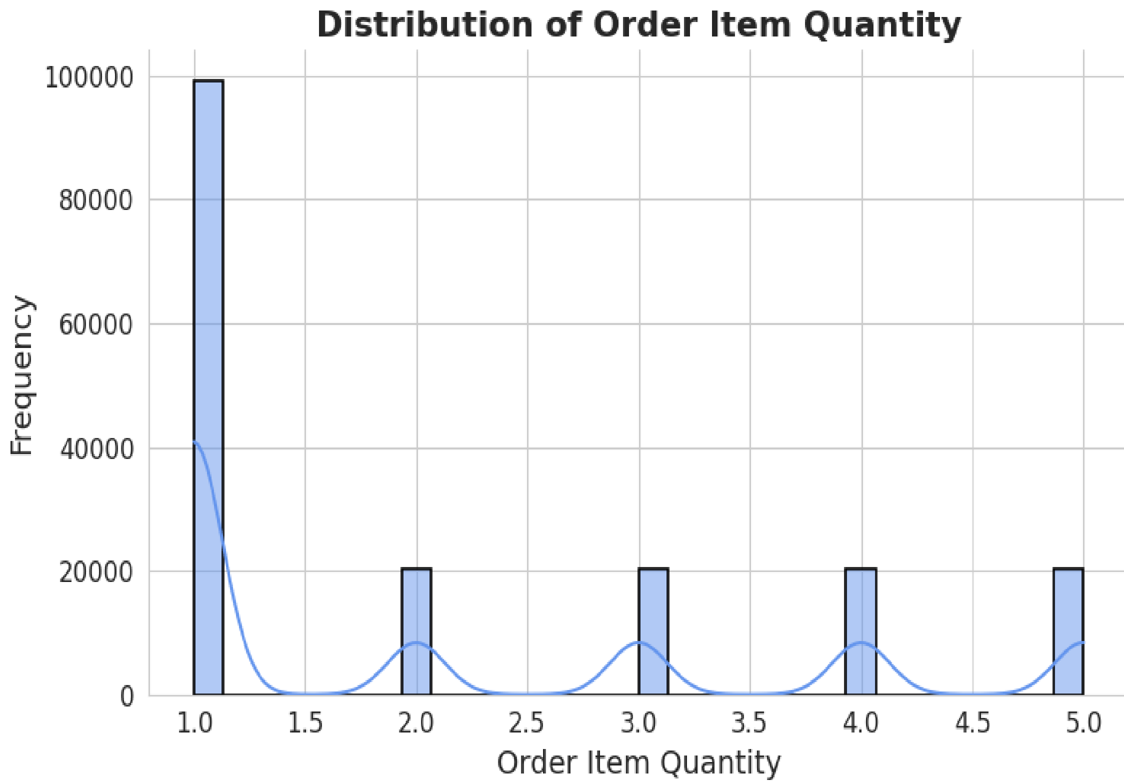

Figure 4.18: Distribution of Order Item Quantity

Figure 4.18 displays the distribution of order item quantities, revealing a strong preference for single-item orders. The data, sourced by the DataCo SMART SUPPLY CHAIN dataset, shows a prominent spike at a quantity of 1, representing approximately 100,000 orders. Smaller, yet notable, peaks appear at quantities of 2, 3, 4, and 5, each hovering around 20,000 orders. This visualization highlights a clear trend: customers predominantly order single items, with a significantly reduced frequency of orders for multiple items, suggesting potential implications for inventory management and order fulfillment strategies within the supply chain.

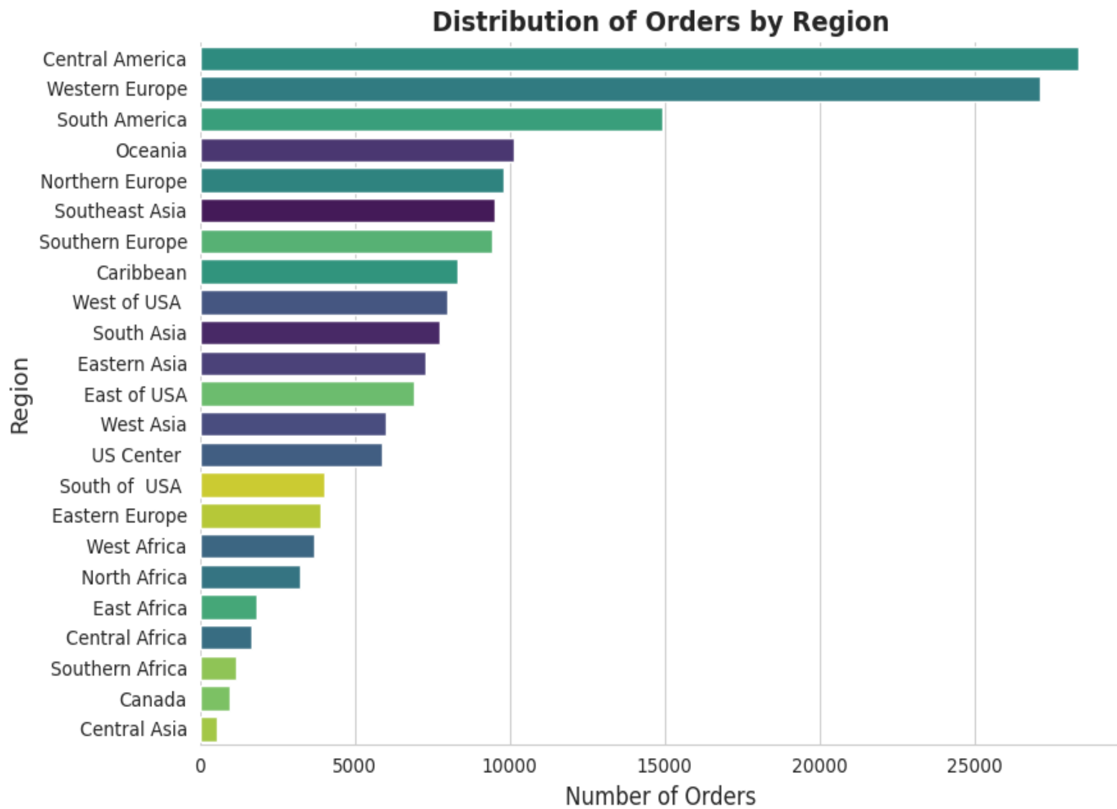

Figure 4.19: Distribution of Orders by Region

Figure 4.19 displayed a Barograph for the distribution of orders across various regions, revealing significant disparities in order volumes. Western Europe leads with approximately 26,000 orders, closely followed by South America with around 16,000. Oceania and Northern Europe show similar order frequencies, nearing 10,000. In contrast, regions like Central Asia, Canada, and Southern Africa exhibit considerably lower order volumes, with Central Asia registering the fewest orders at just over 1,000. This visualization underscores the geographical variations in order patterns, highlighting key markets and areas with potential for growth or requiring targeted strategies.

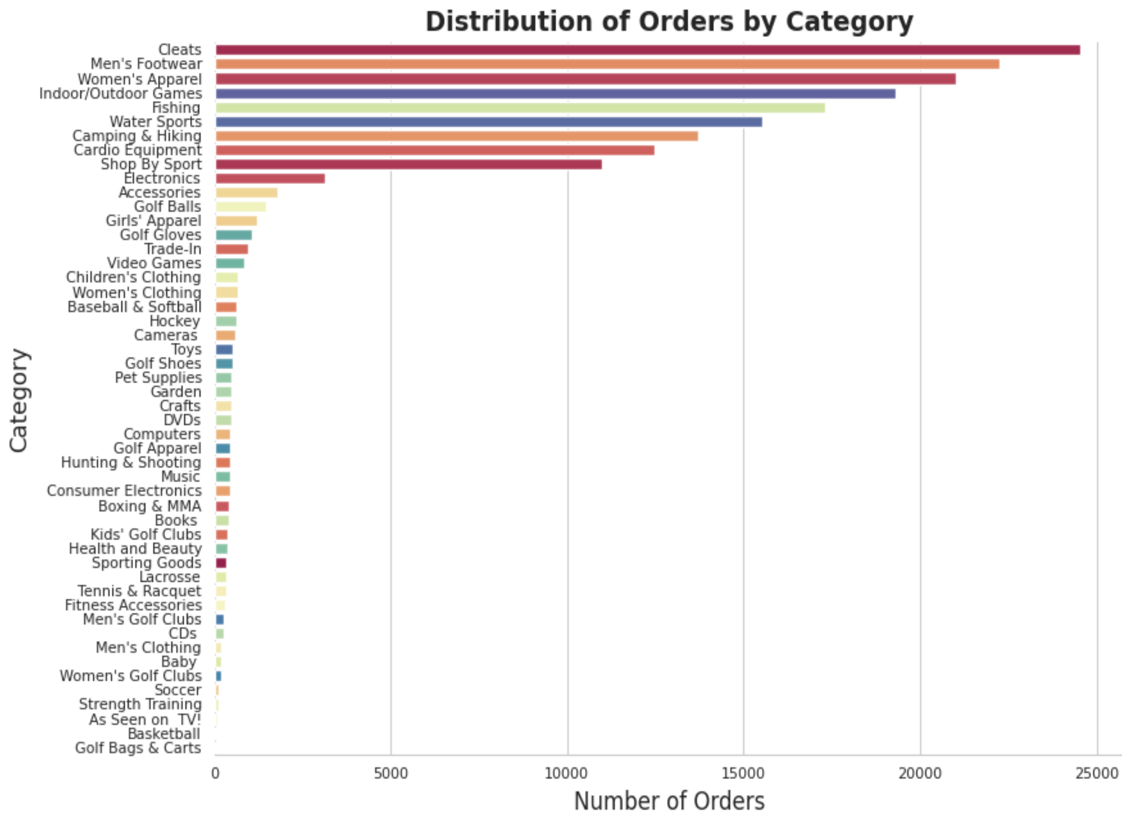

Figure 4.20: Count plot for Distribution of Orders by Category

Figure 4.20 showcases a distribution of orders across various product categories using a count plot. The categories "Cleats," "Men’s Footwear," and "Women’s Apparel" emerge as the top three, each exhibiting order counts exceeding 20,000. Conversely, categorize such as "Golf Bags & Carts," "Basketball," and "As Seen on TV!" show significantly lower order volumes, falling below 5,000. The plot effectively visualizes the varying popularity of different product categories, highlighting best-selling items and those with potentially lower demand, offering valuable insights for inventory management and marketing strategies.

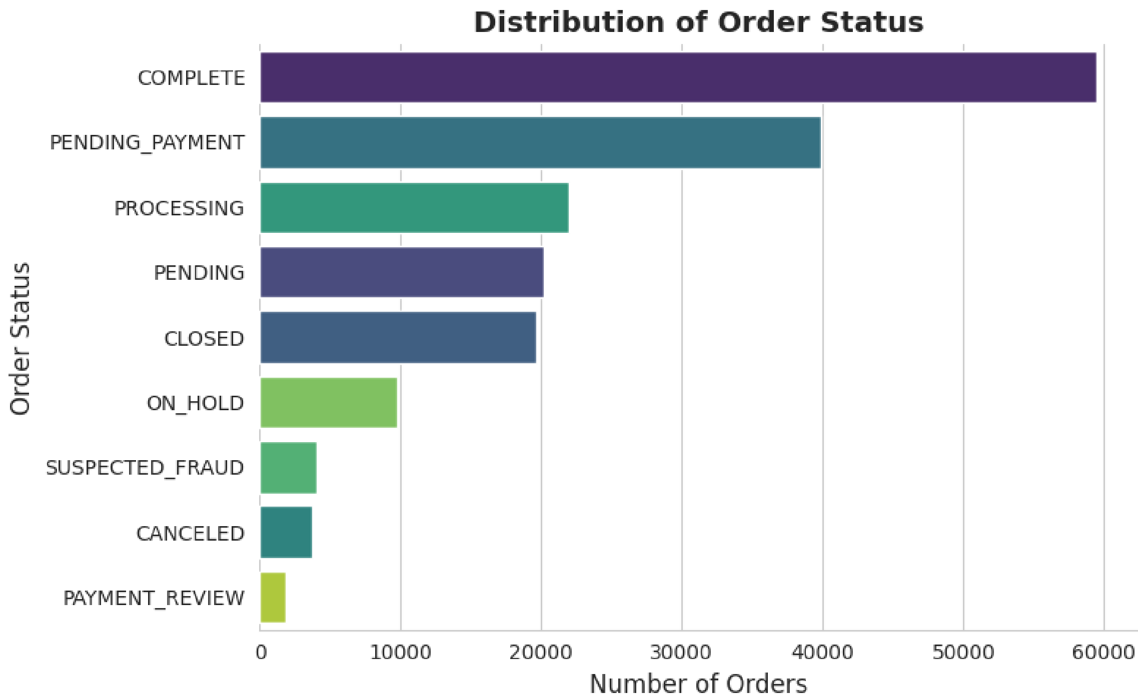

Figure 4.21: Count plot for Distribution of Order Status

Figure 4.21 illustrates a distribution of order statuses, revealing that the majority of orders are "COMPLETE," with a count exceeding 50,000. "PENDING_PAYMENT" represents the next largest category, nearing 40,000 orders. Following this, "PROCESSING" and "PENDING" statuses each account for approximately 20,000 orders. Conversely, statuses like "SUSPECTED_FRAUD," "CANCELED," and "PAYMENT_REVIEW" show significantly lower frequencies, with counts below 10,000, indicating these are less common occurrences. This visualization highlights the overall efficiency of the order process, with a large proportion of completed orders, while also identifying areas, such as pending payments, that may require attention.

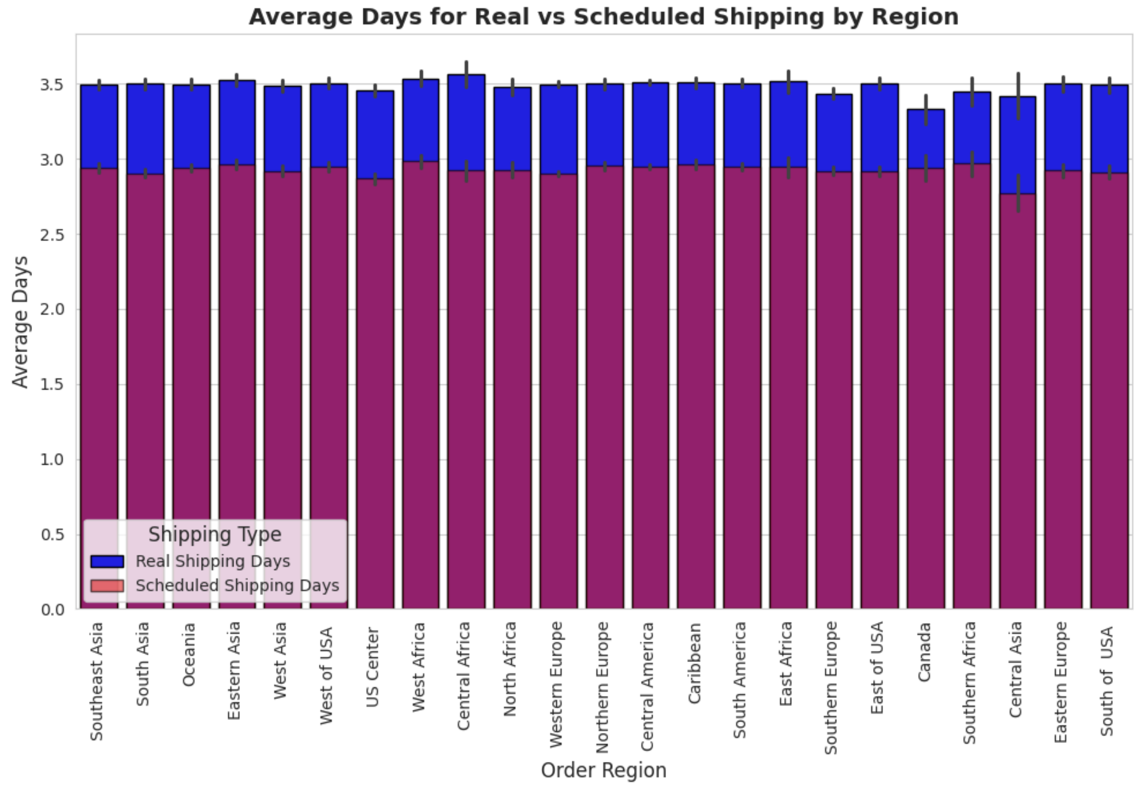

Figure 4.22: Average Days for Real vs Scheduled Shipping By Reason

Figure 4.22 is a bar plot comparing average real shipping days against scheduled shipping days across different regions. The stacked bars, each representing a region, show that in most areas the real shipping time (represented by the bottom portion of the bar in shades of purple/red) is slightly less than the scheduled shipping time (represented by the top portion of the bar in blue). Specifically, regions like Southeast Asia, South Asia, and Oceania show a noticeable difference where scheduled times are longer. The graph also highlights that while there are variations in average shipping times across regions, the scheduled times generally provide a buffer over the actual shipping durations. Additionally, the error bars on the scheduled shipping indicate the variability in scheduled delivery times.

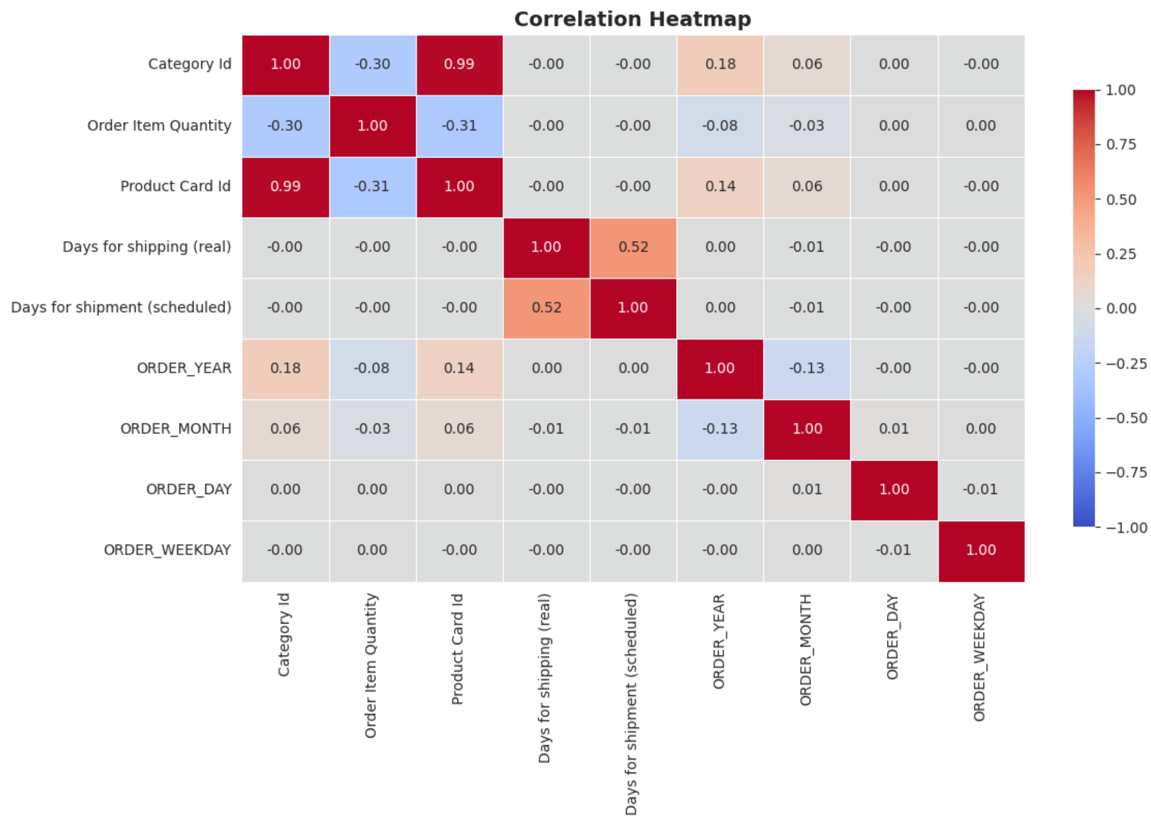

Figure 4.23: Correlation Heatmap For Selected Numerical Features of DataCo Smart Supply Chain Dataset

Figure 4.23 displays a Correlation Heatmap for selected numerical features by the DataCo SMART SUPPLY CHAIN dataset. Using a color-coded scale that goes from -1 to 1, the heatmap graphically displays the correlation coefficients among pairs of variables. Notable positive correlations are observed between 'CategoryId' and 'Product Card Id' (0.99), and between 'Days for shipping (real)' and 'Days for shipment (scheduled)' (0.52). Conversely, a strong negative correlation (-0.75) exists between 'ORDER_WEEKDAY' and 'ORDER_DAY', and a moderate negative correlation (-0.50) is seen between 'Product Card Id' and 'ORDER_MONTH'. The heatmap provides a clear overview of potential relationships between variables, informing feature selection and model development in subsequent analytical tasks.

