Digital Twins: Warehouses of the Future

by

Moutaz Faisal Ali
B.S., American University of Sharjah, Sharjah, UAE, 2012

and

Yumeng Chen

M.S., the University of Adelaide, South Australia, AUS, 2014

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

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

May 2023

Signature of Author: ____
Department of Supply Chain Management
May 12, 2023

Signature of Author: ____
Department of Supply Chain Management
May 12, 2023

Certified by: ____
Dr. María Jesús Sáenz Gil De Gómez
Executive Director, Supply Chain Management Masters Programs
Director, MIT Digital Supply Chain Transformation
Capstone Advisor

Certified by: ____
Dr. Thomas Koch
Postdoctoral Associate
Capstone Co-Advisor

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

Digital Twins: Warehouses of the Future

by

Moutaz Faisal Ali

and

Yumeng Chen

Submitted to the Program in Supply Chain Management
on May 12, 2023, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

As warehouse operations grow in complexity, many organizations turn to digital twins to increase their performance capabilities. Digital twins are virtual replicas of physical entities and their interactions with one another. The technologies in a digital twin capture real-time data to support improvements and decision making. This project focuses on digital twins as a promising solution for enhancing performance metrics within a warehouse operation, including efficiency, productivity, and scalability, particularly in the picking process. Because order picking is one of the most labor-intensive activities in the warehouse, we examined the feasibility of employing machine learning to forecast labor requirements. To develop a digital twin prototype for the order-picking process, we explored several technologies aimed at improving efficiency and productivity: sensors, automated guided vehicles (AGVs), picking robots, and automated storage and retrieval systems (AS/RS). By conducting stakeholder interviews, process mapping, and gathering data pertaining to historical order demand and daily labor hours, we formulated a workforce forecasting model that harnesses machine learning techniques. Leveraging the forecasting model alongside the recommended technologies will allow the warehouse team to enhance their key performance indicators (KPIs) for efficiency and productivity. This project culminates in a comprehensive roadmap for implementing these solutions, with the potential for scaling this digital twin prototype to other processes and warehouses.

Capstone Advisor: Dr. María Jesús Sáenz Gil De Gómez
Title: Executive Director, MIT Supply Chain Management Master Programs
Director, MIT Digital Supply Chain Transformation

Capstone Co-Advisor: Dr. Thomas Koch
Title: Postdoctoral Associate

ACKNOWLEDGMENTS

We thank our advisors, Dr. María Jesús Sáenz Gil De Gómez, and Dr. Thomas Koch. We would also like to thank our sponsor company for providing us with this opportunity.

Furthermore, we would like to acknowledge the contributions of our writing coach, Toby Gooley, whose expertise, and feedback were instrumental in shaping this project.

Lastly, we would like to express our heartfelt thanks to the CTL team and our SCM colleagues for their unwavering help, support, and encouragement.

I would like to express my appreciation to my project partner, Taz. Together, we navigated numerous challenges in this project, enduring late nights, engaging in fruitful discussions, resolving disagreements, and above all, fostering a spirit of collaboration and mutual encouragement. Taz's dedication and contributions were invaluable to the success of our endeavor.

I would also like to dedicate this project to my beloved family and friends from China, Australia, and the United States. Their unwavering love and support have been a constant source of inspiration throughout this journey. I am truly grateful for their presence in my life.

Yumeng (CC)

Table of Contents

List Of Figures

List Of Tables

1 Introduction

1.1 Motivation

In a continuously changing world that is faced with bigger challenges and disruptions every day, tackling the usual business challenges, such as systems complexity, higher customer expectations, and demand and supply imbalance, is not enough. The world of supply chains, logistics, and warehouse operations is having its fair share of increased complexity trying to adjust to the new normal in a post COVID era to be able to achieve the highest service levels, react to raw material and labor shortages, and continue to innovate.

To streamline operations, a lot of warehouse tools and technologies have been developed in the past few decades, i.e., warehouse management systems (WMS), labor management systems (LMS), yard management systems (YMS), transportation management systems (TMS), and enterprise resource planning (ERP) systems. However, using them introduced new complexities to the processes.

Our Sponsoring company has over 150 warehouses across North America, services thousands of customers, and has over 30 years of experience in warehousing and distribution services. Among the services they provide are consolidation, deconsolidation, and fulfillment of goods. Furthermore, they have a global footprint with facilities in numerous regions around the world. The systems mentioned above are working in isolation to perform different tasks in warehouses in different locations, and the company believes that this situation creates inefficiencies. Meanwhile dealing with the vast variety of customers' applications and number of product stock keeping units (SKUs) leads to reduced efficiencies and productivities, difficulty in coping with periods of increased demand which affects service levels, and a less effective utilization of resources. Therefore, the company is looking for solutions to deal with the increasing systems complexity while maintaining targeted growth.

We looked at different capabilities' enhancement options, and the concept of digital twins comes to the top of the list. Digital twins are comprised of a blend of technologies and analytical capabilities that gather data in real-time, and they serve as virtual imitations of real objects and their interactions (Tozanli & Saenz, 2022). Managers can make data-driven decisions on improving warehouse operations, as digital twins allow organizations to create custom warehouse models and generate important and useful insights about the operations and performance with real-time data. Development of custom digital twins for an organization with multiple facilities can also be used to replicate the successful model of one facility to other facilities.

1.2 Problem Statement and Research Questions

The digital twin initiative not only aims to unify a variety of different data sources fed by their transactional systems, but also to utilize novel technology (IoT, sensors, warehouse technologies, mobile experiences) to improve business performance and prepare the organization for the future of warehousing. To define the warehouse operations capabilities that can be enhanced by using digital twins we studied a specific facility's operations. Mapping different processes in that warehouse was a first step toward understanding the possibilities of integrating the different systems. We also identified the technologies that could be added to the existing system to improve performance.

There are three capabilities we will look at in this project. First, efficiency, where we will look at the different operations in the warehouse, and how they can be improved in the different activities performed and the utilization of equipment. Second, productivity, as the higher the throughput of the activities the better service level can be provided. Finally, scalability, as this is an initiative that will be applied on a smaller scale, and the company would want to replicate it in other locations or activities.

To study these capabilities and the possibilities of enhancing them we will need to get data from the existing systems. Gathering existing real-time data from the different systems is a key challenge, as the systems are not linked, data is not unified, and each system controls a specific type of operation. The type of data is also another challenge for this project since every system measures different outputs in the operation cycle.

One of the main areas where we believe a digital twin can yield promising results is the picking process and workforce forecasting within that process. In a warehouse operation where the workforce varies depending on customer's inbound shipments and consumers' outbound orders, the workforce distribution and forecasting it is a major issue that affects the operational teams and processes. Considering that warehouses rely on temporary workers to give the warehouse team the flexibility to scale up or down, depending on the demand periods, an improvement on efficiency and productivity with the possibility of predicting the workforce will be highly useful to the operation. The technologies associated with a digital twin, coupled with a machine learning workforce forecasting model, that leverages on the information from the operation and the sensors that are included in the technologies can help the warehouse team to handle periods of increased demand effectively, and prepare for periods of low demand properly.