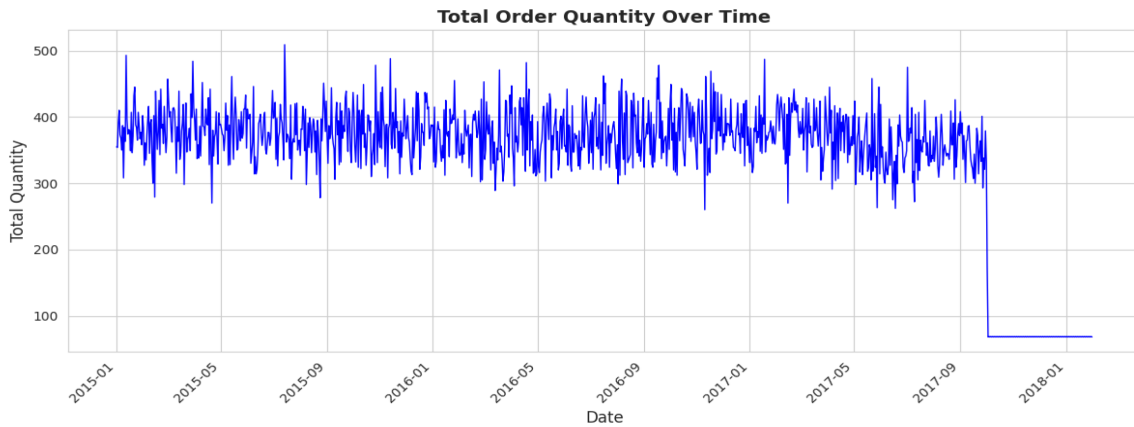

Figure 4.24: Line plot for Total Order Quantity Over Time

Figure 4.24 depicts the total order quantity over time using a line plot, showcasing fluctuations and trends in order volume. The plot reveals a generally stable pattern with order quantities hovering around the 400 marks for much of the period between 2015 and late 2017. However, a noticeable drop occurs towards the end of 2017, where the order quantity sharply declines to approximately 50 units and remains low into 2018. This visualization effectively illustrates the temporal dynamics of order quantities, highlighting a significant decrease in late 2017 which may warrant further investigation into potential causes or contributing factors.

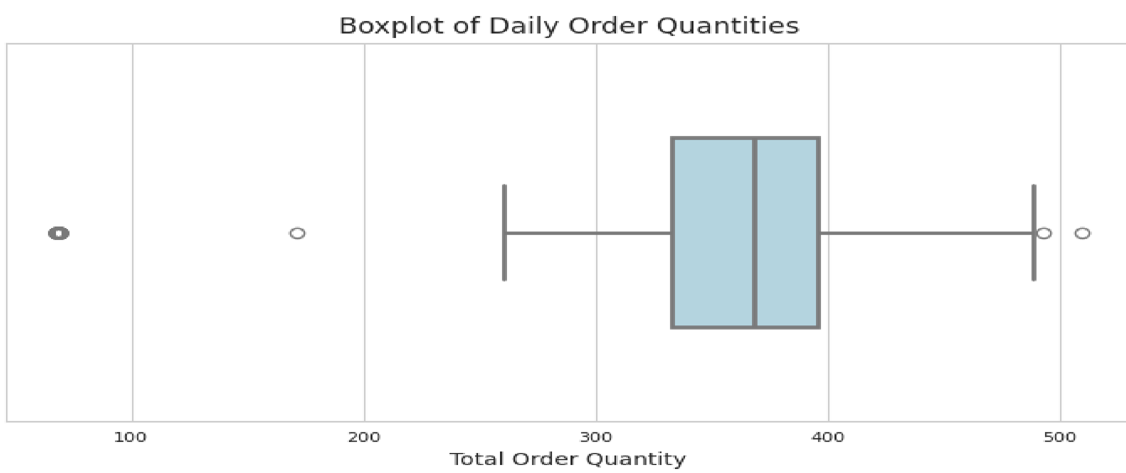

Figure 4.25: Boxplot of Daily Order Quantities

Figure 4.25 presents a boxplot of daily order quantities, visually summarizing the distribution and identifying potential outliers. The box extends from approximately 325 to 425 units, encompassing the IQR and containing the middle 50% of a data. The box has a line that represents the median daily order quantity, which is around 375. The whiskers indicate the range of usual daily order volumes, extending to the lowest and largest values within 1.5 times the IQR by the box. Notably, two distinct outliers are visible: one low value around 75 and one high value just above 500. This visualization highlights the central tendency and spread of daily order quantities, revealing a generally consistent order volume with occasional unusually low or high days.

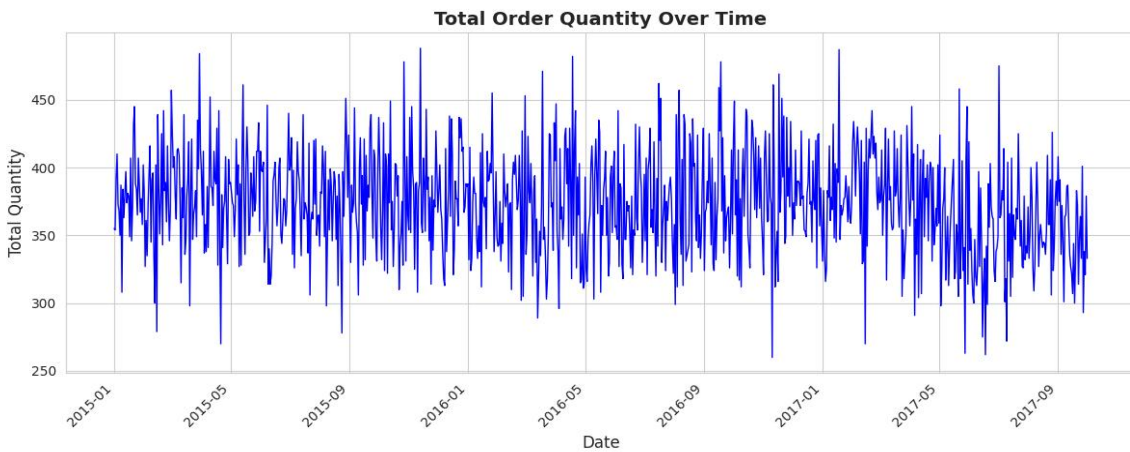

Figure 4.26: Total Order Quantity Over Time (Without Outliers)

Figure 4.26 illustrates the trend of Total Order Quantity Over Time, focusing on the fluctuations in order volume while excluding outliers to provide a clearer view of the overall pattern. The plot shows a generally consistent level of order quantities, oscillating around an average of approximately 400 units throughout the depicted period. While there are minor peaks and valleys indicating variations in demand, the overall trend suggests a relatively stable order volume with no significant upward or downward shifts. This visualization helps in understanding the typical range of order quantities and identifying any potential cyclical patterns or anomalies that deviate from the established norm, excluding extreme values that might skew the interpretation of regular operational performance.

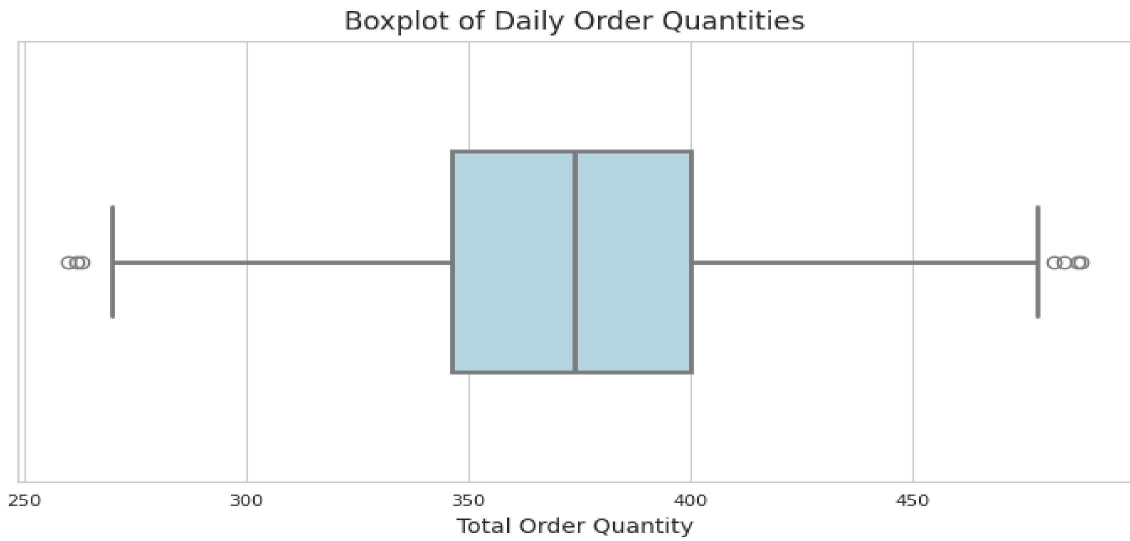

Figure 4.27: Boxplot of Daily Order Quantities (Without Outliers)

Figure 4.27 displays a boxplot of daily order quantities, excluding outliers, to illustrate the typical distribution of order volumes. The center half of the data is included in the IQR, which is represented by the central box, which runs from around 350 to 400 units. Within the box, you can see a line that indicates the median daily order quantity, which is probably about 375. As a visual representation of the range of usual daily order volumes, the whiskers reach the lowest and highest values within 1.5 times the IQR from the box. This visualization emphasizes the consistent nature of daily order quantities, clustering around a central value with a relatively narrow spread, reinforcing the observation of stable demand with minimal fluctuations outside the typical range.

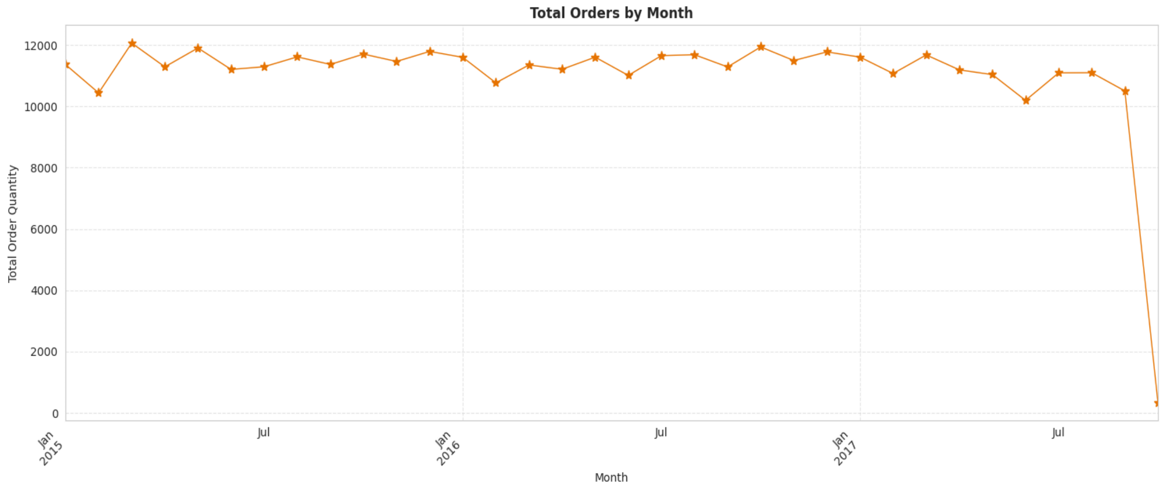

Figure 4.28: Line graph for Total Orders by Month

Figure 4.28 displays the total orders by month using a line graph, revealing a generally stable trend with some fluctuations. Throughout the majority of the period depicted, total orders fluctuate around the 10,000 to 12,000 marks, exhibiting a cyclical pattern of minor peaks and troughs. However, a significant decline is noticeable towards the end of the time frame, with a sharp drop in total orders to below 2,000. This visualization reveals steady order volume across most of the timeseries data with a clear reduction in numbers that hints at order fulfillment issues in that specific time frame.

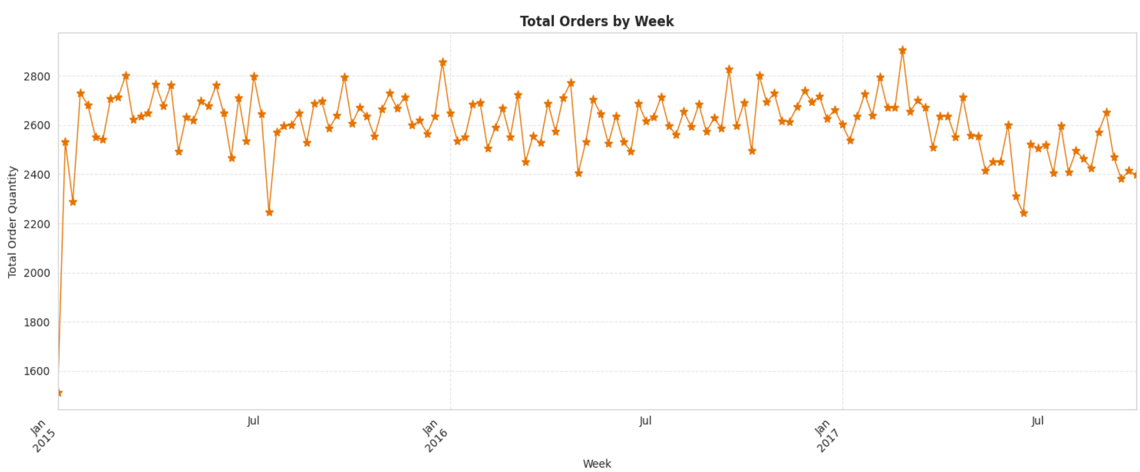

Figure 4.29: Line graph for Total Orders by Week

The total orders by week present a steady pattern with occasional changes according to the line graph in Figure 4.29. The majority of the depicted timeframe shows total orders continuously moving between 2,400 and 2,800 marks while displaying small cyclical patterns. A main decrease emerges initially when orders reach numbers under 2,000 before rising back up fast. The visualization brings to light steady order volume patterns across most of the time period and shows how starting orders suffered a big yet short-term decrease that potentially demonstrates system obstacles during that timeframe.

4.4 Experimental Results

This section evaluates AdaBoost, Cat Boost, and MLP models for delivery status prediction, late risk prediction, and demand forecasting. Classification tasks were assessed using accuracy, precision, recall, and F1score, while demand forecasting used MAE, MSE, and RMSE.

Findings for Delivery Status Prediction

This subsection analyzes the model performance in predicting delivery status (on-time, delayed, etc.). The classification models are assessed using the following metrics: accuracy, precision, recall, and F1score. The results highlight the model that best predicts delivery status, aiding logistics optimization.

Analysis of Proposed AdaBoost Model

Table 4.1: Performance Metrics of AdaBoost Model in Delivery Status Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
AdaBoost	77.64	77.80	77.64	76.99

The above-mentioned Table 4.1 presents a performance metrics of the AdaBoost model specifically for delivery status prediction. The table displays the accuracy, precision, recall, and F1score achieved by the AdaBoost model, with the recorded values being

77.64%, 77.80%, 77.64%, and 76.99% respectively. The AdaBoost model's performance in identifying delivery statuses may be measured quantitatively using these measures.

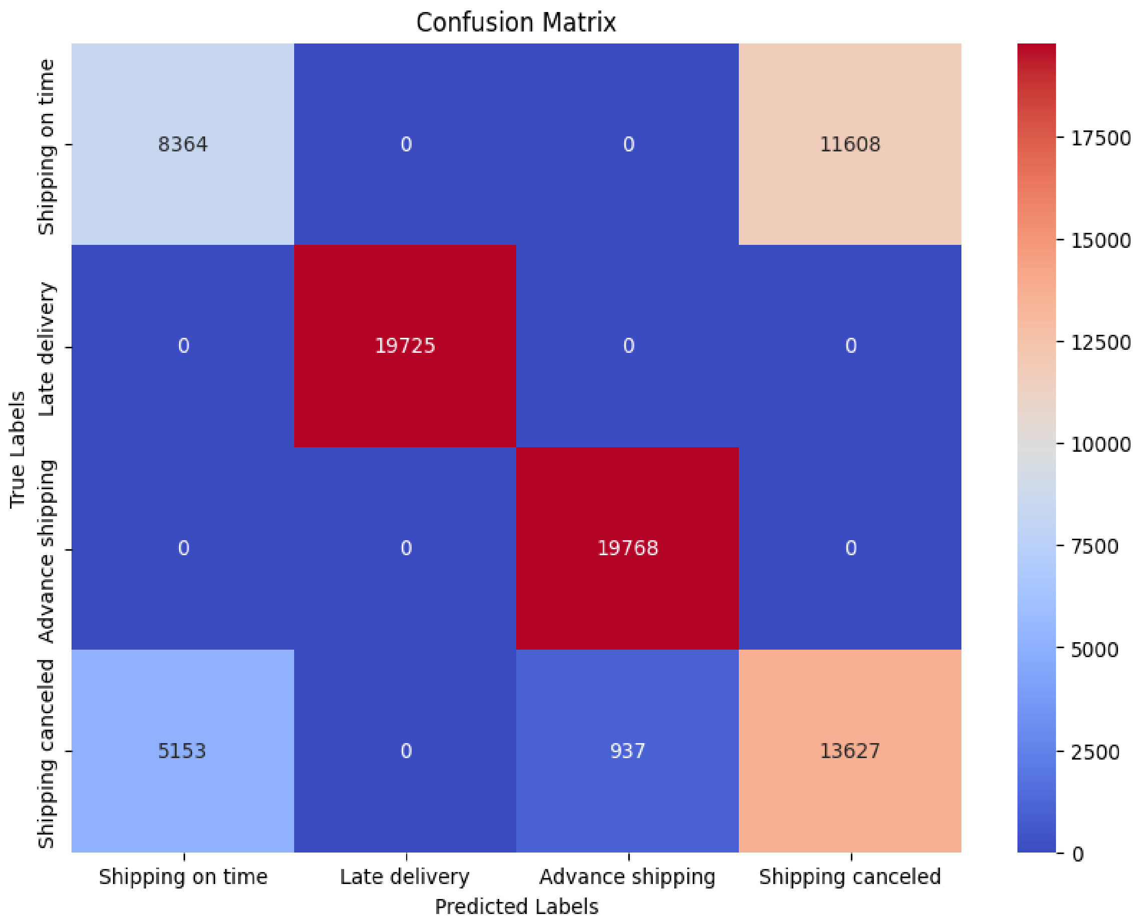

Figure 4.30: Confusion Matrix of AdaBoost Model for Delivery Status Prediction

Figure 4.30 displays the Confusion Matrix for the AdaBoost model and shows the correctly classified instances for each delivery status category. For "Shipping on time", the model correctly predicted 8364 instances. It accurately classified 19725 instances as "Late delivery" and 19768 instances as "Advance shipping". Finally, the model correctly identified 5153 instances as "Shipping canceled".