To implement a digital twin, we need to gather data from the different systems, and that is going to be one of the main challenges for such a project. Data gathering requires significant manual downloading of big data sets with different formats from multiple systems; given the various customers and stock keeping units (SKUs) within the facility, warehouse operations systems complexity, and different Key Performance Indicator (KPI) measurement techniques, this task can be time consuming and incredibly challenging. Then the data will be cleaned and uploaded to a common platform for further analysis.

In this context, the questions to be answered in this project include:

1. How can digital twins be used in warehouse picking process and help enhance capabilities like efficiency, productivity, and scalability?

2. What technology is useful in pursuit of the goal of digital twins for this project?
3. How can we forecast (dependent on data quality) workforce requirements for the picking process in a warehouse using a digital twin?
4. What are the implementation phases for a digital twin solution, its challenges, and potential values?

1.3 Scope: Project Goals and Expected Outcomes

This project aims to pair up technologies with the concept of the warehouse digital twins through 1) identifying key technologies to be used in the twin; and 2) assessing digital twin interactions by measuring the capabilities, in terms of efficiency, productivity and scalability. To be more specific, we will be looking into capturing data by technologies implemented in the physical warehouse, building and validating models, and monitoring performance. Warehouse workforce forecasting is one of the key assessment components as it’s one of the major pain points for the warehouse operations.

The plan of work of this project covers the following:

1. Interview stakeholders and visit the facility.
2. Map the picking process within the warehouse and understand the systems connected to it.
    - Map the As-Is state for the current process.
3. Gather, clean, and analyze the data from the customer and the different systems involved in the process.
    - Statistics analysis and gaps identification.
4. Explore technologies for warehouse capabilities improvements.
5. Build a model for workforce prediction.
6. Create a framework that will assist in the implementation.

7. Visualize performance capabilities, efficiency, productivity, and scalability, through an actionable dashboard to make decisions.
8. Include a roadmap for technologies implementation.

A digital twin model connecting all the relevant systems together will help test the performance enhancing initiatives and will allow the company to be more proactive and enhance its visibility on the operations once implemented. Additionally, the organization will be able to improve service levels and customers satisfaction by making data-driven decisions, reducing the time it takes to target new opportunities and onboard these customers once they have won their business.

2 State of the Art

The two key purposes of this project, as mentioned above, are (1) recommending beneficial warehouse digital twin technologies and (2) forecasting the warehouse picking process workforce requirements through a digital twin. We reviewed literature in several areas. Main topics include digital twins in warehouse applications, enabling technologies for digital twins' implementation, warehouse digital transformation frameworks, and warehouse workforce forecasting methods.

2.1 Digital Twins

A digital twin is a digital representation of a physical asset or system composed of multiple models that change and evolve concurrently with their physical counterparts throughout their lifecycle. Vrabič et al. (2018) explain that these models are integrated through a common model space, which serves as a connection between the digital twin and the physical asset, facilitating the study of complex behaviors. Jones et al. (2020) define a digital twin as a concept encompassing a physical entity, a virtual counterpart, and data connections between the two, which are utilized to enhance the performance of physical entities through the application of computational techniques. It is essential to establish a common understanding of a digital twin before progressing further in the project. Tozanli and Saenz (2022) describe digital twins as virtual replicas of physical entities, such as cities, warehouses, environments, products, and systems, including their interactions.

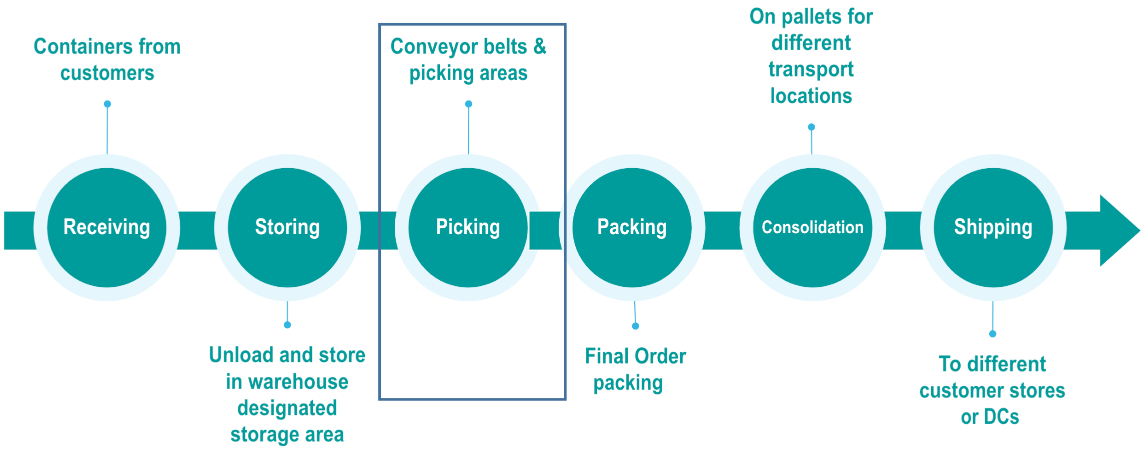
Digital twins can be employed to improve various logistics systems, including warehouses and last-mile delivery networks. Gerlach et al. (2021) highlight that digital twins offer a comprehensive, real-time view of the system, enabling the implementation of system controls and optimization of processes, leading to increased efficiency. Furthermore, digital twins facilitate simulations of different scenarios to determine the most favorable outcomes based on actual data concerning processes and assets.

Digital twins provide a valuable instrument for enhancing logistics systems by offering a comprehensive view of the system in real-time, allowing for system controls to be implemented and process optimization to occur, resulting in increased efficiencies.

2.2 Process Mapping

To understand the warehouse operations and the flow of operations, we conducted a series of interviews with the warehouse operations team. The interviews focused on getting a complete picture of when material arrives at the warehouse in containers until it leaves for its final destination on trucks. The activities covered in these interviews can be summarized in the list below: 1) Receiving; 2) Storing; 3) inventory Replenishment of the conveyer belts & picking; 4) Packing; 5) Consolidating; 6) Shipping, as shown in Figure 1.

Figure 1
Main Warehouse activities



After several interviews and a site visit, it became clear that the replenishment and picking operation should be the focus of our work. This process involves significant manual labor, from the moment sales orders are received in the Warehouse Management System (WMS) until the SKUs/orders reach the packing stage. The warehouse operations team confirmed their interest in seeing increased efficiency and productivity improvements in this area, as it is the main contributor to the workload and workforce forecasting challenges. Consequently, mapping the picking process was the next logical step.

2.3 Digital Twin Framework

Digital twins are being integrated into various industries such as industrial production, medical, smart cities, aerospace, and commercial sectors, and are expected to continue rapid development (Guo & Lv, 2022). Our research on digital twins in warehouses reveals their applications in optimizing layouts, simulating, and testing equipment, predicting maintenance needs, improving inventory management, and optimizing material flows. Ferrari et al (2022) propose a digital twin implementation for an automated storage and retrieval system (AS/RS) in warehousing to enhance automation and digitization advantages in logistics processes. However, literature on simulation-based digital twin concepts for automated warehousing is still limited (Ferrari et al, 2022). Staczek et al (2021) highlight the use of digital twins in testing and improving the design of an autonomous mobile robot (AMR) for intralogistics tasks in challenging production hall environments. The technology proved effective in assessing design assumptions and accelerating the implementation of automated intralogistics systems while reducing costs.

We recognize the need for a structured approach to understand digital twins and their applications in the selected picking process in warehouse operations. Smart warehousing frameworks and technology radar are effective methods for our purpose, which we will discuss in more detail in sections 2.3.1 and 2.3.2.

2.3.1 Smart Warehousing

The evolution of technology has revolutionized the way modern warehouses operate. The emergence of advanced digital technologies and automation has facilitated the integration of efficient and effective systems into warehouse operations. As such, there has been a growing interest in developing a smart warehousing framework that leverages technologies such as digital twins, the Internet of Things (IoT), artificial intelligence (AI), and robotics to optimize warehouse operations. This framework comprises four layers: Sensing, Analyzing, Acting, and Learning (Zhao et al., 2021).

Zhao et al. (2021) explore the application of smart warehousing technologies such as IoT, digital twins, and robotics to improve efficiency, accuracy, and safety in warehouse operations. The technologies enable real-time monitoring, tracking, and data analytics to support better decision-making and overall warehouse management. Similarly, Tekinerdogan et al. (2021) discuss the concept of smart warehousing, which aims to increase service quality, productivity, and efficiency while minimizing costs and failures. Tang et al. (2022) propose a digital twin framework integrating smart warehouse and manufacturing management with the roulette genetic algorithm for demand forecasting in cyclical industries. The proposed framework is demonstrated through a case study of a small-scale textile company, highlighting the importance of inventory optimization in the cyclical industry. Broo et al. (2022) discuss the need to transform passive infrastructure assets into data-centric systems of systems through digital twin technology. They provide literature review of digital twin architecture and a case study of a digital twin implementation in smart infrastructure, highlighting the importance of acquiring a systems perspective, data and information management, and multidisciplinary aspects in digital twin design and implementation. Winkelhaus and Grosse (2022) focus on the role of technology in modern warehouses and identified enabling technologies that categorize into two types: automation technology and digitization technology. They also categorized the key technologies considering the operational process and management purposes, which provides guidance to enabling warehousing technologies such as RFID, smart lighting, blockchain, wearables, automated guided vehicles (AGVs), autonomous robots, drones, etc. Furthermore, the authors examined the impacts of increased digitalization on human factors from a sociotechnical perspective.

The research reveals the importance of technology in optimizing warehouse operations. It highlights the need to leverage advanced digital technologies and automation to increase efficiency, productivity, and safety while minimizing costs and failures. The categorization of enabling technologies according to operational process and management purposes provides guidance in selecting the appropriate technologies for warehouse management. Therefore, the literature provides a useful framework for assessing and implementing smart warehousing technologies in warehouse operations.

We have utilized the categorization of enabling technologies by Winkelhaus and Grosse (2022) to conduct a preliminary assessment of the process in the chosen warehouse, and the results are presented in Table 1 below.

Table 1
Warehouse Picking Process Activities, Resources, and Information

Task	Resources	Information needed to perform task
Receive Sales order: orders received on Company system (WMS)	WMS. Received through the system.	Sales order info from customer (SKUs, qty, delivery date)
Wave orders: Manual assignment of individual orders to pickers	Personnel in charge of waving process	Sales order info from WMS, inventory availability, location on modules to assign pickers in different zones
Check inventory status on modules through WMS	WMS performed task	SKU identification, quantity on modules, overall quantity in WH, location on modules, location on the shelf in case replenishment is needed
Assign location of picking on module	Personnel in charge of waving process	SKU identification, quantity on modules and zones
Print orders sheet with days categorization colors	Personnel in charge of waving process	SKU identification, quantity, delivery date (to link to color coding of the day), final customer information
Assign replenishment route, forklift, quantities, and SKUs	Personnel in charge of replenishment process, forklifts, and forklift drivers	SKUs needed, quantities, location on the shelf, route assigned, priority
Scan the picked-up items by the forklift to change locations	Forklift driver/picker in the zone assigned, scanning device linked to WMS	SKUs needed, quantities, location on the shelf
Scan when the forklift arrives to conveyer feeder to assign module location	Forklift driver/modules feeder personnel, scanning device linked to WMS	module zone location
Assign pickers in different zones on the modules	Pickers in charge of completing the picking process	Demand level on the zone, pickers experience to assign to the right zones depending on demand level
Segregate items based on individual customer orders	Pickers in charge of completing the picking process	Sales orders sheets with SKUs information, labels
Ensure customer order is fulfilled and send to packing	Pickers in charge of completing the picking process	Sales orders sheets with SKUs information, labels
Pack items according to their labels	Packers in charge of completing the packing process	Final labels for shipment, packing specification

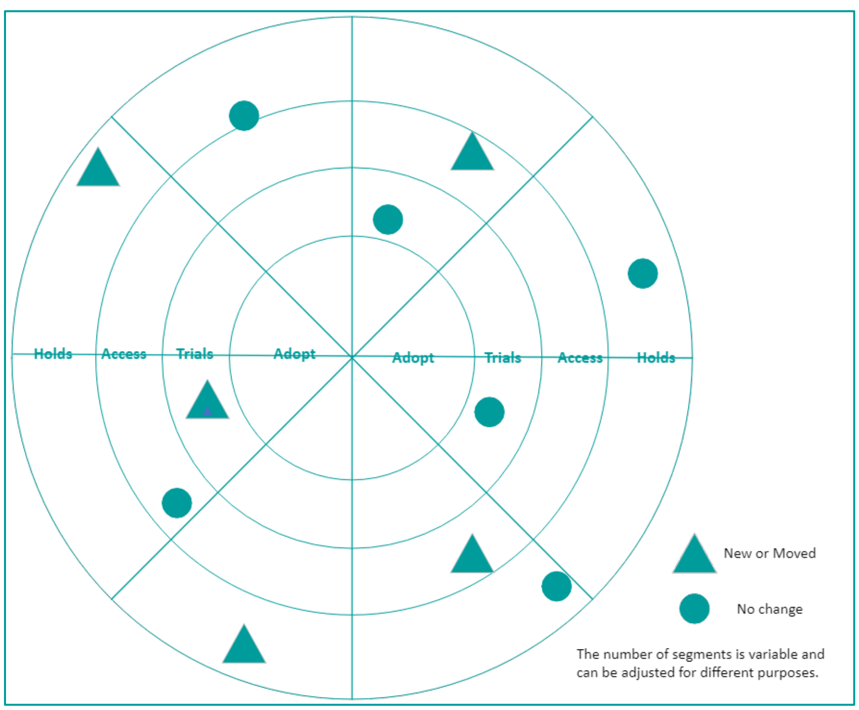
There is a wide range of technologies that can be used with varying effects on efficiency and productivity. Therefore, we needed to limit our work on certain areas to provide a more structured approach for this project. Hence, we resorted to process mapping of the warehouse activities to understand where we can focus.