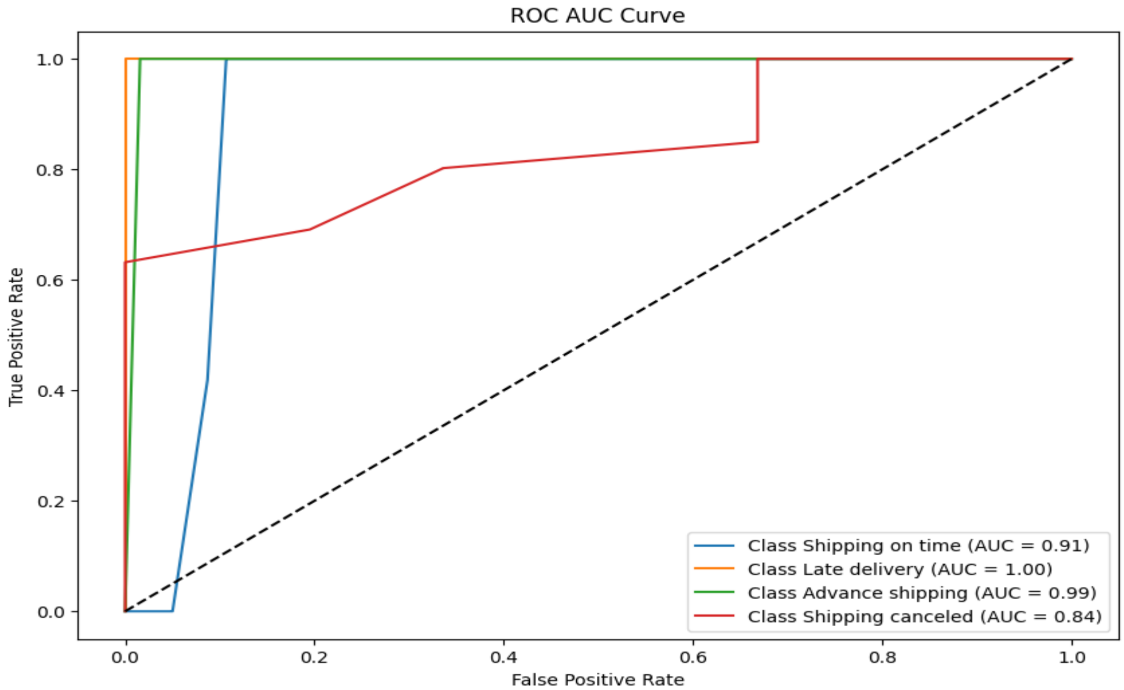

Figure 4.31: ROC Curve of AdaBoost Model for Delivery Status Prediction

Figure 4.31 displayed the ROC curves for the AdaBoost model's delivery status predictions across four classes. Every curve plots the TPR against the FPR for various classification thresholds. The AUC values indicate a model's discriminative ability for every class: Shipping on time (0.91), Late delivery (1.00), Advance shipping (0.99), and Shipping canceled (0.84), reflecting strong overall performance.

Analysis of Proposed Cat Boost Model

Table 4.2: Performance Metrics of Cat Boost Model in Delivery Status Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Cat Boost	100	100	100	100

The above-mentioned Table 4.2 displays the performance evaluation metrics for the Cat Boost model applied to delivery status prediction. The chart shows that the Cat Boost model performed very well in identifying delivery statuses, as it got perfect scores across all assessed parameters, including accuracy (100), precision (100), recall (100), and F1score (100).

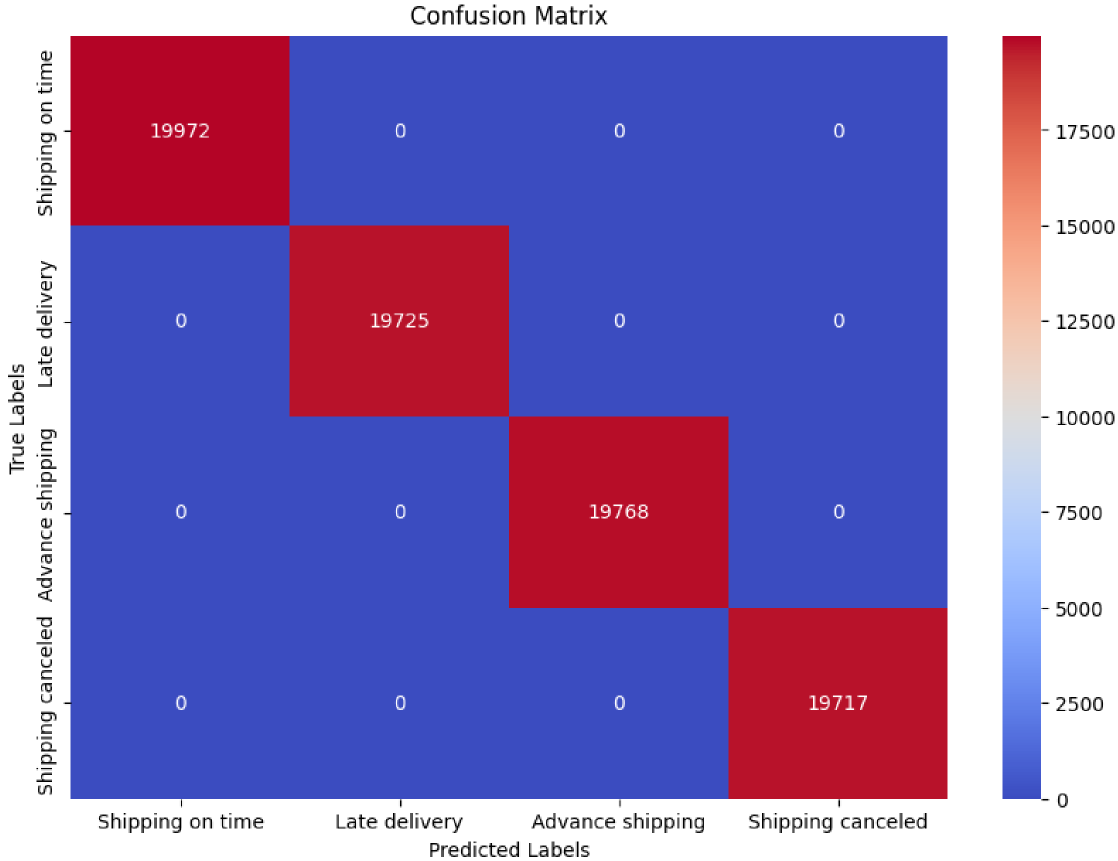

Figure 4.32: Confusion Matrix of CatBoost Model for Delivery Status Prediction

Figure 4.32 displays the confusion matrix for the Cat Boost model's performance in delivery status prediction. The number of cases that were accurately identified for each category is highlighted in the matrix: 19972 for "Shipping on time", 19725 for "Late delivery", 19768 for "Advance shipping", and 19717 for "Shipping canceled". These diagonal entries represent the accurate predictions made by the model across all delivery outcomes.

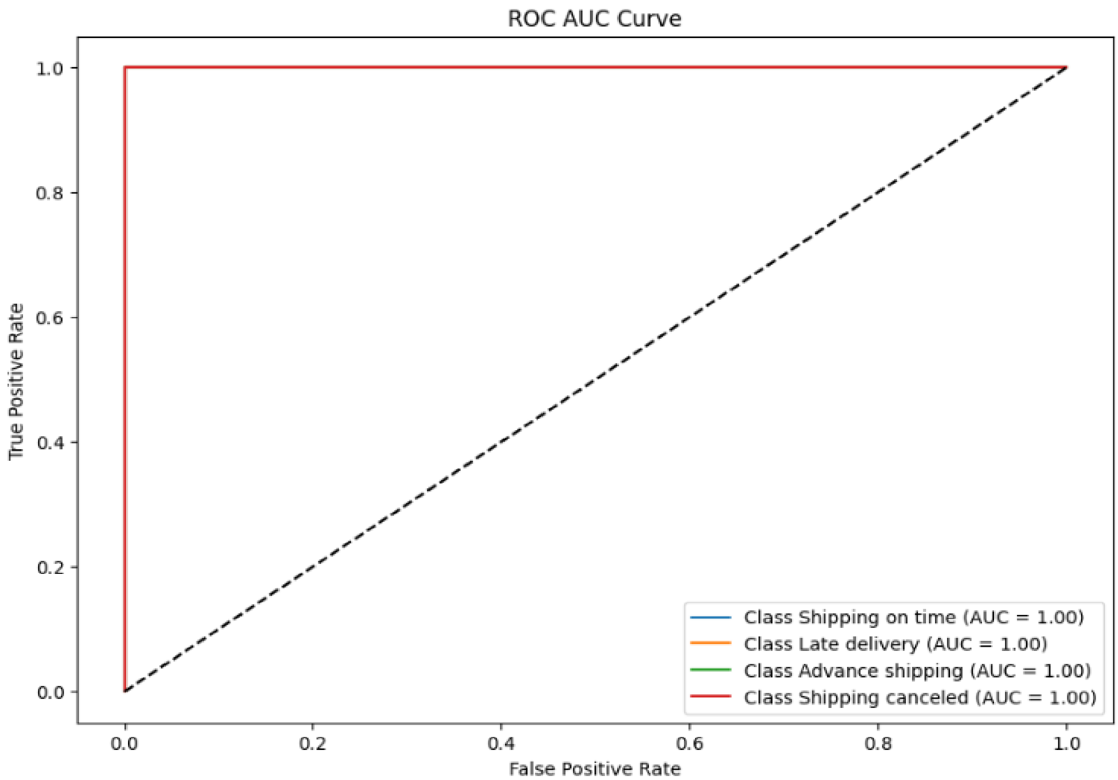

Figure 4.33: ROC Curve of CatBoost Model for Delivery Status Prediction

Figure 4.33 illustrates the ROC curves for the Cat Boost model's performance in predicting four delivery statuses. Each class, "Shipping on time", "Late delivery", "Advance shipping", and "Shipping canceled", achieves a perfect AUC score of 1.00. This proves that the Cat Boost model is quite good at differentiating between good and bad examples across all types of delivery outcomes.

Analysis of Proposed MLP Model

Table 4.3: Performance Metrics of MLP Model in Delivery Status Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
MLP	99.98	99.98	99.98	99.98

The above-mentioned Table 4.3 presents the performance metrics obtained by the Multi-Layer Perceptron (MLP) model for the task of delivery status prediction. The table summarizes the model's performance across 4 key metrics: accuracy, precision, recall, and F1score, all of which are reported as 99.98%, indicating a near-perfect classification performance on the delivery status prediction task.

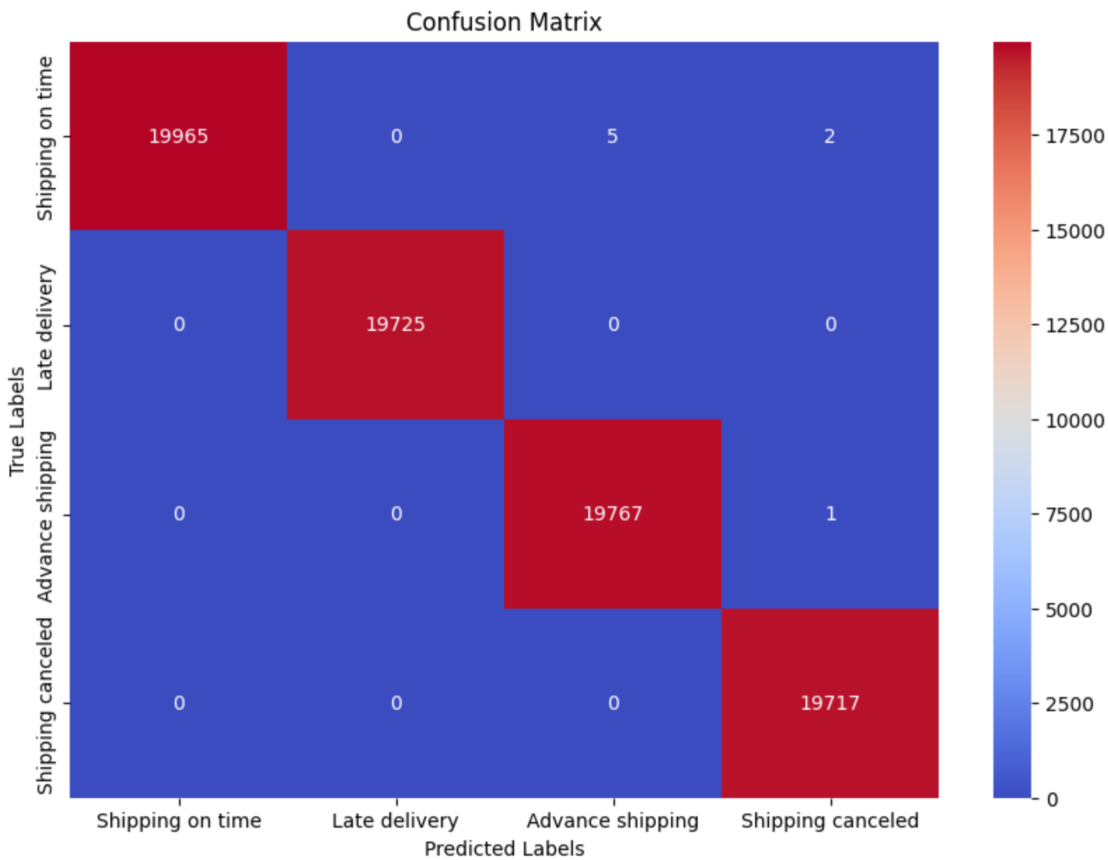

Figure 4.34: Confusion Matrix of MLP Model for Delivery Status Prediction

Figure 4.34 displayed the Confusion Matrix for the MLP model's delivery status predictions. The diagonal values represent correctly classified instances: 19965 for "Shipping on time", 19725 for "Late delivery", 19767 for "Advance shipping", and 19717 for "Shipping canceled". These numbers indicate the model's accuracy in correctly identifying each delivery outcome.

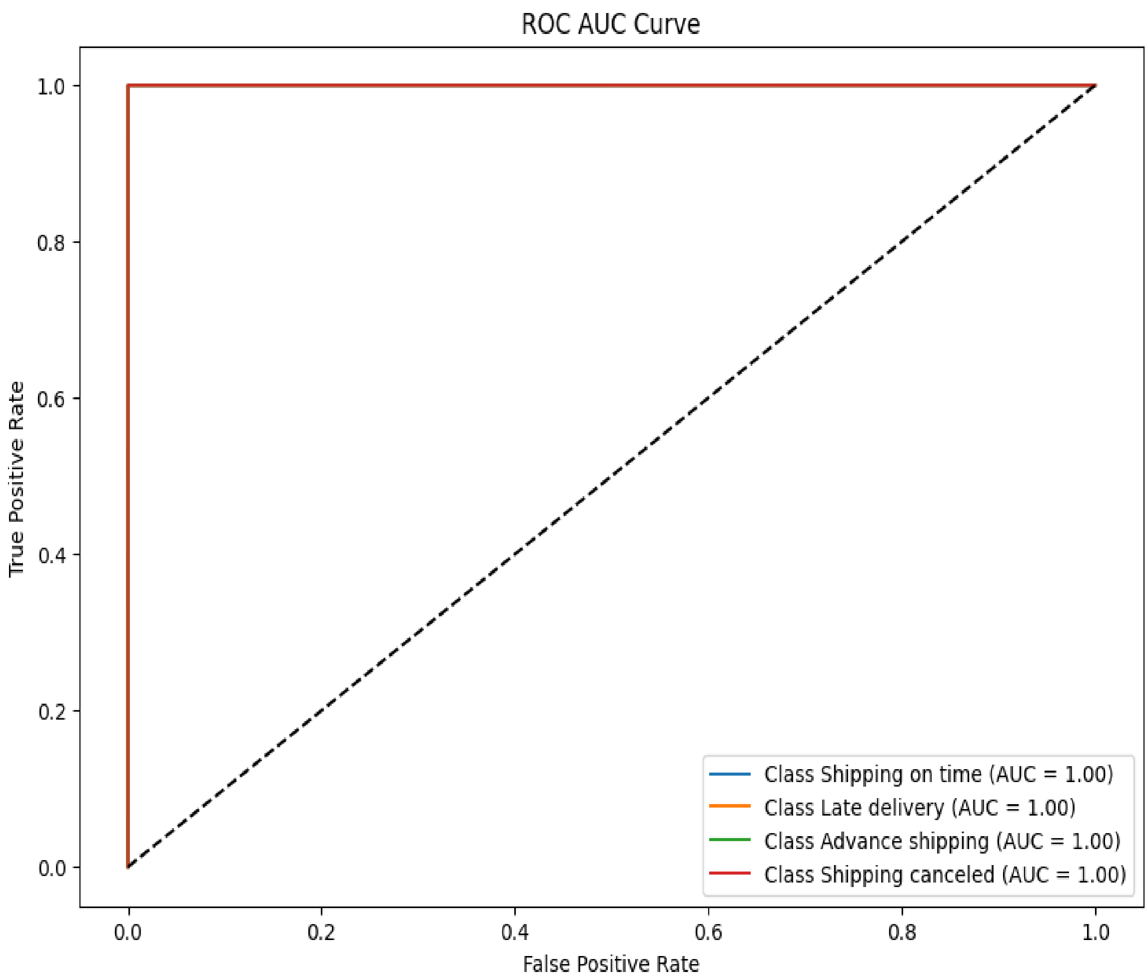

Figure 4.35: ROC Curve of MLP Model for Delivery Status Prediction

Figure 4.35 displayed the ROCAUC curves for the MLP model's delivery status prediction across four classes. Each class, namely "Shipping on time", "Late Delivery", "Advance Shipping", and "Shipping Canceled", exhibits a perfect AUC of 1.00. This indicates that the MLP model achieved ideal discriminative performance, effectively distinguishing among positive and negative instances for all delivery status categories.