2.3.2 Technology Radar

A technology radar is another useful tool. From general browsing of the internet and some articles from the MIT technology review, we gathered that the concept of technology radar was popularized by thoughtworks, a global software consultancy company. Thoughtworks started publishing their technology radar report in 2010, which serves as a regularly updated guide for businesses and technology leaders to make informed decisions about evolving and emerging technologies (Mugrage, 2022).

The thoughtworks Technology Radar is created by a group of senior technologists from the company, who gather inputs from their experiences, clients, and industry trends. The radar categorizes technologies into four groups: Techniques, Platforms, Tools, and Languages & Frameworks. Additionally, it places each technology into one of four rings that represent different levels of adoption recommendation: Adopt, Trial, Assess, and Hold (Mugrage, 2022). A simplified template has been built as per below Figure 2.

Figure 2
Technology Radar Template



Note: an adapted version of Technology Radar by Thoughtworks

The technology radar framework provides a guide for analyzing and evaluating different technology trends across four different categories (Smith, 2018):

1. Emerging Technologies: These are new and innovative technologies that are not yet fully developed or tested but have the potential to significantly impact the market.
2. High-Potential Technologies: These are technologies that have shown promise and are worth monitoring, but they may not yet be ready for widespread adoption.
3. Mature Technologies: These are established technologies that are widely used and have proven their value in the market.
4. Declining Technologies: These are technologies that are becoming obsolete or losing market share.

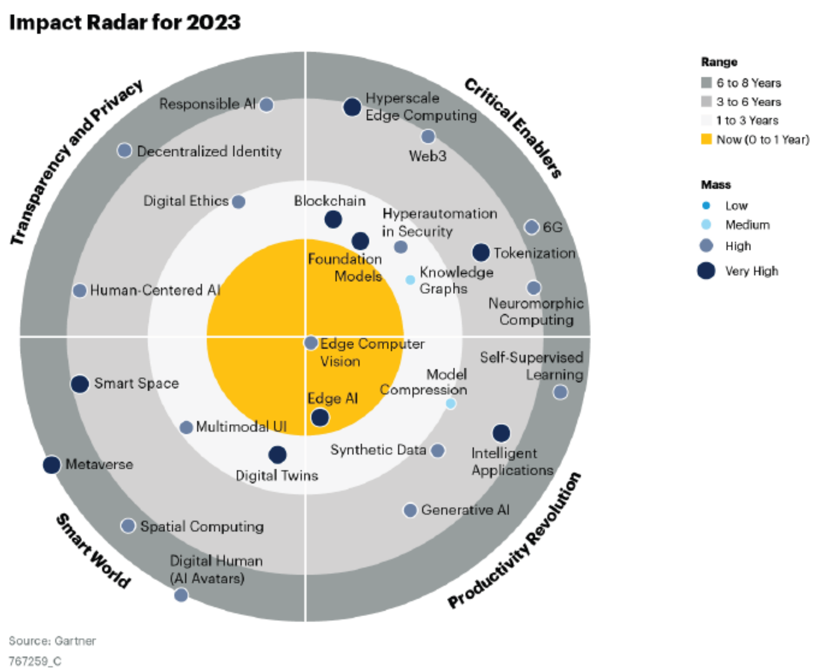
From a general approach, Nguyen et al. (2023) identify four key themes that define the most important emerging technologies for technology providers in 2023. These themes are the smart world, productivity revolution, transparency and privacy, and critical enablers. The report highlights specific emerging technologies within each theme, such as digital humans and the metaverse in the smart world, edge AI and synthetic data in the productivity revolution, digital ethics and responsible AI in transparency and privacy, and foundation models and neuromorphic computing in critical enablers.

The report provides recommendations for product leaders to capitalize on these emerging technologies, including investing in smart world foundations, exploring new tools to increase the value of AI applications, proactively investing in technologies that promote transparency and privacy, and assessing the impact of critical enablers on their product or service offerings (Nguyen et al. 2022).

Overall, the technology radar is a useful tool for organizations to stay up to date on emerging technologies and trends in the IT industry. By categorizing technologies into different categories and themes, it helps businesses to prioritize their investments in technology and stay ahead of the curve in a rapidly changing industry. The recommendations provided in the report can help product leaders make informed decisions about their technology investments and capitalize on market opportunities.

Below Figure 3 is the latest impact radar from Emerging Tech Impact Radar: 2023.

Figure 3
Emerging Tech Impact Radar: 2023



Note: Nguyen, T., Jump, A., & Casey, D. (2022). Emerging Tech Impact Radar: 2023

2.4 Warehouse Picking Process Technologies for Digital Twins

The picking process due to its manual nature and its involvement of multiple movements is one of the most labor-intensive processes. Therefore, if designed in an efficient way and adopted technology and automation, it could significantly improve productivity (Dallari et al., 2008). The picking process in the warehouse being studied in this project is done in a completely manual way. There are many technologies in the market today that can be used with varying levels of automation, cost, and productivity improvement levels. However, our focus will be on autonomous guided vehicles (AGVs), picking robots, and automated storage and retrieval systems (AS/RS), as we believe they can affect enhance performance in a better way for this warehouse. These three technologies are primarily focused on enhancing the picking process productivity and efficiency. They also vary in their complexity of adoption and cost, with AGVs being the simplest and lowest in cost and AS/RS being the most complex and highest in cost. In addition to improving the picking the process they will have a positive influence on the other activities in the warehouse, as they will be able to generate valuable data through their sensors while operating which can be used to improve the overall flow in the warehouse.

2.4.1 AGVs

Automated Guided Vehicles (AGVs) have become a popular technology in the supply chain network due to their ability to provide high accuracy, efficiency, and sustainable solutions. One important aspect of their implementation is their impact on warehousing design and day-to-day operations. One potential application of AGVs is in the replenishment process, which is currently done manually using forklifts. This process is heavily dependent on the forklift driver's capabilities and can lead to safety issues during peak times. AGVs, on the other hand, can be integrated with the warehouse management system (WMS) to ensure a continuous supply to the pickers' storage zones in real-time. They can also communicate with each other to avoid congestion or collisions and always ensure the use of the most optimized routes for the replenishment process (Cupek et al., 2020).

The replenishment is done through forklifts that require a driver and instructions to go pick up the SKUs from the shelves and put them in the picking zones close to the conveyors for the pickers to pick them when needed. The WMS checks if there is enough inventory next to the pickers and sends the instructions to replenish these shelves in case there is not. This manual process of replenishment is heavily dependent on the forklift drivers' capabilities, correct identification of needed items, and productivity. It could also lead to safety issues at peak times if the routes are not coordinated properly, whether it being congestion in heavily demanded zones that could delay the process, collisions in case of human error, or bad practices when reaching for items stored at higher levels on the shelves. AGVs that are connected to the network and the WMS will have the information in real-time. They will get their assigned routes and they can communicate with one another to avoid congestions or collisions. In addition to that they can ensure a continuous supply to the picker's storage zones (Cupek et al., 2020). The integration of AGVs with humans in a warehouse can lead to reduction of transportation time inside the facility and by its connectivity to the network they can always ensure the use of the most optimized routes to do the replenishment process (Yao et al., 2018).

AGVs have become increasingly popular in warehouse operations due to their ability to increase efficiency, reduce labor costs, and enhance safety. In this section, we will delve into the benefits of AGVs, the different types of AGVs, and the priority choices for the picking process.

Some benefits of AGVs and how it can enhance workload forecasting (Kumar et al., 2015), are as follows:

1. Increased efficiency: AGVs can work continuously without breaks, allowing for a more streamlined and efficient warehouse operation. They are also able to move at precise speeds, optimizing travel times and reducing idle times. This can result in significant productivity gains for the warehouse.
2. Reduced labor costs: AGVs can take over manual tasks, reducing the need for human intervention and associated labor costs. By automating repetitive tasks, AGVs can free up the workforce for more value-added tasks, leading to better overall workforce utilization.
3. Enhanced safety: Equipped with sensors such as laser scanners and cameras, AGVs can navigate the warehouse environment while avoiding collisions and minimizing the risk of accidents. This not only protects warehouse workers but also reduces the potential for damage to goods and infrastructure.
4. Improved inventory management: AGVs can help maintain accurate stock levels by scanning barcodes or QR codes and updating the inventory system. This real-time tracking of inventory

levels allows for better demand forecasting and reduces the likelihood of stockouts or overstock situations.

5. Optimized replenishment: AGVs can move goods from storage areas to picking locations, ensuring items are readily available for order fulfillment. This facilitates faster order processing and improves customer satisfaction.

Various types of AGVs cater to different needs within warehouse operations (Moshayedi et al., 2019). Moshayedi et al. highlight some of the common types:

1. Tugger AGVs: Tugger AGVs are designed to tow multiple carts or trailers in a train, making them ideal for high-volume material transport. They are commonly used in manufacturing and distribution centers.

2. Pallet AGVs: Pallet AGVs are designed to transport individual pallets, allowing for precise and accurate placement. They are commonly used in warehouses, distribution centers, and manufacturing plants.

3. Unit-load AGVs: Unit-load AGVs are designed to handle large and bulky materials, such as heavy machinery or large crates. They are commonly used in manufacturing and distribution centers.

4. Forklift AGVs: Forklift AGVs are designed to mimic the operation of a human-operated forklift, allowing them to lift and transport heavy loads. They are commonly used in manufacturing and distribution centers.

2.4.2 Picking Robots

After the material is replenished and placed on the smaller shelves next to the picker zones and the conveyor built, the pickers will have to take their orders information and walk through their assigned zones to pick the items in their orders one by one. This process involves the pickers walking back and forth within the zone and picking the items by hand or a cart. This is highly dependent on the picker's ability to memorize their items' locations to choose the most efficient route and how much they can carry or push in the cart. In addition to that, when the picker has many items, they resort to scanning them and updating their location when they go back to the conveyor built to place them in their respective boxes, which results in a delay in updating inventory location. A picking robot or a collaborative order picking robot (cobot) can increase the efficiency and productivity of the process by assisting humans in this process. The cobot will have the items information from the network and will have the ability to choose the best route and sequence to pick the items (Lambrechts et al., 2021). These robots can detect the humans around them and navigate through the route without affecting them. They are also able to provide better results when the items are heavier and reduce the risk of accidents in the workplace (Tutam, 2021). Additionally, they can continuously move with new orders information, which can shift the focus on the humans on the items' placements on the conveyor for orders fulfillment (Lambrechts et al., 2021).

Picking robots are the second technology we'd like to recommend. Picking robots are designed to automate the picking process by identifying and selecting items from warehouse shelves. By using sensors such as cameras and grippers, picking robots can capture data on product dimensions, weight, and location. This data can be used to optimize picking routes, reduce picking time, and prevent product damage. All in all, they are able to automate the picking process and improve overall efficiency. In this discussion, we will explore the benefits of picking robots, the different types available, and the priority choices for the picking process.

Some benefits of picking robots are as per below (Tutam, 2021):

1. Optimized picking routes: Picking robots use captured data to optimize routes, reducing picking time and increasing overall efficiency.

2. Reduced product damage: By using sensors and grippers, picking robots can handle items carefully, minimizing the risk of product damage.
3. Improved workforce productivity: Picking robots can take over repetitive tasks, freeing up the workforce for more value-added activities.
4. Enhanced accuracy: Picking robots can significantly reduce picking errors, leading to improved order fulfillment and customer satisfaction.

Some different Types of Picking Robots (Bormann et al., 2019):

1. Robotic Arm Pickers: These versatile robots are equipped with multi-axis robotic arms that can reach, grasp, and manipulate items with precision. With advanced gripping technologies and sensors, robotic arm pickers can adapt to different item shapes, sizes, and weights, making them suitable for a wide range of warehouse operations.
2. Piece Picking Robots: Designed specifically for handling individual items, these robots often employ a combination of computer vision, artificial intelligence, and advanced gripping mechanisms. They can identify and pick items from shelves with high accuracy, making them ideal for order fulfillment in warehouses with diverse product assortments.
3. Carton Picking Robots: Specializing in handling cartons or boxes, these robots typically use vacuum grippers or mechanical clamps to securely pick and move packaged goods. Carton picking robots can be employed in warehouses that deal with high volumes of boxed items, such as e-commerce fulfillment centers.
4. Collaborative Robots (Cobots): Designed to work alongside human operators, Cobots use advanced sensors and safety features to ensure that they can operate in close proximity to workers without posing a risk. Cobots can be used for tasks such as picking, packing, or sorting, and they can enhance human productivity by taking over repetitive or physically demanding tasks.

2.4.3 Automated Storage and Retrieval Systems (AS/RS)

Automated storage and retrieval systems (AS/RS), are systems that use a combination of equipment and controls to automatically handle, store, and retrieve materials such as components, tools, raw materials, and subassemblies with high speed and precision. They are commonly used in industrial settings to efficiently manage products and make the most of time, space, and equipment (Manzini et al., 2006).

AS/RS technologies typically utilize automated machines that move vertically through one or multiple storage levels to handle and store items. These technologies promote efficiency in space utilization, increase storage capacities, and aid in long-term planning. They can enhance inventory management and control processes, providing a flexible system that quickly locates, stores, and retrieves materials or products. The real-time information provided by these systems reduces inventory shortages and improves the rotation of stored items. They also offer versatility in storage design and can easily integrate with other inventory management technologies, such as automated guided vehicles. Additionally, they help minimize costs by reducing energy and heating expenses and limiting waste through automation (Kahraman et al., 2020).

AS/RS is the third technology recommendation. It's more expensive and might take longer time to achieve return on investment. The benefits of AS/RS, the various types available, and the priority choices for the picking process are presented as per below.

Some benefits of AS/RS are (Kahraman et al., 2020):

1. Optimized storage locations: AS/RS can analyze captured data to identify optimal storage locations, leading to more efficient use of warehouse space.
2. Reduced storage space requirements: By utilizing high-density storage solutions, AS/RS can significantly decrease the overall space required for storage.