Comparative Analysis of Proposed ML Models for Delivery Status Prediction

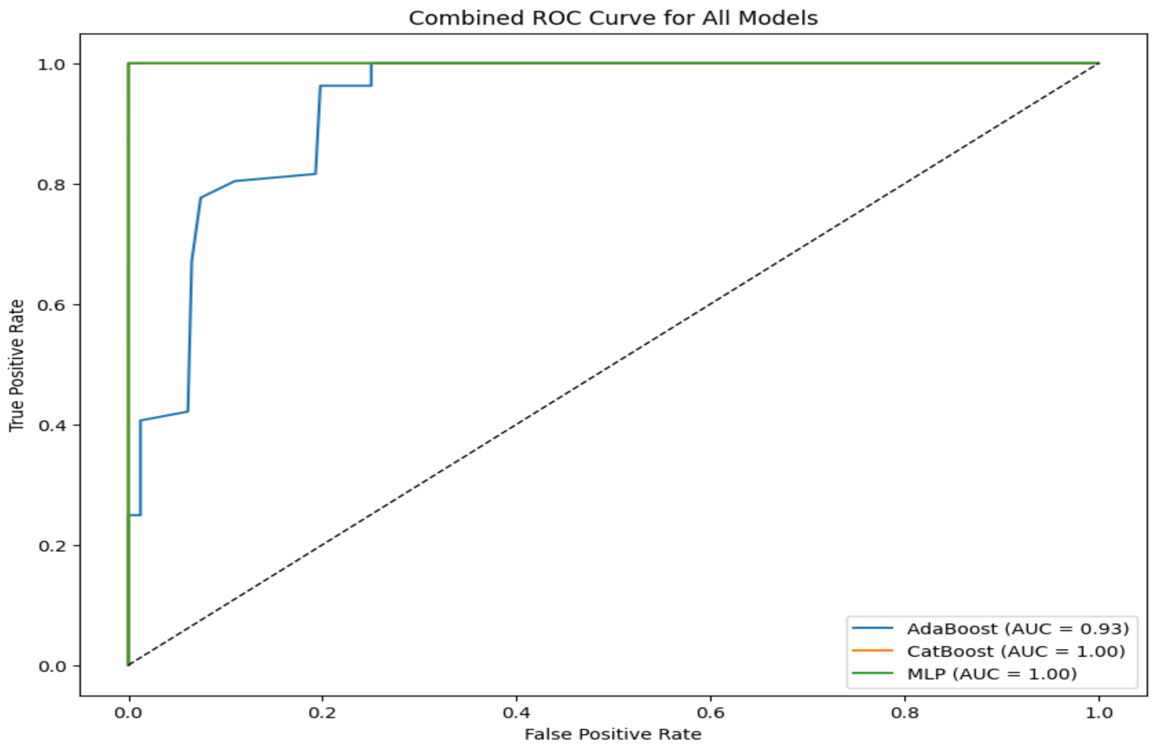

Figure 4.36: Comparison of Proposed ML Models for Delivery Status Prediction

Figure 4.36 presents a comparative analysis of the AdaBoost, Cat Boost, and MLP models for delivery status prediction using ROC curves. The plot overlays the ROC curves for each model, with corresponding AUC scores indicating their overall performance. Cat Boost and MLP both achieve perfect AUC scores of 1.00, demonstrating superior discriminative ability compared to AdaBoost, which has an AUC of 0.93. This visualization allows for a direct comparison of the models' ability to distinguish between positive and negative delivery status outcomes across various classification thresholds.

Findings for Late Risk Prediction

This section focuses on predicting the risk of late deliveries using classification models. The analysis highlights the model that best identifies potential delays, helping in proactive decision-making to minimize disruptions.

Analysis of Proposed AdaBoost Model

Table 4.4: Performance Metrics of AdaBoost Model in Late Risk Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
AdaBoost	97.52	95.68	100	97.79

The above-mentioned Table 4.4 presents the performance metrics of the AdaBoost model specifically for late risk prediction. The table shows the accuracy achieved by the model is 97.52%, the precision is 95.68%, the recall is a perfect 100%, and the resulting F1score is 97.79%. These metrics collectively indicate a strong performance of the AdaBoost model in identifying instances of late risk.

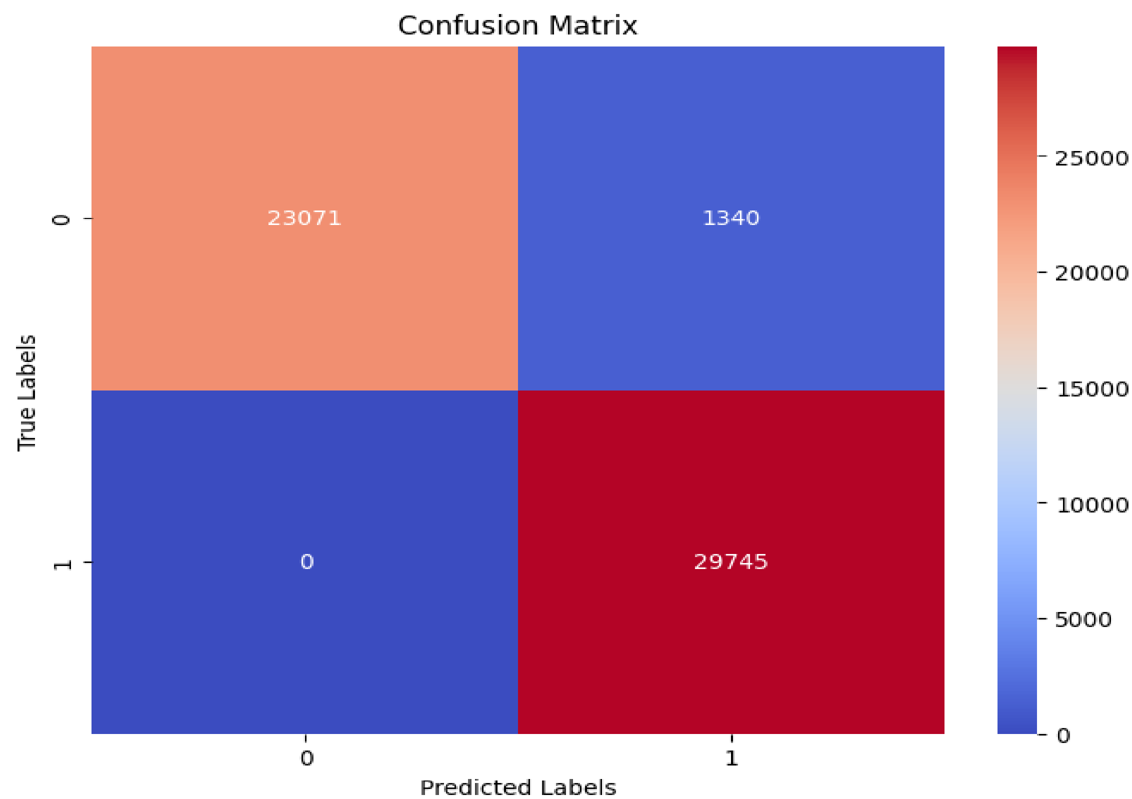

Figure 4.37: Confusion Matrix of AdaBoost Model for Late Risk Prediction

Figure 4.37, the Confusion Matrix for the AdaBoost model in late risk prediction, shows the correctly classified instances. The model accurately predicted 23071 instances of the non-late risk class (0) and 29745 instances of the late risk class (1). These diagonal values represent the number of TN and TP achieved by the AdaBoost model for this prediction task.

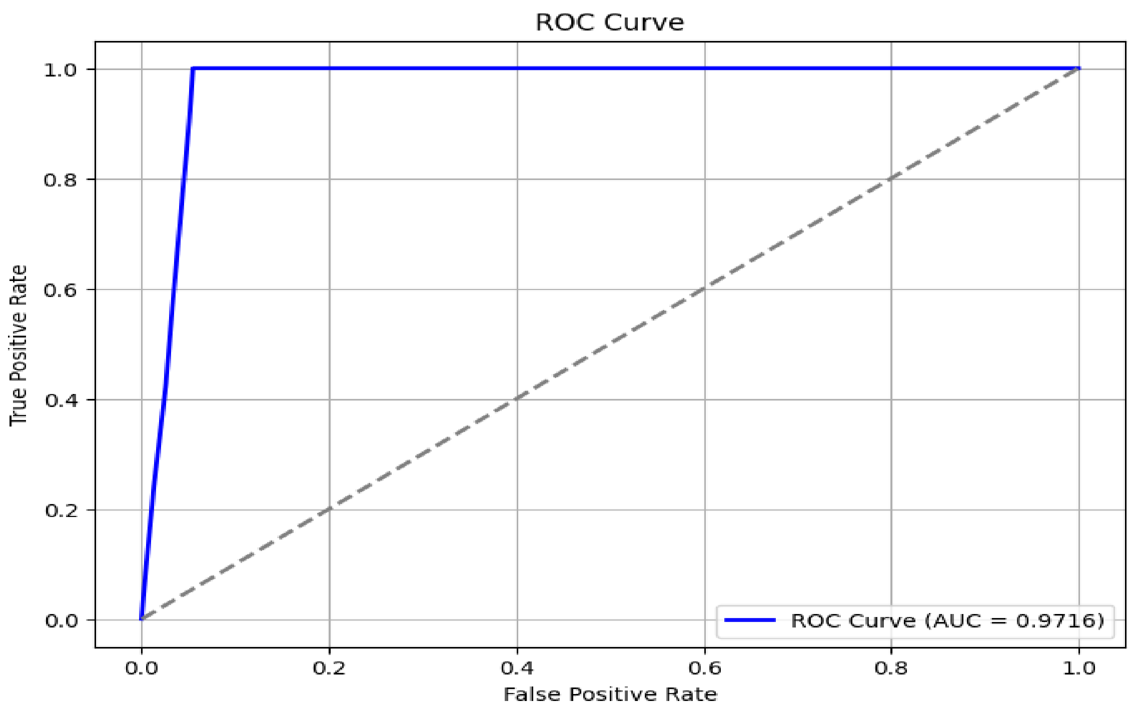

Figure 4.38: ROC Curve of AdaBoost Model for Late Risk Prediction

Figure 4.38 displayed the ROC curve for the AdaBoost model in late risk prediction. The curve plots the TPR against the FPR, yielding an AUC of 0.9716. The AdaBoost model's high AUC shows that it can accurately distinguish between late-risk and non-late-risk occurrences.

Analysis of Proposed Cat Boost Model

Table 4.5: Performance Metrics of Cat Boost Model in Late Risk Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**

Cat Boost	97.52	95.68	100	97.79

Table 4.5, which was stated before, showed several metrics that measured how well the Cat Boost model predicted late-stage risks. The table indicates that the model achieved an accuracy of 97.52%, a precision of 95.68%, a recall of 100%, and an F1score of 97.79%. These metrics suggest a strong performance by the Cat Boost model in identifying instances of late risk, with perfect recall indicating that all actual late risk cases were correctly identified.

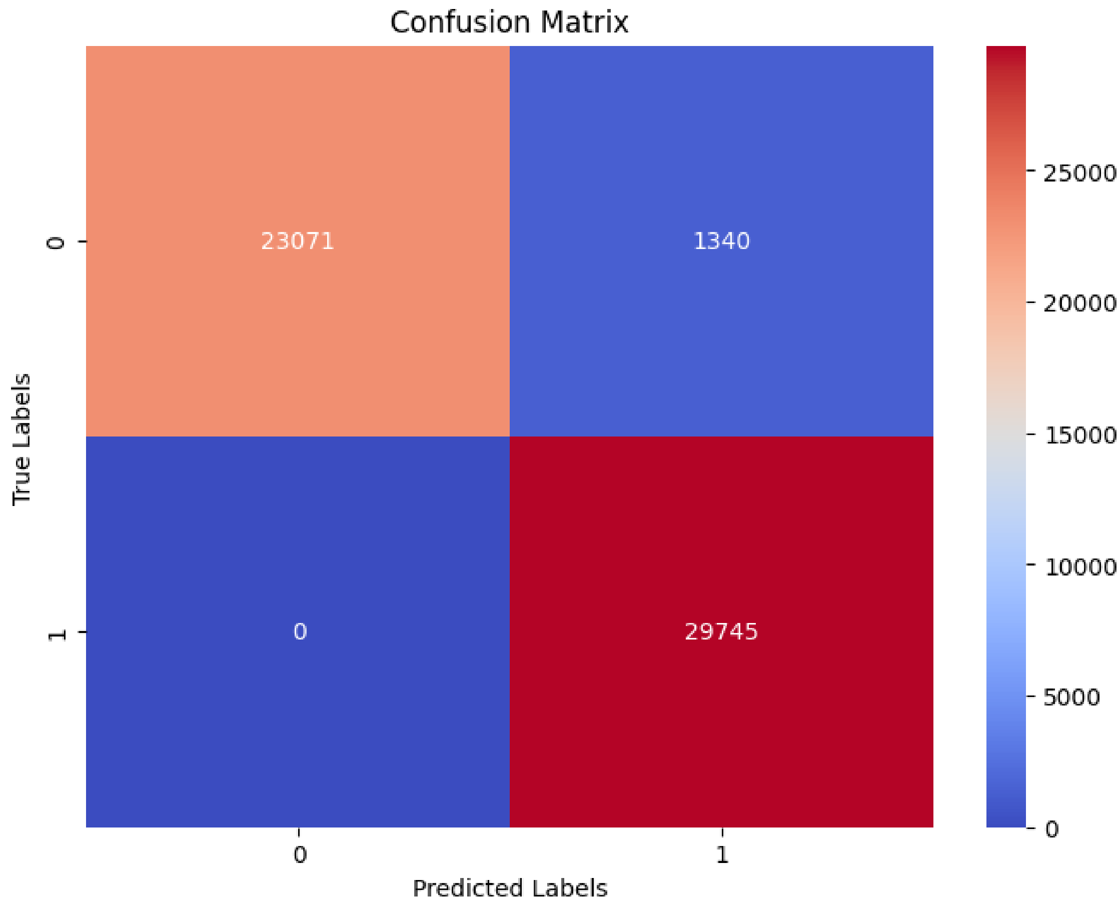

Figure 4.39: Confusion Matrix of Cat Boost Model for Late Risk Prediction

Figure 4.39 displays the confusion matrix for the Cat Boost model's performance in late risk prediction. The model correctly classified 23071 instances as non-late risk (0) and

29745 instances as late risk (1). These values on the diagonal represent the TN and TP achieved by the Cat Boost model.

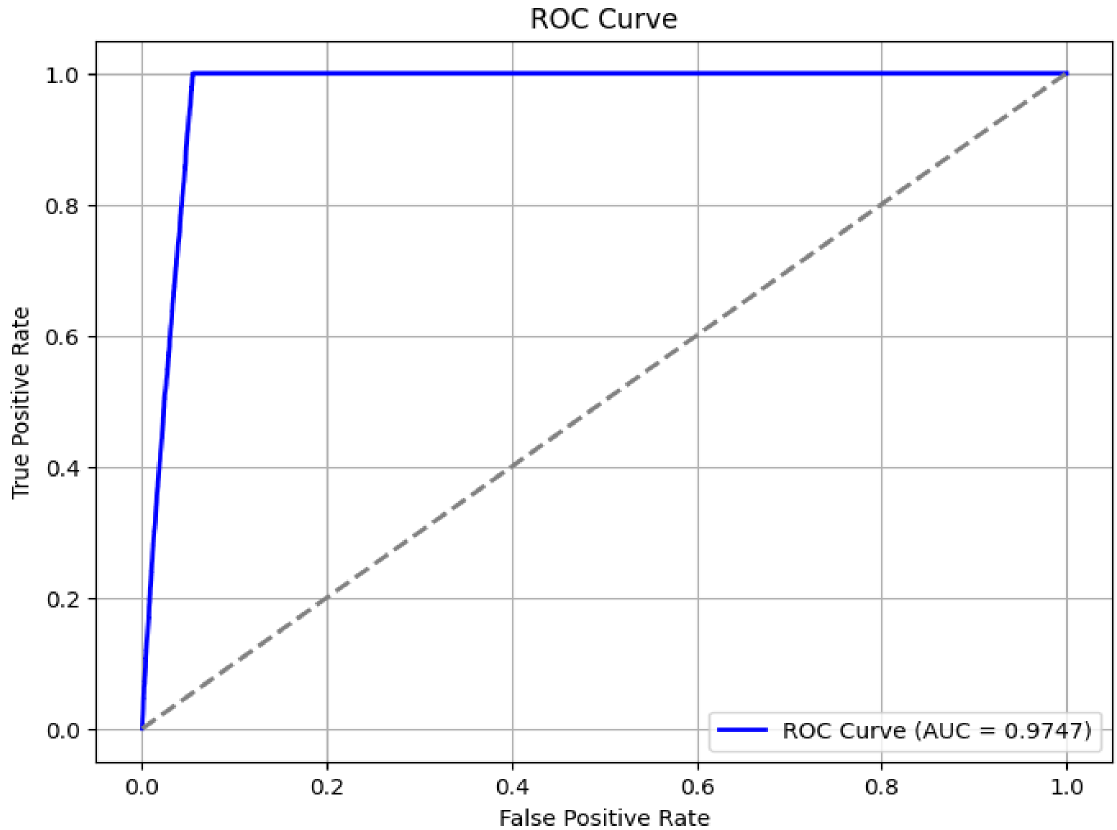

Figure 4.40: ROC Curve of CatBoost Model for Late Risk Prediction

Figure 4.40 displays the ROC curve for the Cat Boost Model in predicting late risk. The model achieves a high AUC of 0.9747, indicating strong discriminatory power between late and non-late risk instances.

Analysis of Proposed MLP Model

Table 4.6: Performance Metrics of MLP Model in Late Risk Prediction

Model	Accuracy	Precision	Recall	F1-score
MLP	97.52	95.68	100	97.79

Table 4.6 displays the MLP model's performance metrics for late risk prediction, as described before. The table details the model's achieved accuracy of 97.52%, a precision

of 95.68%, a recall of 100%, and a resulting F1score of 97.79%. These metrics thoroughly assess how well the MLP model detects and categorizes incidents linked to late risk.