3. Enhanced inventory control: AS/RS enables real-time monitoring of product location, quantity, and movement, ensuring accurate inventory management.
4. Increased productivity: Automating storage and retrieval processes reduces manual labor, allowing staff to focus on more value-added tasks.
5. Improved safety: By minimizing human intervention in storage and retrieval processes, AS/RS can help reduce workplace accidents.

Introducing some AS/RS types:

1. Unit Load AS/RS: These systems are designed to handle heavy and bulky loads such as pallets or large containers. They typically use a crane or similar mechanism to move items vertically and horizontally within the storage area. The crane is guided by a computerized control system, which ensures precise positioning of the items and maximizes storage density. Unit Load AS/RS can handle various load sizes and can be customized to accommodate specific warehouse requirements, making them suitable for a wide range of applications.
2. Mini Load AS/RS: Mini Load AS/RS are designed for managing smaller items or loads. They operate similarly to Unit Load AS/RS but are tailored to handle lighter items efficiently and accurately. These systems often use a combination of conveyor systems, robotic arms, and shuttle devices to transport and store items. Mini Load AS/RS are ideal for situations where a large number of small items need to be stored and retrieved quickly, such as in e-commerce fulfillment centers or parts distribution warehouses.
3. Vertical Lift Modules (VLMs): VLMs are enclosed storage systems that consist of two parallel columns of trays or shelves, with an automated lift mechanism in the center. The lift moves vertically to access the stored items and can bring them to an ergonomic workstation for retrieval. VLMs are designed to maximize vertical space usage while providing quick access to stored items.

They are particularly useful for storing small parts or items with a high turnover rate and can be easily integrated into existing warehouse workflows.

4. Horizontal and Vertical Carousels: Carousels are rotating storage systems that provide fast and efficient storage and retrieval of items. Horizontal Carousels consist of a series of bins or shelves mounted on an oval track, while Vertical Carousels consist of trays or shelves mounted on a vertical loop. Both types are motorized and can be operated using a computerized control system, allowing for quick access to stored items. Carousels are ideal for high turnover items or applications where the rapid retrieval of items is critical.

2.5 Predictive Analytics

2.5.1 Workforce Forecasting Through Digital Twins

Digital twin technology has emerged as a valuable tool in optimizing warehouse operations by leveraging real-time data to simulate and monitor warehouse processes. In section 2.4, we explored various enabling technologies for smart warehousing, including digital twins, and their application in improving efficiency, productivity, accuracy, and safety in warehouse operations. However, as the adoption of digital twin technology increases, it is essential to understand its impact on the workforce in warehouses. In this section, we will explore different warehouse workforce forecasting models, and further discuss the significance of warehouse workforce forecasting in digital twin implementation, highlighting the potential benefits and challenges in enhancing workforce productivity and job satisfaction while minimizing risks associated with automation and digitalization. By understanding the role of digital twins in warehouse operations and its impact on the workforce, we can develop strategies to effectively integrate this technology and optimize warehouse operations for maximum efficiency and effectiveness.

Workforce planning is one of the most important issues to resolve for the warehouse operations teams and its management. The warehouse suffers from the usual challenges in operation, like fluctuating demand causing sudden labor requirements to go up or down. In addition to that, there are other factors, such as location of the facility, as it is far away from other warehouses. This makes personnel sharing with other facilities impractical. Furthermore, new staff training and skilled staff loss are two other contributors. Skilled staff might be lost due to sudden low demand over a certain period, and new staff training normally requires several weeks (Van Gils et al., 2016).

The real underlying challenge is maintaining a certain level of expertise in each picking zone, and with good tactical forecasting of demand. Over forecasting of labor requirements in the warehouse will cause dramatically increased cost. On the other hand, under forecasting leads to reduced service levels as it leads to delays in orders.

Most literature on warehouse workforce forecast assume the sales orders are known, such as: (Van Gils et al. (2016) and Ho et al. (2022). This will be the same case in our study, as the customer, in our case, provides the sales forecasting which sets the expectations and can be used to measure service levels. Later in the modelling part, we intend to analyze the correlations between (1) daily sales data, in terms of historical actual demand and future forecast received from customer and (2) daily number of hours spent on the picking process in the warehouse.

We looked at different predictive modeling methods to be explored and we will go into details into a couple of them that are relevant for our project, time series forecasting, and machine learning.

2.5.2 Time Series Forecasting

Time series forecasting is a significant forecasting technique where a variable's past observations and a random error factor are used to predict it. It identifies historical trends and patterns and predicts future trends. Literature widely discusses exponential smoothing and seasonal autoregressive integrated moving average models (SARIMA) models. Time series forecasting is widely applied beyond warehousing, e.g., in water and energy consumption. There are various time series forecasting methods that differ in how they handle trends and seasonal patterns. These methods include the Naive method, moving average, exponential smoothing, SARIMA, and composite forecasting (CF), which combines other models (Van Gils et al., 2016). Taking the information of demand for past periods and workforce figures will allow us to use time series to build a model that can help predict future workforce requirements.

Van Gils et al. (2016) show in their study real cases where time series forecasting yielded positive results in improving workforce forecasting in the picking process. The study's results indicate that some of the forecasting models outperformed the baseline, which were the supervisor's non-statistical predictions. In warehouses, order picking zones are created to improve order picking efficiency, and order picker planning is done at the zone level. Hence, order forecasting needs to be broken down. A bottom-up forecasting approach generates more precise forecasts, unlike a top-down approach. The CF and exponential smoothing models have been shown to be particularly effective in forecasting orders at the zone level.

2.5.3 Machine Learning

Machine learning (ML), a subfield of Artificial Intelligence concerned with developing algorithms that can learn from data. This includes pattern recognition and computational learning theory. ML is widely used in supply chain management applications across different industries and in different areas, such as supplier selection, supplier segmentation, demand estimation, inventory management, and warehouse operations. ML techniques include inductive learning, artificial neural networks, case-based reasoning, support vector machines, and reinforcement learning. There are two main types of ML learning: supervised, where the data is structured and labeled, and unsupervised, where the data is unlabeled and used to find patterns. There are many ML algorithms, including Neural Networks, Support Vector