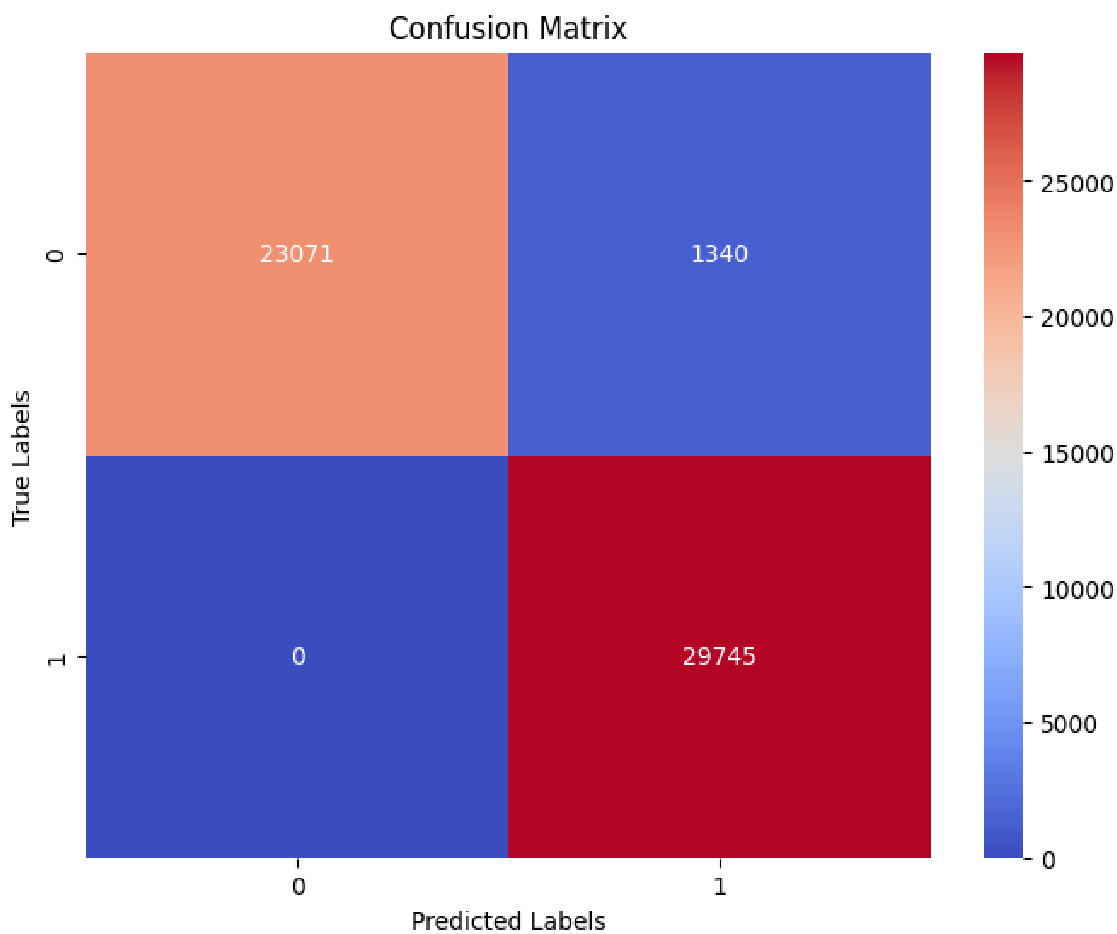

Figure 4.41: Confusion Matrix of MLP Model for Late Risk Prediction

Figure 4.41 displayed the Confusion Matrix for the MLP model's late risk prediction. The correctly classified instances are 23071 for the non-late risk class (0) and 29745 for the late risk class (1). These diagonal values represent the true negatives and true positives achieved by the MLP model.

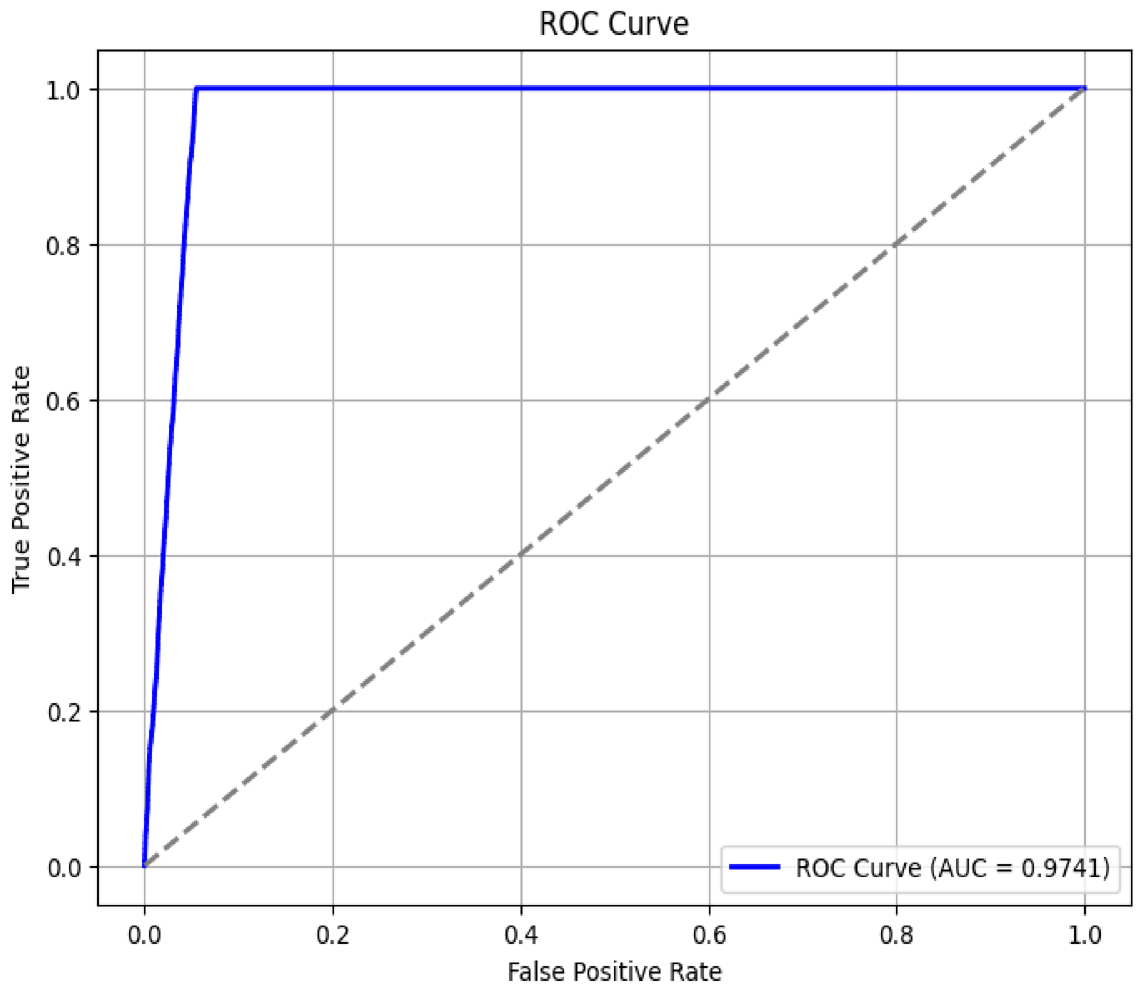

Figure 4.42: ROC Curve of MLP Model for Late Risk Prediction

Figure 4.42 presents the ROC curve for the MLP model's performance in late risk prediction. The MLP model strongly differentiates between late and non-late risk cases, as seen by the high AUC of 0.9741 on the curve.

Comparative Analysis of Proposed ML Models for Late Risk Prediction

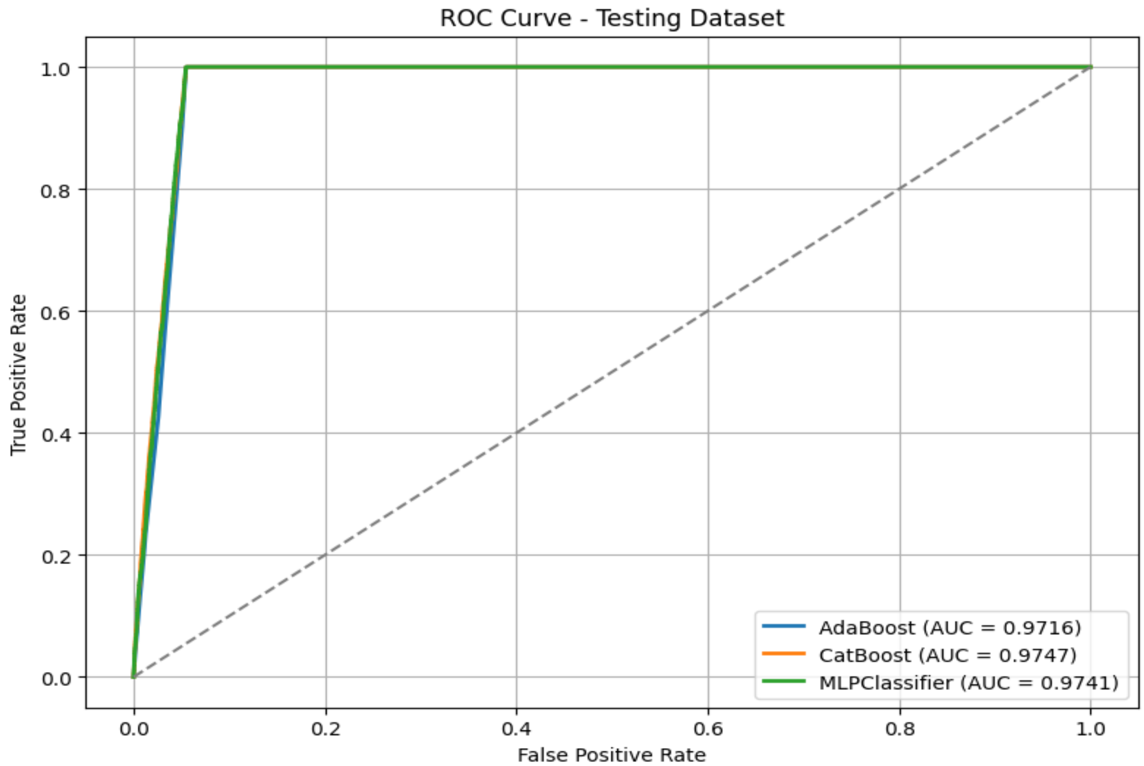

Figure 4.43: Comparative Analysis of Proposed ML Models for Late Risk Prediction

Figure 4.43 presents a comparative ROC curve analysis of AdaBoost, Cat Boost, and MLP Classifier models for late risk prediction on the testing dataset. The plot shows the TPR against the FPR for each model, with corresponding AUC values. Cat Boost achieves the highest AUC of 0.9747, closely followed by MLP Classifier at 0.9741, while AdaBoost has a slightly lower AUC of 0.9716, indicating that all three models demonstrate strong and comparable performance in discriminating between late and non-late risk instances.

Findings for Demand Forecasting

This subsection evaluates the demand forecasting model using MAE, MSE, and RMSE. The consequence identifies the reliability of demand forecasts, which helps to enhance inventory management and SCM.

Analysis of Proposed Prophet Regression Model

Table 4.7: Performance Metrics of Prophet Regression Model for Demand Forecasting

Model	**MAE**	**MSE**	**RMSE**
Prophet Regression	0.1391	0.0260	0.1612

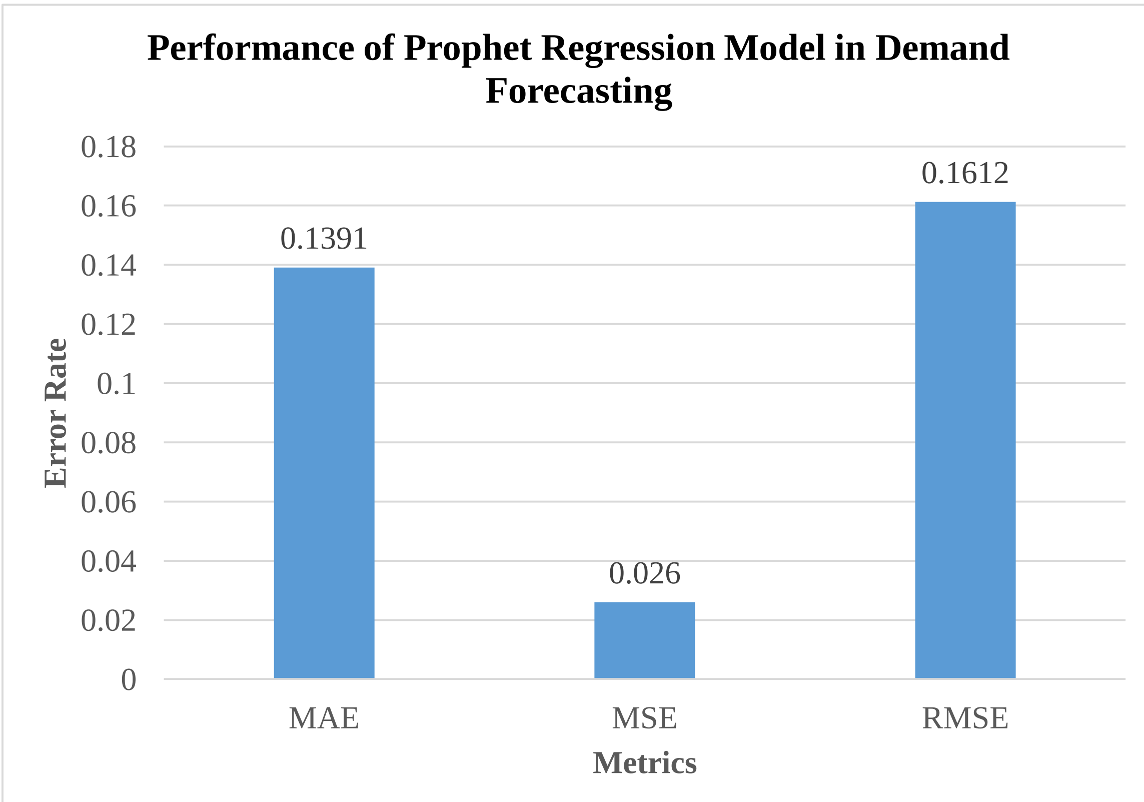

Figure 4.44: Performance of Prophet Regression Model in Demand Forecasting

A above-mentioned Table 4.7 and Figure 4.44 displays a bar chart illustrating a performance of the Prophet regression model in demand forecasting on the DataCo SMART SUPPLY CHAIN dataset. The model shows three primary metrics with values of MAE at 0.1391 while MSE equals 0.026 and RMSE stands at 0.1612. The chart demonstrates both accuracy and prediction capabilities of the model by presenting error rate statistics for demand forecasting results.

Analysis of Weekly Orders Forecast

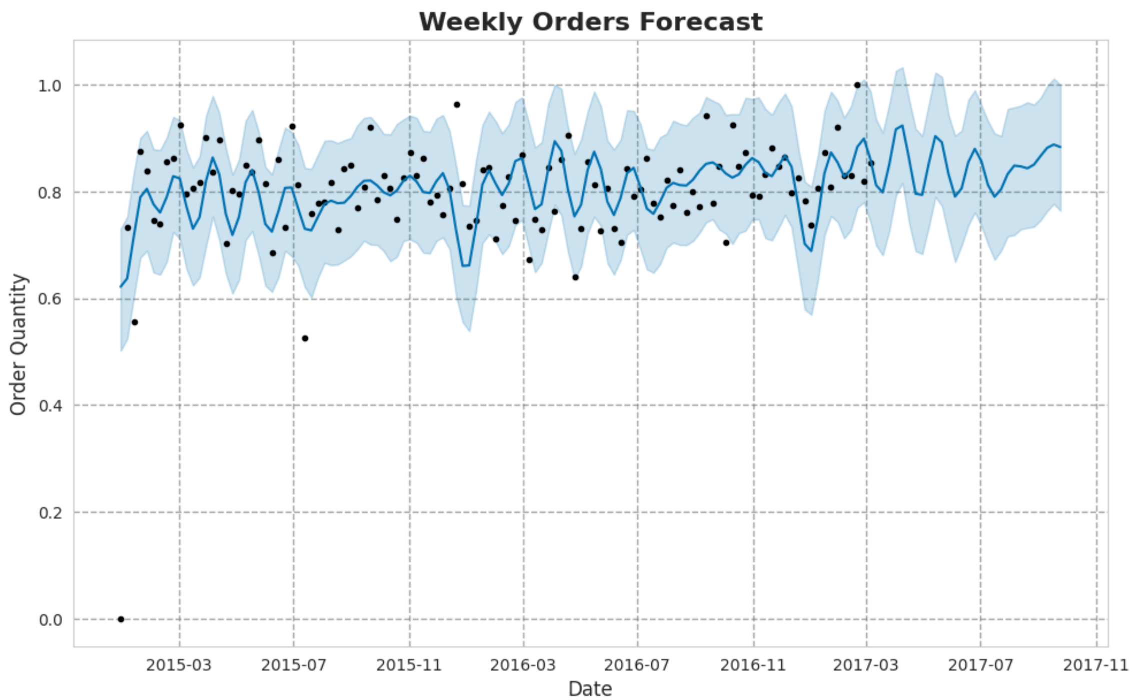

Figure 4.45: Prophet Regression Model Weekly Orders Forecast

The forecast of weekly orders through Prophet regression appears in Figure 4.45 where historical and projected trends in order volumes across time are displayed. The analysis uses a date scale on the x-axis and a normalized order quantity scale reaching 1 across the y-axis as its axes. The graph contains actual black dots to display order volume variations through weekly observations. A solid blue line presents the model estimation that includes seasonal patterns and long-term order quantity movement. The uncertainty interval generates confidence ranges using the light blue shaded section of the chart. Most of the forecasting period shows order counts rising steadily between 0.6 and 1.0 according to the model. Numerous distinct seasonal patterns appear throughout the data because it highlights cyclic changes in demand levels. From mid-2015 until the end of 2017 the model reveals rising order quantities that later experience periodic decreases yet recover to previous levels. The overall upward direction of weekly order trends remains consistent because the data shows periodic upward movements.

Analysis of Forecast Components

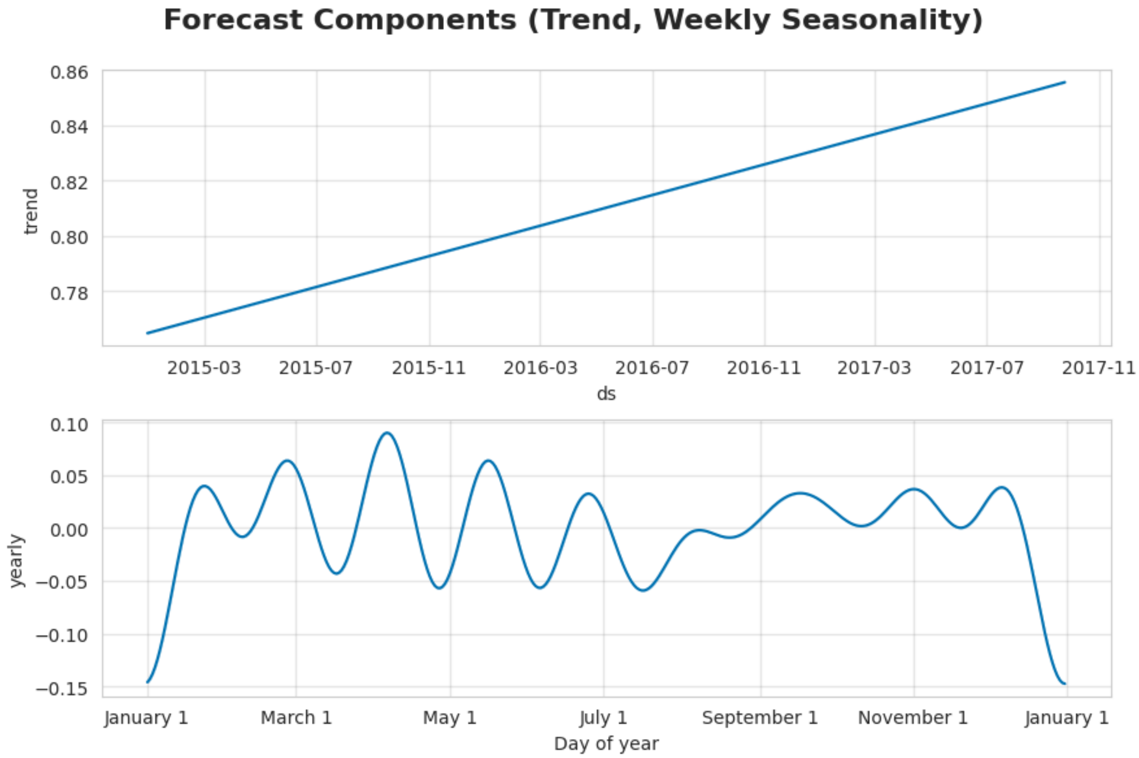

Figure 4.46: Forecast Components (Trend, Weekly Seasonality)

Figure 4.46 presents the forecast components of the Prophet model, decomposing the predictions into trend and weekly seasonality components for demand forecasting. The trend component shows a continuous upward movement from 0.77 during early 2015 to 0.86 in mid-2017 based on the data displayed in the top graph. The bottom graph shows how demand waves up and down throughout the weeks of each year. Periodic peaks achieve a maximum level of 0.10 demand during regular intervals while troughs show nearly -0.15 demand at the yearly beginning and end. This decomposition helps in understanding the cyclical trends and fluctuations for long term and looking into the cyclicality of the demand patterns which helps in demand planning and resource forecasting.

Analysis for Actual vs Predicted Weekly Orders

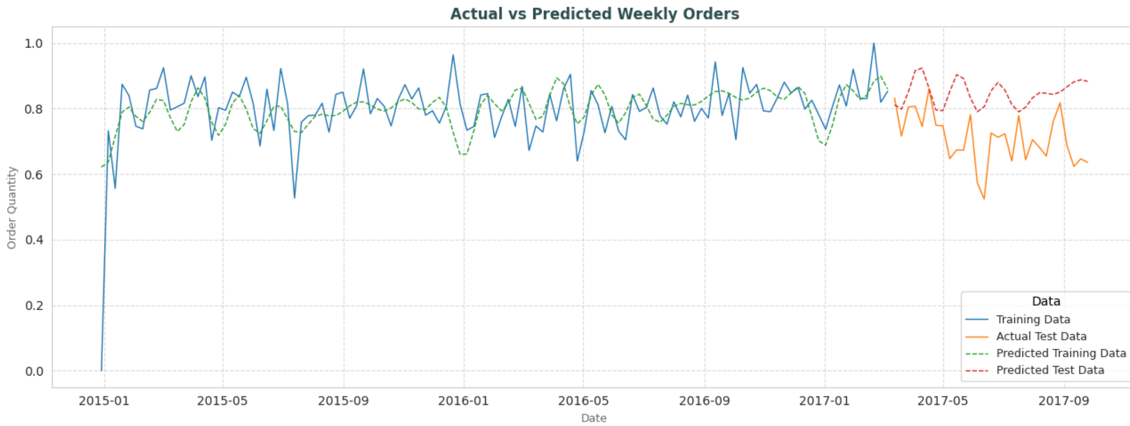

Figure 4.47: Actual vs Predicted Weekly Orders

Figure 4.47 illustrates the performance of the Prophet model in forecasting weekly orders by comparing actual versus predicted values. The blue line represents the actual training data and is shown in the range of the order quantity of approximately 0.6 up to 0.9 for the period Q1 2015 up to Q4 2017. Based on the actual test data indicated by a solid orange line, the value oscillates between 0.6 and 0.9 from 2017-01 up to 2017-09, with sharp decline in 2017-06. The green dashed line in the figure represents the predicted training data and it can be observed that it lies very close to the actual training data. This is as shown in the red dotted line that representing the predicted test data closely follows that of the actual test data, ranging between 0.7 and 0.9 with some small deviations particularly at the point of the dip in the actual test data to show that it is capable of predicting the flow of even though it does not accurately predict the exact values.

CHAPTER V:

ANALYSIS AND DISCUSSION

5.1 Comparative Analysis

The analysis examines various machine learning methods which optimize supply chain operations through the evaluation of delivery status prediction and late risk prediction alongside demand forecasting. This research evaluates the base and proposed models using the DataCo Smart Supply Chain for Big Data Analysis dataset to determine their success in forecasting accuracy and decision support. Comparing base models with machine learning proposed techniques, the study pointed how much the efficient has been increased in supply chain operations. The results suggest that analytical methods enhance decision making, increase effectiveness, decrease risks, and facilitate the planning and management of the supply chain activities.

Comparison of Base and Proposed Models for Delivery Status Prediction

Table 5.1: Comparative Analysis of Base and Proposed Model for Delivery Status Prediction

Parameters	**Base Model**			**Proposed Model**		
	Decision Tree (Pattnaik et al., 2024)	**KNN** (Pattnaik et al., 2024)	**Naïve Bayes**	**AdaBoost** (Pattnaik et al., 2024)	**Cat Boost**	**MLP**
AUC	97.00	70.00	96.00	93.00	100.00	100.00

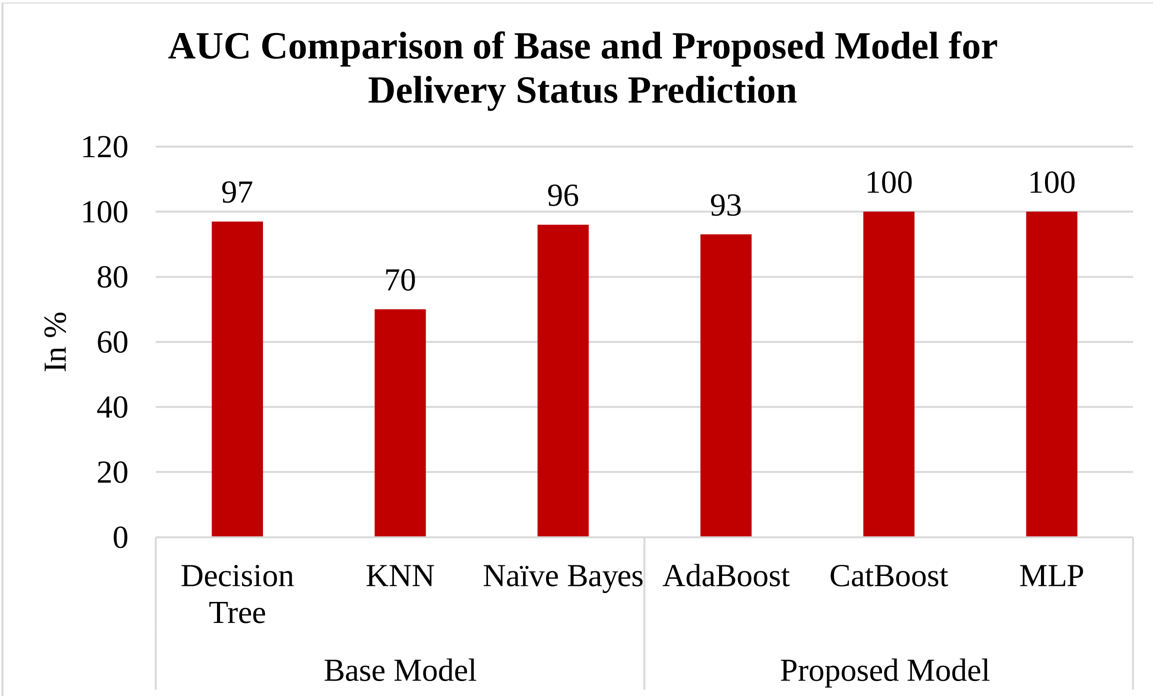

Figure 5.1: AUC Comparison of Base vs. Proposed Models for Delivery Status Prediction

The above-mentioned Figure 5.1 and Table 5.1 show a comparative analysis displaying an AUC performance of different machine learning models for delivery status prediction using the DataCo Smart Supply Chain for Big Data Analysis dataset. The different machine learning models which include base models (Decision Tree, KNN, and Naïve Bayes), AdaBoost, Cat Boost, and MLP are grouped along the x-axis. The y-axis displays the AUC score as a percentage. Among the base models, Decision Tree achieves an AUC of 97%, Naïve Bayes scores 96%, while KNN performs the lowest at 70%. AdaBoost outperforms the other models to achieve 93% while Cat Boost and MLP reach 100% AUC score of 100%. The research outcomes demonstrate how ensemble learning along with deep learning achieves better delivery status prediction accuracy, thus providing valuable potential benefits to supply chain optimization.

Comparison of Base and Proposed Models for Late Risk Prediction

Table 5.2: Comparative Analysis of Base and Proposed Models for Late Risk Prediction

Parameters	**Base Model**			**Proposed Model**		
	XGBoost (Pattnaik et al., 2024)	**LR** (Pattnaik et al., 2024)	**DT** (Pattnaik et al., 2024)	**AdaBoost**	**Cat Boost**	**MLP**
AUC	95.00	58.00	77.00	97.16	97.47	97.41

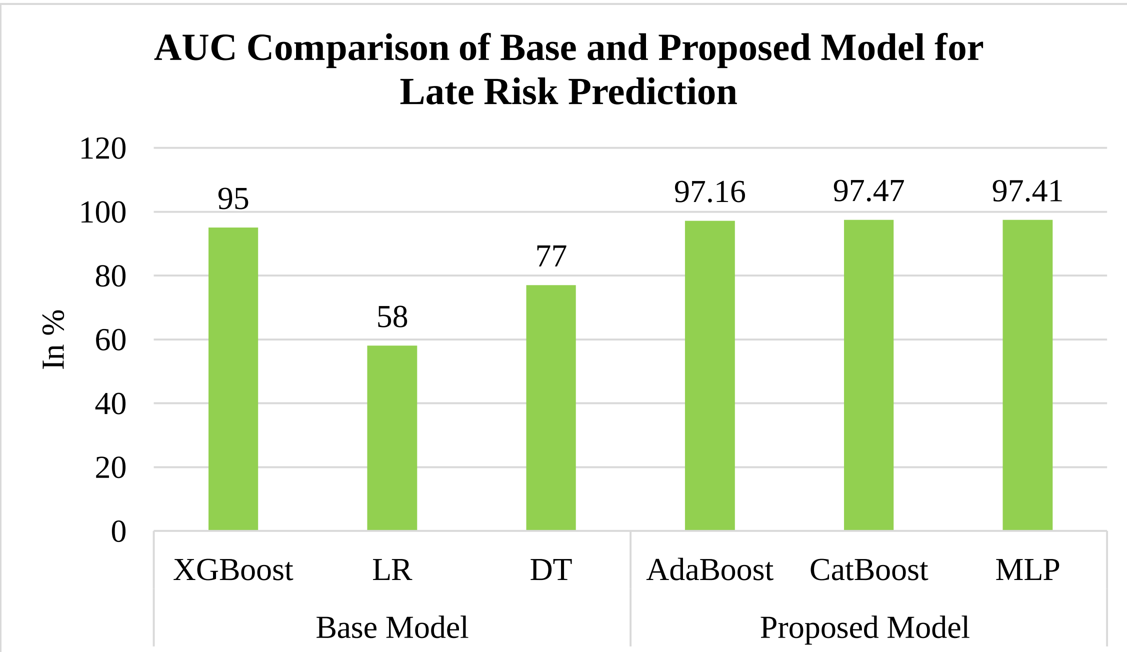

Figure 5.2: AUC Comparison of Base vs. Proposed Models for Late Risk Prediction

Table 5.2 and Figure 5.2 show the results of comparing the base and suggested models' performance for late risk forecasting using the AUC measure. The image provides a bar graph with the base and recommended models shown on the x-axis and their performance as a percentage on the y-axis. The analysis includes base models XGBoost, Logistic Regression (LR) and Decision Tree (DT) where XGBoost demonstrated the best result (95.00% AUC) followed by Decision Tree (77.00% AUC) yet Logistic Regression performed at (58.00% AUC). The proposed AdaBoost and Cat Boost and MLP models outperformed other options for predictions because it produced AUC scores of 97.16% and 97.47% and 97.41%, respectively. The better results from proposed models demonstrate ensemble learning and deep learning have shown their power to optimize late-risk prediction abilities. The experimental findings indicate boosting algorithms along with neural networks provide superior accuracy and reliability when used for predicting late risks within supply chain management compared to traditional prediction models.

Comparison of Base and Proposed Models for Demand Forecasting

Table 5.3: Comparative Analysis of Base and Proposed Models for Late Risk Prediction

Parameters	**Base Model**			**Proposed Model**
	Lasso Regression (Haque & Kelly, n.d.)	**Random Forest** (Haque & Kelly, n.d.)	**Linear Regression** (Submitted et al., 2024)	**Prophet**
MAE	1.3374	0.2014	0.4587	0.1391
RMSE	2.0951	1.7779	0.6527	0.1612

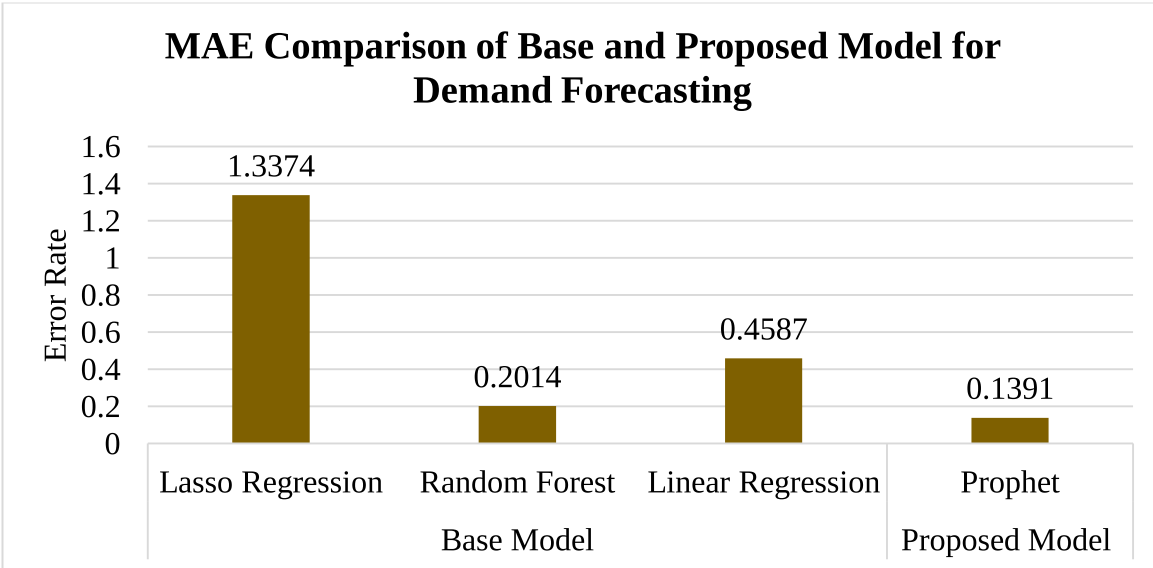

Figure 5.3: MAE Comparison of Base vs. Proposed Models for Demand Forecasting

Table 5.3 and Figure 5.3 provide a bar chart that compares the MAE of four alternative demand forecasting models using the DataCo Smart Supply Chain for Big Data Analysis dataset. The chart shows the MAE values for Lasso Regression, Random Forest, and Linear Regression as base models, and Prophet as the proposed model. The MAE values are displayed above each bar, indicating that Prophet has the lowest error rate (0.1391), followed by Random Forest (0.2014), Linear Regression (0.4587), and Lasso Regression with the highest error rate of 1.3374. The Prophet model produces superior results than base models when used for forecasting demand within this dataset.

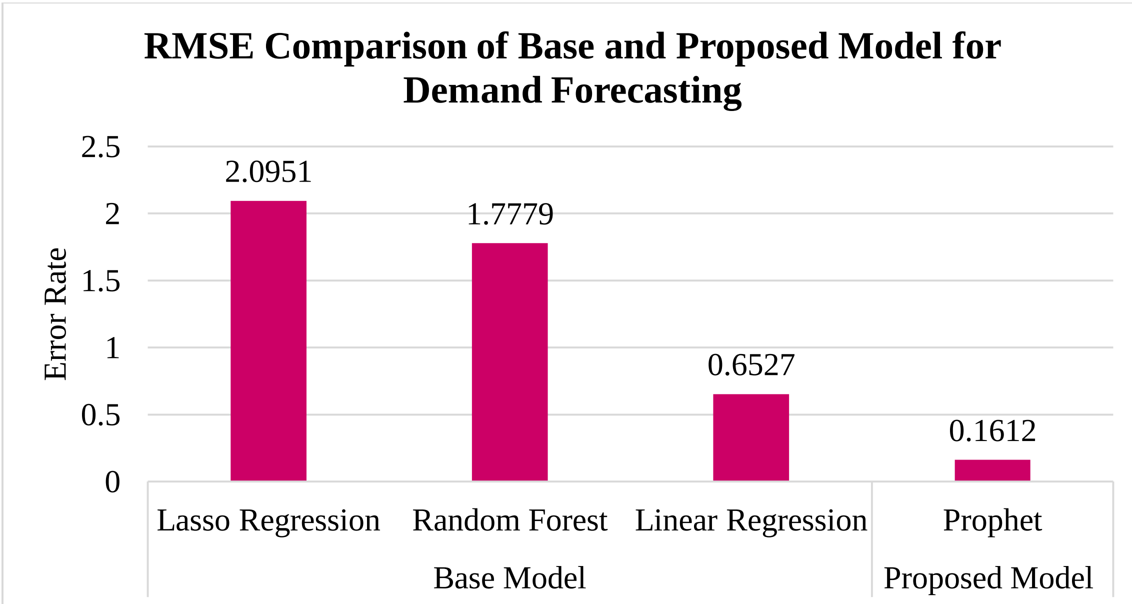

Figure 5.4: RMSE Comparison of Base vs. Proposed Models for Demand Forecasting

An above-mentioned Table 5.3 and Figure 5.4 presents a bar chart comparing the RMSE of four various models used for demand forecasting on the DataCo Smart Supply Chain for Big Data Analysis dataset. The chart displays the RMSE values for Lasso Regression, Random Forest, and Linear Regression as base models, and Prophet as the proposed model. The RMSE values are shown above each bar, indicating that Prophet has the lowest error rate (0.1612), followed by Linear Regression (0.6527), Random Forest (1.7779), and Lasso Regression with the highest error rate of 2.0951. The Prophet model leads other tested base models in prediction accuracy regarding demand forecasts because it demonstrates the least RMSE value.

5.2 Discussion

The comparative study of supply chain model using the ML approach reveals that organization can benefit from the latest ensemble learning type and deep learning algorithms compared to the traditional models. Finally, this section presents the chief research findings and relates these to supply chain management.

Delivery Status Prediction

The findings reveal that the new models featured in this study including AdaBoost, Cat Boost, and MLP are more accurate in predicting delivery status higher when compared to Decision Tree, Naïve Bayes, and KNN models. The fact that Cat Boost bagged 100% AUC score means that it is able to identify intricate patterns within delivery data as well as the MLP model. However, this high performance is deceptive in a way because it might mean overfitting especially if the dataset is not diverse. Future research could explore methods such as cross-validation and additional regularization techniques to ensure generalizability across different supply chain scenarios.

Late Risk Prediction

The analysis demonstrates that ensemble learning techniques (AdaBoost and Cat Boost) and deep learning models (MLP) improve the accuracy of late risk prediction, with AUC scores surpassing traditional models. The highest-performing model, CatBoost (97.47%), suggests that gradient boosting techniques effectively handle imbalanced datasets and capture nonlinear relationships in supply chain delays. However, real-world supply chains involve dynamic and evolving conditions, which might not be fully captured in historical data. Future improvements could integrate real-time data streams and reinforcement learning techniques to enhance predictive capabilities.