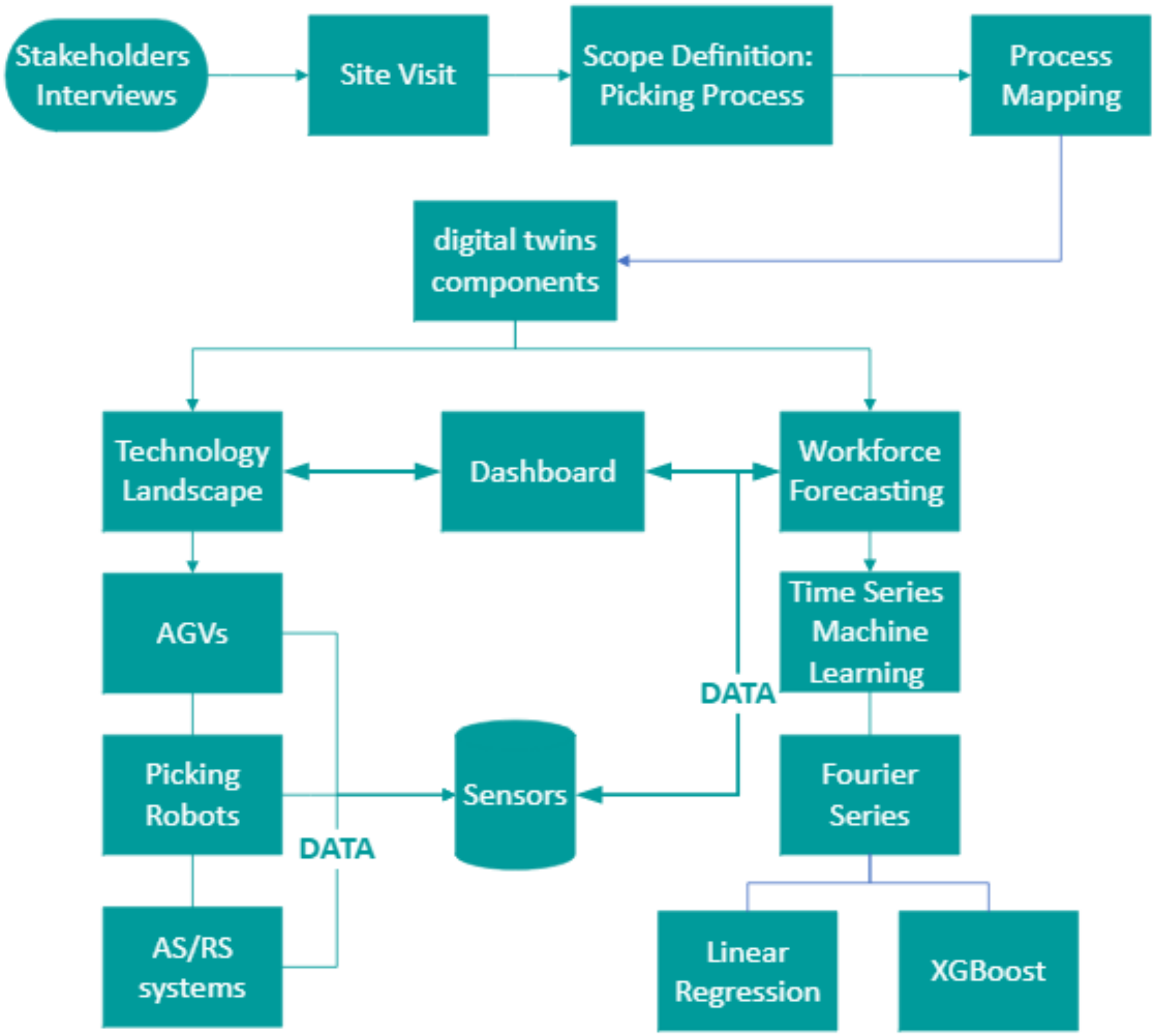
Machines, Regression, Decision Trees, Random Forests, Association Rule Learning, classifiers, and k-means algorithms (Ghaouta et al., 2021). The use of ML can be convenient when it comes to big data, and when it has several sources. The computer assisted model generated by using ML techniques will be data driven and fully automated, which can generate great efficiency in adjusting the workforce forecasting (Faloutsos et al., 2018).

A case study that trains machine learning models to predict the workload of a picking process in a warehouse uses a decision tree classifier to evaluate the significance of input variables. The results indicate that the volume, weight, popularity, and seasonality of SKUs play a crucial role in making accurate predictions (Ghaouta et al., 2021).

3 Data and Methodology

In this chapter we describe the steps that we took to tackle the workforce forecasting problem, along with the three technologies that we need to implement to achieve an improvement in the operational capabilities of efficiency, productivity, and scalability. In addition to that in Figure 5 we show the framework for the project and overview of the steps that we took. At the end of the chapter, we present the digital twin layers that tie together the workforce forecasting model with the technologies.

Figure 4
Project Framework



3.1 Stakeholders Interviews

To understand the warehouse processes and the flow of operations, we conducted a series of interviews with the warehouse operations team. The interviews focused on getting a complete picture, starting from when material arrives to the warehouse in containers until it leaves to its final destination on trucks. The activities covered in these interviews can be summarized in the list below:

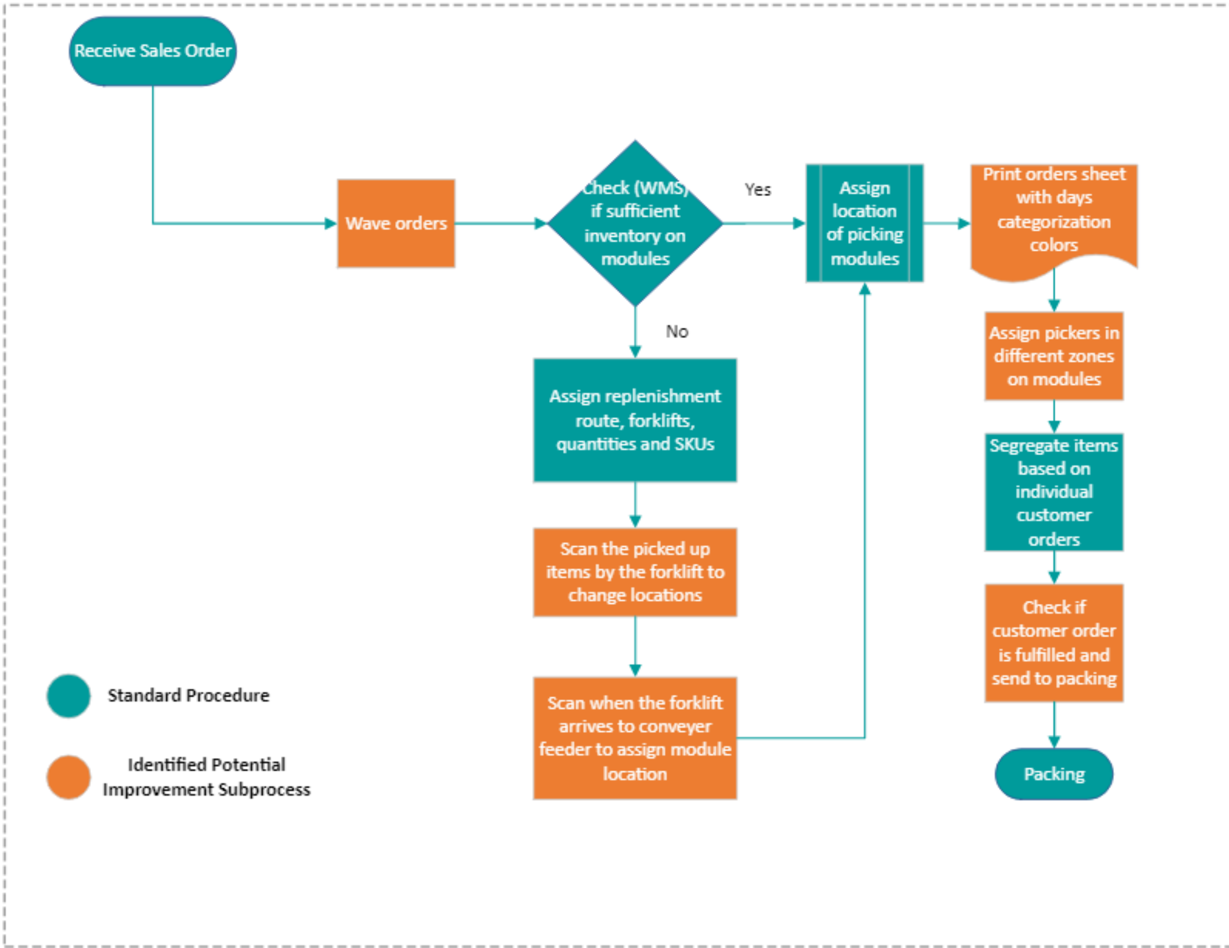
1. Receiving
2. Storing inventory
3. Replenishment of the conveyor belts & picking
4. Packing
5. Consolidating
6. Shipping