Demand Forecasting

The Prophet model consistently outperformed traditional forecasting methods such as Linear regression, lasso regression, and random forest, as indicated by lower MAE and RMSE scores. This suggests that Prophet's ability to model seasonal trends and handle missing data makes it a robust choice for demand forecasting. However, one limitation of Prophet is its assumption of trend continuity, which may not always hold in highly volatile markets. Further research could examine hybrid forecasting models that combine Prophet with neural networks or Bayesian methods to improve adaptability.

Implications for Supply Chain Management

The findings underscore the importance of leveraging machine learning for supply chain optimization. By improving delivery status prediction, reducing late risks, and enhancing demand forecasting, businesses can minimize operational inefficiencies, reduce costs, and improve customer satisfaction. However, the successful implementation of these models requires:

- High-quality, real-time data to improve model accuracy and adaptability.
- Integration with existing supply chain systems to enable seamless decision-making.
- Scalability and robustness testing to ensure models perform well under diverse market conditions.

CHAPTER VI:
SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

The complexities in supply chains also the increased globalization of supply chain management has made it imperative for organizations to look for better, informed methods to control or even reduce the challenges of operational inefficiency, lead time, among other factors. Over the past few years, the application of ML in combination with

Big Data Analytics has been identified to address such complexities. This paper also aims to determine the use of ML models in improving the crucial decision-making areas of the supply chain by classifying the delivery status, predicting the possible risks of delayed delivery, as well as in demand prediction. In the case of DataCo SMART SUPPLY CHAIN, the examined models are AdaBoost, Cat Boost, MLP, and Prophet regression.

Classification performance by Cat Boost for delivery status prediction was exceptional, achieving 100% accuracy, precision, recall, and F1-score. Slight overfitting was observed by using the MLP model closely followed by the desired accuracy of 99.98% for all the classification metrics. The AdaBoost model, though not as strong, achieved good performance with 77.64% for accuracy, 77.80% for precision, 77.64% for recall and an F1-score of 76.99% meaning moderate level of predictive ability for this particular task.

The credit scoring results of the three models comparing the late delivery risk classification with the hold-out set were as follows, for all three models, that is AdaBoost, Cat Boost, and MLP, achieved an accuracy level of 97.52%, precision of 95.68 % a recall value of 100% and the f1 score proved to have value of 97.79%. The high recall value is a notable one since it signifies the ability of the models to detect all the instances of the risk of late delivery, which is essential for immediate management in supply chain management.

For demand forecasting the Prophet regression model was used and for assessment of the model's performance regression metrics were used. The model achieved a MAE of 0.1391, MSE of 0.026, and RMSE of 0.1612. The model demonstrates excellent performance in predicting product demand because its error values remain low which leads to improved planning and better inventory control. Overall, this study confirms the powerful role of ML in supply chain management.

Classification tasks were the most optimal ones for Cat Boost and MLP, whereas Prophet demonstrated high effectiveness in time-series analysis. It is possible to emphasize that application of such models can contribute to development of better decision-making, decrease of uncertainty, and improving supply chain flexibility in contemporary networks.

6.2 Implications

The findings of this research outline the role of machine learning as a groundbreaking tool in contemporary supply chain management in the areas such as delivery status prediction, identification of the late risk, and demand forecast. Advanced machine learning techniques represented by Cat Boost and MLP together with AdaBoost and Prophet demonstrate superior performance compared to traditional Decision Trees, Naïve Bayes, and Linear Regression in both accuracy measurement and reliability statistical analysis as well as error control metrics. The developed performance improvements provide both technical benefits and real-world functional advantages. Market managers exploit accurate delivery status along with late risk predictions to actively adjust shipping routes while delivering resources for more efficient operations which leads to delay reductions and better customer satisfaction. Accurate demand forecasting enables better control of inventory combined with cost-effective procurement by reducing situations of both excessive and insufficient stock levels. The implementation success of these models allows organizations to maintain competitive positions through their ability to make quick data-based choices in real time. The study supports operational systems by integrating ensemble learning and deep learning techniques because this integration allows automation and error reduction that improves supply chain agility. These implications have high importance today because rapid markets with ongoing supply chain disruptions and volatility require them. The research supports organizations in adopting machine learning technology at all business scales for deploying intelligent optimization solutions to their supply chain. The study proves that organizations can achieve substantial improvements in predictive potential alongside operational excellence and market resilience by deploying data effectively through advanced algorithms in data-driven global markets.

6.3 Recommendations for Future Research

The findings of this work create potential for future study that aims to develop advanced methods toward superior machine learning models for supply chain management prediction capabilities. Historical and spatial relationships within extensive supply chain data can be detected through deep learning methods particularly LSTM and CNN-based approaches when performing tasks including demand forecasting. The creation of hybrid models which unite ensemble learning methods with neural networks and traditional algorithms will produce more strong and adaptive solution systems. Sophisticated preprocessing methods which include feature selection and dimensionality reduction with outlier detection and time series decomposition enhance model execution and textual understanding. The development of flexible models must prioritize scalability together with interpretability features which must match real-time conditions. Additional research aims to develop automatic model operation together with methodologies to implement multicompany use and integrate privacy and ethical guidelines.

Key directions for future research include:

- **Application of Deep Learning Models:** Future research can benefit from employing deep learning architectures like LSTM, GRU, CNN, and Transformer models to better capture temporal patterns, sequential dependencies, and spatial features in supply chain data. These models are particularly effective in demand forecasting and anomaly detection due to their ability to learn complex, non-linear relationships.

- **Development of Hybrid Models:** Combining traditional algorithms with ensemble methods and neural networks can lead to more accurate and robust models. Hybrid approaches allow leveraging the strengths of multiple techniques, thereby improving prediction reliability and adapting to various types of supply chain challenges.
- **Use of Advanced Data Preprocessing Techniques:** The implementation of feature engineering combined with PCA and SMOTE as well as time series decomposition and outlier detection methods create improvements in both data quality and model performance. The preprocessing methods clean the input data while maintaining balance and extracting maximum value from its informative features.
- **Exploration of Explainable AI (XAI):** Research efforts should concentrate on developing explainable AI techniques to build transparency in model decision making processes. Explainable Artificial Intelligence enables supply chain managers to understand model decisions thus building trust in AI systems therefore facilitating operational adoption.
- **Integration into Real-Time Systems:** The implementation of predictive models should happen directly inside enterprise systems such as ERP or WMS so users can access immediate insight. Such integration enables proactive business decisions through alerts which automatically execute both logistical operations and inventory tasks.

- **Focus on Cross-Domain Validation:** Multiple industry testing of models provides both generalizable results and strong model stability. To assess scalability and adaptability supply chain models must be deployed into healthcare fields and manufacturing together with retail.

6.4 Conclusion

In conclusion, this research aimed at revealing how machine learning and big data analytics may contribute to the improvement of the supply chain, with special reference to delivery status prediction, late risk factors, and demand forecasting. In this research, the DataCo Smart Supply Chain for Big Data Analysis data set was used in conducting the analysis and training to compare the models like baseline, AdaBoost, Cat Boost, multilayer perceptron and Prophet. The experiments showed that all the proposed models were better than the base models in all tasks. In delivery status prediction, it is observable that Cat Boost and MLP yielded a 100% accuracy which demonstrates their reliability in the classification. All the models proposed in the late risk prediction domain which include the AdaBoost, Cat Boost, and MLP performed very well with the AUC score above 97% high precision, and perfect recall with a score of (100%) which underpin the strategic importance of early detection of delivery delays through timely risk prediction. The Prophet model was found to give the lowest MAE (0.1391) and RMSE (0.1612) among the base models which highlighted its effectiveness in demand forecasting especially for time-series data and in inventory planning and resource allocation. These observations imply that incorporating superior predictive models into supply chain processes does more than improving precision and performance but also facilitating wise decision-making and system longevity. The implications derived from these models are that the organizations could be in a better position to forecast demand fluctuations, manage risks in real time and improve the efficiency of logistics. In sum, the best supply chain is intelligent, responsive and resourceful and this research confirms the importance of AI and big data as strategic tools in this area. Future work should focus on extending the current model into real-time systems; incorporating external sources of data in the model; and testing the model applied to different industries to achieve the similar positive results stated in this work.

REFERENCES

Abbas, K., Afaq, M., Khan, T. A., & Song, W. C. (2020). A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. *Electronics (Switzerland)*. https://doi.org/10.3390/electronics9050852

Ahmed, A., Dabral, S., Bahuguna, D., Kaur, J., & Singh, B. (2023). Artificial Intelligence’s Integration in Supply Chain Management: A Comprehensive Review. *Issn*, *13*(3), 2323–5233.

Ahmed, S. S., Akter, T., & Ma, Y. (2018). Green Supply Chain Management (GSCM) Performance Implemented by the Textile Industry of Gazipur District, Dhaka. *Logistics*. https://doi.org/10.3390/logistics2040021

Ali, S. I., Ali, A., AlKilabi, M., & Christie, M. (2021). Optimal supply chain design with product family: A cloud-based framework with real-time data consideration. *Computers and Operations Research*. https://doi.org/10.1016/j.cor.2020.105112

ANANTHARAMAN PRAKASH, C.B. SENTHIL KUMAR, D. K. J. (2025). *THE IMPACT OF ARTIFICIAL INTELLIGENCE (AI) ON SUPPLY CHAIN DECISION-MAKING: A COMPREHENSIVE ANALYSIS*.

Armstrong, M., Lazio, F., Herrmann, D., & Duckworth, D. (2018). Capturing value from the smart packaging revolution. *Deloitte Insights*, 22.

Ascher, D., Dubois, P. F., Hinsen, K., Hugunin, J., & T, O. (2001). *Numerical Python*.

Ashraf, M., Eltawil, A., & Ali, I. (2024). Disruption detection for a cognitive digital supply chain twin using hybrid deep learning. In *Operational Research* (Vol. 24, Issue 2). Springer Berlin Heidelberg. https://doi.org/10.1007/s12351-024-00831-y

Asim, Z., Shamsi, I. R. Al, Wahaj, M., Raza, A., Abul Hasan, S., Siddiqui, S. A., Aladresi, A., Sorooshian, S., & Seng Teck, T. (2022). Significance of Sustainable Packaging: A Case-Study from a Supply Chain Perspective. *Applied System Innovation*. https://doi.org/10.3390/asi5060117

Awasthi, S. (2024). Artificial Intelligence in Supply Chain Management. *Journal of Student Research*, *13*(1), 780–782. https://doi.org/10.47611/jsrhs.v13i1.5996

Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2019.07.059

Bassiouni, M. M., Chakrabortty, R. K., Sallam, K. M., & Hussain, O. K. (2024). Deep learning approaches to identify order status in a complex supply chain. *Expert Systems with Applications*, *250*(April), 123947. https://doi.org/10.1016/j.eswa.2024.123947

Biji, K. B., Ravishankar, C. N., Mohan, C. O., & Srinivasa Gopal, T. K. (2015). Smart packaging systems for food applications: a review. *Journal of Food Science and Technology*, *52*(10), 6125–6135. https://doi.org/10.1007/s13197-015-1766-7

Cao, S., Zeng, Y., Yang, S., & Cao, S. (2021). Research on Python Data Visualization Technology. *Journal of Physics: Conference Series*, *1757*(1). https://doi.org/10.1088/1742-6596/1757/1/012122

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. https://doi.org/10.1613/jair.953

Chen, S., Brahma, S., Mackay, J., Cao, C., & Aliakbarian, B. (2020). The role of smart packaging system in food supply chain. In *Journal of Food Science*. https://doi.org/10.1111/1750-3841.15046

Chitra, K., M, R. S., & Arivoli, S. (2022). a Review on Traditional and Modern Food Packaging Technologies. *International Journal of Novel Research and Development (Www.Ijnrd.Org)*, *7*(9), 769–775. https://doi.org/IJNRD2209089

Chong, J. W., Kim, W., & Hong, J. S. (2022). Optimization of Apparel Supply Chain

Using Deep Reinforcement Learning. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3205720

Dalal, S., Lilhore, U. K., Simaiya, S., Radulescu, M., & Belascu, L. (2024). Improving efficiency and sustainability via supply chain optimization through CNNs and BiLSTM. *Technological Forecasting and Social Change*, *209*(October), 123841. https://doi.org/10.1016/j.techfore.2024.123841

Danach, K., Dirani, A. El, & Rkein, H. (2024). Revolutionizing Supply Chain Management with AI: A Path to Efficiency and Sustainability. *IEEE Access*, *October*, 188245–188255. https://doi.org/10.1109/ACCESS.2024.3474531

DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS Dataset. (n.d.).

De Santis, R. B., De Aguiar, E. P., & Goliatt, L. (2017). Predicting material backorders in inventory management using machine learning. *2017 IEEE Latin American Conference on Computational Intelligence*, *LA-CCI 2017 - Proceedings*, *2017-Novem*, 1–6. https://doi.org/10.1109/LA-CCI.2017.8285684

Deborah, N., & Sugihartanto, M. F. (2024). Assessment of Sustainable Packaging Supply Chain Management Using the Life Cycle Assessment Method (Case Study: FMCG Company in Indonesia). *Procedia Computer Science, 234,* 654–662. https://doi.org/10.1016/j.procs.2024.03.051

Del Giudice, M., Chierici, R., Mazzucchelli, A., & Fiano, F. (2020). Supply chain management in the era of circular economy: the moderating effect of big data. *International Journal of Logistics Management*. https://doi.org/10.1108/IJLM-03-2020-0119

Dey, A. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174–1179.

Dong, C., Akram, A., Andersson, D., Arnäs, P. O., & Stefansson, G. (2021). The impact of emerging and disruptive technologies on freight transportation in the digital era: current state and future trends. *International Journal of Logistics Management*, *32*(2), 386–412. https://doi.org/10.1108/IJLM-01-2020-0043

Dörr, L., Brandt, F., Pouls, M., & Naumann, A. (2020). Fully-Automated Packaging Structure Recognition in Logistics Environments. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*. https://doi.org/10.1109/ETFA46521.2020.9212152

Douaioui, K., Oucheikh, R., & Mabrouki, C. (2024). Enhancing Supply Chain Resilience: Rime-Clustering and Ensemble Deep Learning Strategies for Late Delivery Risk Prediction. *Logforum*, *20*(1), 55–70. https://doi.org/10.17270/J.LOG.001007

Dumitrascu, O., Dumitrascu, M., & Dobrotă, D. (2020). Performance evaluation for a sustainable supply chain management system in the automotive industry using artificial intelligence. *Processes*, *8*(11), 1–20. https://doi.org/10.3390/pr8111384

Eni, L. N., Groenewald, E. S., & Hamidi, I. A. (2024). Optimizing Supply Chain Processes through Deep learning Algorithms: A Managerial Approach. *Journal of Informatics Education and Research*, *February*. https://doi.org/10.52783/jier.v4i1.567

Ergun, O., Karakus, G., Keskinocak, P., Swann, J., & Villarreal, M. (2009). Humanitarian Supply Chain Management – An Overview. *Dagstuhl Seminar Proceedings*.

Falkner, D., Bögl, M., Gattinger, A., Stainko, R., Zenisek, J., & Affenzeller, M. (2024). Integrating Machine Learning into Supply Chain Management: Challenges and Opportunities. *Procedia Computer Science*, *232*(2023), 1779–1788. https://doi.org/10.1016/j.procs.2024.01.176

Farooq, M., & Yee Yen, Y. (2024). *Artificial Intelligence in Supply Chain Management:*

A Comprehensive Review and Framework for Resilience and Sustainability.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Florescu, M. S., Ceptureanu, E. G., Cruceru, A. F., & Ceptureanu, S. I. (2019). Sustainable supply chain management strategy influence on supply chain management functions in the oil and gas distribution industry. *Energies*. https://doi.org/10.3390/en12091632

Fosso Wamba, S., Gunasekaran, A., Papadopoulos, T., & Ngai, E. (2018). Big data analytics in logistics and supply chain management. *The International Journal of Logistics Management*, *29*(2), 478–484. https://doi.org/10.1108/IJLM-02-2018-0026

Garvey, M. D., Carnovale, S., & Yeniyurt, S. (2015). An analytical framework for supply network risk propagation: A Bayesian network approach. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2014.10.034

Gayathri, R., Rani, S. U., Čepová, L., Rajesh, M., & Kalita, K. (2022). A Comparative Analysis of Machine Learning Models in Prediction of Mortar Compressive Strength. *Processes*. https://doi.org/10.3390/pr10071387

Gegeckienė, L., Venytė, I., Karpavičė, J., Tambo, T., Vaitasius, K., & Pauliukaitis, D. (2022). NEAR FIELD COMMUNICATION (NFC) TECHNOLOGY IN THE PACKAGING INDUSTRY. *International Symposium on Graphic Engineering and Design*. https://doi.org/10.24867/GRID-2022-p54

Ghaani, M., Cozzolino, C. A., Castelli, G., & Farris, S. (2016). An overview of the intelligent packaging technologies in the food sector. In *Trends in Food Science and Technology*. https://doi.org/10.1016/j.tifs.2016.02.008

Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A

Review. *Information (Switzerland)*, *15*(4). https://doi.org/10.3390/info15040235

Gupta, R. (2024). *An In-depth Study of Sustainable Packaging Strategies in Supply Chains: Perspectives from Industry Experts* (pp. 1–11). https://doi.org/10.21203/rs.3.rs-4284108/v1

Hà Huy, P. (2021). Technology in Supply Chain Management: a Trend Analysis for the Fresh Connection. *ARNHEM, HAN UNIVERSITY OF APPLIED SCIENCES*, *December*. https://doi.org/10.13140/RG.2.2.17161.36966/1

Han, J. H. (2005). New technologies in food packaging: Overiew. *Innovations in Food Packaging*, 3–11. https://doi.org/10.1016/B978-012311632-1/50033-4

Han, J. W., Ruiz-Garcia, L., Qian, J. P., & Yang, X. T. (2018). Food Packaging: A Comprehensive Review and Future Trends. In *Comprehensive Reviews in Food Science and Food Safety*. https://doi.org/10.1111/1541-4337.12343

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00369-8

Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, *44*(3), 348–361. https://doi.org/10.3102/1076998619832248

Haque, R., & Kelly, J. (n.d.). *Forecasting Sales and Inventory in Supply Chain using Machine Learning Methods MSc Research Project MSc in Data Analytics Rian Dwi Putra School of Computing National College of Ireland Supervisor :*

Hendriksen, C. (2023). Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption? *Journal of Supply Chain Management*. https://doi.org/10.1111/jscm.12304

Hinton, G. (2009). Deep belief networks. *Scholarpedia*, *4*(5), 5947. https://doi.org/10.4249/scholarpedia.5947

Huang, X. (2023). Introduction to Programming in Python. *Land Carbon Cycle Modeling*, 343–351. https://doi.org/10.1201/9780429155659-50

Jackson, I., Ivanov, D., Dolgui, A., & Namdar, J. (2024). Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. *International Journal of Production Research*, *62*(17), 6120–6145. https://doi.org/10.1080/00207543.2024.2309309

Jahin, M. A., Naife, S. A., Saha, A. K., & Mridha, M. F. (2023). *AI in Supply Chain Risk Assessment: A Systematic Literature Review and Bibliometric Analysis*.

Jawad, E. (2023). the Deep Neural Network-a Review. *Ijrdo -Journal of Mathematics*, *9*(9), 1–5. https://doi.org/10.53555/m.v9i9.5842

Jianying, F., Bianyu, Y., Xin, L., Dong, T., & Weisong, M. (2021). Evaluation on risks of sustainable supply chain based on optimized BP neural networks in fresh grape industry. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2021.105988

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Karpavice, J., Hafith, I. A., Tambo, T., Chinello, F., Venyte, I., & Gegeckiene, L. (2023). Experimental approaches to NFC-enabled packaging for UX / CX of physical artefacts: A technology maturity study. *Procedia Computer Science*, *219*(2022), 577–585. https://doi.org/10.1016/j.procs.2023.01.326

Keskin, B. (2022). Digital supply chain transformation at Jabil: integrated supply and demand planning via concurrent planning. *The Case For Women*. https://doi.org/10.1108/cfw.2022.000005

Khedr, A. M., & S, S. R. (2024). Enhancing supply chain management with deep learning and machine learning techniques: A review. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(4). https://doi.org/10.1016/j.joitmc.2024.100379

Kollia, I., Stevenson, J., & Kollias, S. (2021). Ai-enabled efficient and safe food supply chain. In *Electronics (Switzerland)*. https://doi.org/10.3390/electronics10111223

Kosasih, E. E., & Brintrup, A. (2022). A machine learning approach for predicting hidden links in supply chain with graph neural networks. *International Journal of Production Research*, *60*(17), 5380–5393. https://doi.org/10.1080/00207543.2021.1956697

Kot, S. (2018). Sustainable supply chain management in small and medium enterprises. *Sustainability (Switzerland)*. https://doi.org/10.3390/su10041143

Kozik, N. (2020). Sustainable packaging as a tool for global sustainable development. *SHS Web of Conferences*. https://doi.org/10.1051/shsconf/20207404012

Kumari, N., Chaudhary, D., Kaur, H., & Yadav, A. L. (2023). Artificial Intelligence in Supply Chain Optimization. *2023 International Conference on IoT, Communication and Automation Technology, ICICAT 2023*. https://doi.org/10.1109/ICICAT57735.2023.10263631

Laldin Ismaeil, M. K., & Lalla, A. F. (2024). The Role and Impact of Artificial Intelligence on Supply Chain Management: Efficiency, Challenges, and Strategic Implementation. *Journal of Ecohumanism*, *3*(4), 89–106. https://doi.org/10.62754/joe.v3i4.3461

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Lee, S. K., & Rah, M. (2024). *DEEP LEARNING IN RETAIL SUPPLY CHAIN*

MANAGEMENT : AN. 0–4.

Lei, Y., Qiaoming, H., & Tong, Z. (2023). Research on Supply Chain Financial Risk Prevention Based on Machine Learning. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2023/6531154

Li, D. C., Wang, S. Y., Huang, K. C., & Tsai, T. I. (2022). Learning class-imbalanced data with region-impurity synthetic minority oversampling technique. *Information Sciences*. https://doi.org/10.1016/j.ins.2022.06.067

Li, J., Zhang, R., Jin, Y., & Zhang, H. (2022). Optimal Path of Internet of Things Service in Supply Chain Management Based on Machine Learning Algorithms. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2022/4844993

Lis, A., Sudolska, A., & Tomanek, M. (2020). Mapping research on sustainable supply-chain management. *Sustainability (Switzerland)*, *12*(10). https://doi.org/10.3390/SU12103987

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When Machine Learning Meets Privacy: A Survey and Outlook. *ACM Computing Surveys*, *54*(2), 1–35. https://doi.org/10.1145/3436755

Lu, L. X., & Swaminathan, J. M. (2015). Supply Chain Management. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, *December 2015*, 709–713. https://doi.org/10.1016/B978-0-08-097086-8.73032-7

MacCarthy, B. L., & Ivanov, D. (2022). The Digital Supply Chain—emergence, concepts, definitions, and technologies. In *The Digital Supply Chain*. https://doi.org/10.1016/B978-0-323-91614-1.00001-0

Mageto, J. (2021). Big data analytics in sustainable supply chain management: A focus on manufacturing supply chains. In *Sustainability (Switzerland)*.

https://doi.org/10.3390/su13137101

Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, *9*(1), 381–386. https://doi.org/10.21275/art20203995

Makkar, S., & Solanki, V. K. (2020). ICICCT 2019 – System Reliability, Quality Control, Safety, Maintenance and Management. *ICICCT 2019 – System Reliability, Quality Control, Safety, Maintenance and Management*, *January*. https://doi.org/10.1007/978-981-13-8461-5

Mallesham, G. (2022). The Role of AI and ML in Revolutionizing Supply Chain Management. *International Journal of Scientific Research and Management (IJSRM)*, *10*(06), 918–928. https://doi.org/10.18535/ijsrm/v10i6.ec05

Malviya, L., Chittora, P., Chakrabarti, P., Vyas, R. S., & Poddar, S. (2022). Backorder prediction in the supply chain using machine learning. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2020.11.558

Mastos, T. D., Nizamis, A., Terzi, S., Gkortzis, D., Papadopoulos, A., Tsagkalidis, N., Ioannidis, D., Votis, K., & Tzovaras, D. (2021). Introducing an application of an industry 4.0 solution for circular supply chain management. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2021.126886

McCane, B. (2009). Introductory Programming with Python Curriculum. *The Python Papers Monograph*, *1*, 1–11.

McKinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing*, *January 2011*, 1–9.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). DEFINING SUPPLY CHAIN MANAGEMENT. *Journal of Business Logistics*. https://doi.org/10.1002/j.2158-1592.2001.tb00001.x

Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information*, *15*(9), 517. https://doi.org/10.3390/info15090517

Mohamed-Iliasse, M., Loubna, B., & Abdelaziz, B. (2022). Machine Learning in Supply Chain Management: A Systematic Literature Review. *International Journal of Supply and Operations Management*, *9*(4), 398–416. https://doi.org/10.22034/ijsom.2021.109189.2279

Mohammed, I. A., Energy, F., & Kothinti, R. R. (2025). *Predictive Analytics in Supply Chain Management : The Role of AI and Machine Learning in Demand Forecasting Advances in Consumer Research Predictive Analytics in Supply Chain Management : The Role of AI and Machine Learning in Demand Forecasting*. *February*.

Morashti, J. A., An, Y., & Jang, H. (2022). A Systematic Literature Review of Sustainable Packaging in Supply Chain Management. *Sustainability (Switzerland)*, *14*(9). https://doi.org/10.3390/su14094921

Mugoni, E., Kanyepe, J., & Tukuta, M. (2024). Sustainable Supply Chain Management Practices (SSCMPS) and environmental performance: A systematic review. *Sustainable Technology and Entrepreneurship*, *3*(1), 100050. https://doi.org/10.1016/j.stae.2023.100050

Muhammed, M. A., Bülent, S., & Balcıoğlu, Y. S. (2025). MACHINE LEARNING APPLICATIONS IN PACKAGING INDUSTRY OPTIMIZATION: A SYSTEMATIC LITERATURE REVIEW AND FUTURE RESEARCH DIRECTIONS. *12th International New York Academic Research Congress on Social, Humanities, Administrative, and Educational Sciences*, *January*. https://doi.org/10.30546/19023.978-9952-8545-4-1.20980

Nasurudeen Ahamed, N., & Karthikeyan, P. (2020). A Reinforcement Learning Integrated in Heuristic search method for self-driving vehicle using blockchain in supply chain management. *International Journal of Intelligent Networks*. https://doi.org/10.1016/j.ijin.2020.09.001

Nosrati, M. (2011). Python: An appropriate language for real world programming. *World Applied Programming*, *1*(2), 110–117.

Oyewola, D. O., Dada, E. G., Omotehinwa, T. O., Emebo, O., & Oluwagbemi, O. O. (2022). Application of Deep Learning Techniques and Bayesian Optimization with Tree Parzen Estimator in the Classification of Supply Chain Pricing Datasets of Health Medications. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app121910166

Ozcan, A. (2020). New approaches in smart packaging technologies. *International Symposium on Graphic Engineering and Design*. https://doi.org/10.24867/GRID-2020-p1

Pandey, B. K., Paramashivan, M. A., Mahajan, R., Mahajan, D. A., Behare, N., Gowwrii, G., & Pramanik, S. (2024). *Applications of Artificial Intelligence and Machine Learning in Supply Chain Management*. *July*, 74–90. https://doi.org/10.4018/979-8-3693-1347-3.ch006

Park, K. J. (2021). Determining the tiers of a supply chain using machine learning algorithms. *Symmetry*. https://doi.org/10.3390/sym13101934

Pasupuleti, V., Thuraka, B., Kodete, C. S., & Malisetty, S. (2024). Enhancing Supply Chain Agility and Sustainability through Machine Learning: Optimization Techniques for Logistics and Inventory Management. *Logistics*, *8*(3), 73. https://doi.org/10.3390/logistics8030073

Pattnaik, S., Liew, N., Kures, A. O., Pinsky, E., & Park, K. (2024). *Catalyzing Supply*

Chain Evolution: A Comprehensive Examination of Artificial Intelligence Integration in Supply Chain Management. 57. https://doi.org/10.3390/engproc2024068057

Pimentel, J. F., Murta, L., Braganholo, V., & Freire, J. (2019). A large-scale study about quality and reproducibility of jupyter notebooks. *IEEE International Working Conference on Mining Software Repositories*, *2019-May*, 507–517. https://doi.org/10.1109/MSR.2019.00077

Polydoros, A. S., & Nalpantidis, L. (2017). Survey of Model-Based Reinforcement Learning: Applications on Robotics. *Journal of Intelligent and Robotic Systems: Theory and Applications*. https://doi.org/10.1007/s10846-017-0468-y

Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, *4*, 19–29. https://doi.org/https://doi.org/10.1016/j.dsm.2021.12.002

Qu, C., & Kim, E. (2024). Reviewing the Roles of AI-Integrated Technologies in Sustainable Supply Chain Management: Research Propositions and a Framework for Future Directions. *Sustainability (Switzerland)*, *16*(14). https://doi.org/10.3390/su16146186

Qu, K. (2024). Research on linear regression algorithm. *MATEC Web of Conferences*, *395*, 01046. https://doi.org/10.1051/matecconf/202439501046

Radivojević, G., Mitrović, M., & Popović, D. (2022). *Overview of Criteria and Methods of Machine Learning for Supplier Selection*.

Radmanesh, S. A., Haji, A., & Fatahi Valilai, O. (2023). Blockchain-Based Architecture for a Sustainable Supply Chain in Cloud Architecture. *Sustainability (Switzerland)*, *15*(11). https://doi.org/10.3390/su15119072

Rajeev, S. (2023). Importance of Packaging in Supply Chain Management. *Asian Journal* *of Management*, *16*(1), 17–23. https://doi.org/10.52711/2321-5763.2024.00003

Ramya, A. A., & Stephan Thangaiah, I. S. (2022). A Study to Enhance Supply Chain Performance in Online Merchandising by Integrating Digitalization to SCOR Model using Artificial Neural Network (ANN). *International Journal of Intelligent Systems and Applications in Engineering*, *10*(4), 23–28.

Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017). Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. https://doi.org/10.1109/JCDL.2017.7991618

Rawat, A. (2020). A Review on Python Programming. *International Journal of Research in Engineering, Science and Management*.

Rezaee, Z. (2018). Supply chain management and business sustainability synergy: A theoretical and integrated perspective. *Sustainability (Switzerland)*. https://doi.org/10.3390/su10010275

Riansyah, M., Suwilo, S., & Zarlis, M. (2023). Improved Accuracy In Data Mining Decision Tree Classification Using Adaptive Boosting (Adaboost). *SinkrOn*. https://doi.org/10.33395/sinkron.v8i2.12055

Rojek, I., Jasiulewicz-Kaczmarek, M., Piszcz, A., Galas, K., & Mikołajewski, D. (2024). Review of the 6G-Based Supply Chain Management within Industry 4.0/5.0 Paradigm. *Electronics (Switzerland)*, *13*(13). https://doi.org/10.3390/electronics13132624

Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, *61*(20), 7151–7179. https://doi.org/10.1080/00207543.2022.2140221

Sakas, D. P., Giannakopoulos, N. T., Terzi, M. C., & Kanellos, N. (2023). Engineering Supply Chain Transportation Indexes through Big Data Analytics and Deep Learning. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app13179983

Samuels, A. (2024). Examining the integration of artificial intelligence in supply chain management from Industry 4.0 to 6.0: a systematic literature review. *Frontiers in Artificial Intelligence*, *7*(January), 0–17. https://doi.org/10.3389/frai.2024.1477044

Sankhye, S., & Hu, G. (2020). Machine Learning Methods for Quality Prediction in Production. *Logistics*. https://doi.org/10.3390/logistics4040035

Sardar, S. K., Sarkar, B., & Kim, B. (2021). Integrating machine learning, radio frequency identification, and consignment policy for reducing unreliability in smart supply chain management. *Processes*. https://doi.org/10.3390/pr9020247

Sayyad, J., Attarde, K., & Yilmaz, B. (2024). Improving Machine Learning Predictive Capacity for Supply Chain Optimization through Domain Adversarial Neural Networks. *Big Data and Cognitive Computing*, *8*(8). https://doi.org/10.3390/bdcc8080081

Sewak, M., Sahay, S. K., & Rathore, H. (2020). An overview of deep learning architecture of deep neural networks and autoencoders. *Journal of Computational and Theoretical Nanoscience*. https://doi.org/10.1166/jctn.2020.8648

Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00329-2

Shamsuddoha, M., Khan, E. A., Chowdhury, M. M. H., & Nasir, T. (2025). Revolutionizing Supply Chains: Unleashing the Power of AI-Driven Intelligent Automation and Real-Time Information Flow. *Information (Switzerland)*, *16*(1), 1–19. https://doi.org/10.3390/info16010026

Shankha Shubhra Goswami, S. M., Sarkar, S., Gupta, and K. K., Sahoo, S. K., & Halder, R. (2024). *Artificial Intelligence Enabled Supply Chain Management : Unlocking New Opportunities and Challenges*. *September 2023*, 1–16. https://doi.org/10.47852/bonviewAIA42021814

Sharma, P., & Dash, B. (2022). Smart SCM Using AI and Microsoft 365. *IJARCCE*. https://doi.org/10.17148/ijarcce.2023.12106

Shebeshe, E. N., & Sharma, D. (2024). Sustainable supply chain management and organizational performance: the mediating role of competitive advantage in Ethiopian manufacturing industry. *Future Business Journal*, *10*(1). https://doi.org/10.1186/s43093-024-00332-6

Song, M., Fisher, R., de Sousa Jabbour, A. B. L., & Santibañez Gonzalez, E. D. R. (2022). Green and sustainable supply chain management in the platform economy. *International Journal of Logistics Research and Applications*, *25*(4–5), 349–363. https://doi.org/10.1080/13675567.2022.2045763

Soundarraj, P., Aarif, M., Gangadharan, S., Naqvi, S. R., Assihalaf, N. K., & Salih Mahdi, A. (2023). Smart Product Packing and IoT Marketing: Enhancing Customer Interaction. *Proceedings of the 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2023*, *December 2023*, 1–6. https://doi.org/10.1109/ICSES60034.2023.10465408

Sruthy, S. K. (2024). Enhancing Supply Chain Resilience through Artificial Intelligence: A Strategic Framework for Executives. *Emerging Science Journal*, *8*(4), 1462–1473. https://doi.org/10.28991/ESJ-2024-08-04-013

Stoychev, V. (2023). *The potential benefts of implementing machine learning in supply chain management*. *00124824*.

Submitted, T., Fulfillment, P., & Studies, P. (2024). *Optimizing Supply Chain Dynamics* *using Machine Learning*.

Summerfield, B. M. (2011). *Programming in Python 3: A Complete Introduction to the Python Language*.

Sun, C. (2012). Application of RFID Technology for Logistics on Internet of Things. *AASRI Procedia*, *1*, 106–111. https://doi.org/10.1016/j.aasri.2012.06.019

Tajima, M. (2007). Strategic value of RFID in supply chain management. *Journal of Purchasing and Supply Management*. https://doi.org/10.1016/j.pursup.2007.11.001

Terrada, L., El Khaili, M., & Ouajji, H. (2022). Demand Forecasting Model using Deep Learning Methods for Supply Chain Management 4.0. *International Journal of Advanced Computer Science and Applications*, *13*(5), 704–711. https://doi.org/10.14569/IJACSA.2022.0130581

Tirkolaee, E. B., Sadeghi, S., Mooseloo, F. M., Vandchali, H. R., & Aeini, S. (2021). Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas. In *Mathematical Problems in Engineering*. https://doi.org/10.1155/2021/1476043

Tundys, B. (2020). Sustainable supply chain management – past, present and future. *Prace Naukowe Uniwersytetu Ekonomicznego We Wrocławiu*, *64*(3), 187–207. https://doi.org/10.15611/pn.2020.3.15

Van Rossum, G., & Drake, F. L. . (2006). *An introduction to Python, release 2.5*.

Verma, P. (2024). *Transforming Supply Chains Through AI: Demand Forecasting, Inventory Management, and Dynamic Optimization*. *September*.

Walker, K. W., & Jiang, Z. (2019). Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: A machine-learning approach. *Journal of Academic Librarianship*. https://doi.org/10.1016/j.acalib.2019.02.013

Wan, F. (2021). XGBoost Based Supply Chain Fraud Detection Model. *2021 IEEE 2nd*

International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2021. https://doi.org/10.1109/ICBAIE52039.2021.9390041

Wang, H., Alidaee, B., & Sua, L. S. (2023). *Enhancing supply chain security with automated machine learning*.

Wang, J., Li, L., & Zeller, A. (2020). Better Code, Better Sharing: On the Need of Analyzing Jupyter Notebooks. *Proceedings - 2020 ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results, ICSE-NIER 2020*, *1*, 53–56. https://doi.org/10.1145/3377816.3381724

Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open-Source Software*. https://doi.org/10.21105/joss.03021

Wenzel, H., Smit, D., & Sardesai, S. (2019). A literature review on machine learning in supply chain management Standard-Nutzungsbedingungen. *Proceedings of the Hamburg International Conference of Logistics (HICL)* .

Yang, M., Lim, M. K., Qu, Y., Ni, D., & Xiao, Z. (2023). Supply chain risk management with machine learning technology: A literature review and future research directions. *Computers and Industrial Engineering*. https://doi.org/10.1016/j.cie.2022.108859

Yashan, N., Materials, Y., & Limited, E. P. (2024). Blockchain and AI Integration: Transforming Transparency in Supply Chain Management. *European Economic Letters*, *September*. https://doi.org/10.52783/eel.v14i3.1885

Yun, J. J., Kim, D., & Yan, M. R. (2020). Open innovation engineering—preliminary study on new entrance of technology to market. In *Electronics (Switzerland)*. https://doi.org/10.3390/electronics9050791

Zaadnoordijk, L., Besold, T. R., & Cusack, R. (2022). Lessons from infant learning for unsupervised machine learning. *Nature Machine Intelligence*, *4*(6), 510–520.

https://doi.org/10.1038/s42256-022-00488-2

Žunić, E., Korjenić, K., Hodžić, K., & Đonko, D. (2020). Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data. *International Journal of Computer Science and Information Technology*. https://doi.org/10.5121/ijcsit.2020.12203