The interviews revealed that the replenishment and picking operation is where we can focus our work. This process involved a lot of manual work from the moment the sales orders are received into the WMS until the SKUs/orders reach the packing stage. Furthermore, the warehouse operations team confirmed that this where they would like to see more efficiencies and improvements in productivity, as this is the process that mainly contributes to the workload and workforce forecasting problem.

3.2 Process Mapping

In order to have a better understanding of the current conditions of the picking activities, we started with mapping the processes in the warehouse to deepen our understanding and help pinpoint the processes that require our focus. We, once again, resorted to interviews with the operations team of the distribution center to get that understanding. Based on the outcome of these interviews we were able to draw Figure 6 for the picking activities and highlighted the sub-processes that can benefit from digital twin technologies and a predictive forecasting model. Figure 6 shows the As-Is process mapping with the identified sub-processes for improvement. The sub-processes in orange color are the ones that we identified and believed will benefit from a digital twin solution to generate an improvement in efficiency and productivity, as they are done today in a manual manner.

Figure 5
Picking Process Mapping - AS IS with Sub-processes Identified for Improvement



3.3 Data Gathering

Gathering existing real-time data from the different systems is a key challenge, as the systems are not linked, data is not unified, and each system controls a specific type of operation. The type of data is also another challenge for this project since every system measures different outputs in the operation cycle. Besides, we also need to consider the virtual data, as Zhang et al. (2022) suggest, good digital twins' data should be a combination of data from both the physical world and virtual scenarios. As mentioned before, for the sake of limiting the scope we will be focusing on the picking related activities to have a specific process where we will be diving in depth into understanding and analyzing it.

Current data acquisition and preparation steps are summarized as per below Table 2

Table 2
Data Gathering & Preparation

Data Gathering & Preparation	Details
Historical sales data	2021-2022 daily sales
Labor management system (LMS) data	2021-2022 daily labor hours
Data preparation	Warehouse team assistance to clean the data

3.3.1 Historical Sales Data

The distribution center has one major customer that occupies around 90% of the capacity and workload. Therefore, we focused our data gathering on that customer. The customer of our sponsor company is an American shoe manufacturer that sells their products through different channels. The shoes arrive to the distribution center, then depending on the type of orders they are sent to different locations across the US. The facility handles business to business (B2B) and business to consumer (B2C) orders. B2B orders go to other distribution centers owned by retailers or directly to retailers. B2C orders through different e-commerce platforms, such as Amazon or Zappos, end up going directly to the final consumer.

The customer sends the company's distribution center a monthly forecast of orders they expect to receive from their B2B and B2C channels. This forecast is fully controlled by the customer, and it only shows a monthly quantity for B2B and B2C, with no weekly or daily split. In a warehouse operation, the forecast of workload and workforce is done on a daily frequency. Therefore, the company's warehouse operation team had to rely on their experience and historical information to create a forecast for their operational needs, in terms of people and equipment.

For the use of our predictive model, we needed to look at the historical information to understand the behavior. The distribution center team shared with us the actual sales information for 2021 and 2022. The datasets that we received included a lot of information, however, the following features were the ones that were important for our purpose:

A. Customer information
B. Order type
C. Order reception date
D. Order shipment date
E. Quantity

3.3.2 Data from Labor Management System (LMS)

To create a predictive model for workforce forecasting, we needed to study the historical information for daily labor hours. We asked the company for data that matches the time horizon we received for the sales. We received the daily hours punched into the labor management system by each employee, whether it is a permanent employee or a temporary one. As much as the sales information is important, it is going to be used an input into our model. We assumed that the sales forecast is a feature that the model used to predict the daily hours. The daily hours information is the most crucial for the workforce forecasting model. The data we received included, employees' daily hours logged into the system for the years 2021 and 2022.

3.3.3 Data Preparation

The warehouse operations team was our focal point to clarify, understand, and help us in the process of cleaning the data. With the help of the team, we were able to simplify the datasets to focus only the data points that are needed for the model and create simplified versions of both the sales data, and the labor data. As we mentioned in 3.3.1, our model needed specific features for the purpose of the project, which meant that we had to remove all unnecessary features to the analysis.

3.4 Technologies Implementation

Our model has workforce as its target variable and primarily considers sales orders as a feature. A comprehensive workforce forecasting model usually considers a range of factors to address efficiency, productivity, and external variables that may impact workforce requirements.

Some of these key factors include staff turnover rates, which can help predict fluctuations in workforce size; employee skill levels, which affect overall performance and productivity; historical data on employees’ skills, which provides insights into potential skill gaps or surpluses; and business objectives such as the company's growth plans, which can influence workforce expansion or contraction. Additionally, the model should consider factors like product and service offerings, as well as potential market developments that may impact staffing needs.

By integrating these diverse data points into the model, it becomes more responsive to the intricacies of the workforce and the ever-changing business environment. This in turn leads to more accurate workforce forecasting, enabling better resource allocation, and facilitating more informed decision-making for warehouse managers and other key stakeholders.

To restate the digital twin concept for our project, it involves creating virtual replicas of the warehouse and all associated activities. This encompasses the integration of current and future technologies, real-time data collection from IoT devices, sensors, and software systems like WMS and LMS. Data analytics serve to provide valuable business insights.

As time progresses, continuous data collection feeds into various models, such as the workforce forecasting model in this case. The model learns from the newly acquired information, interacts with both humans and other IoT devices, and strives to continuously improve forecast accuracy. This, in turn, supports warehouse managers in making informed decisions.

By leveraging the digital twin, the warehouse environment becomes more dynamic and adaptive, enabling better management of resources, optimization of processes, and anticipation of potential challenges. Additionally, it provides a platform for testing and validating new strategies and solutions before implementation, reducing risks, and enhancing overall operational efficiency.

The use of sensors is one of the most effective ways to capture different types of data that can improve warehouse workload forecasting models. Sensors can be integrated with various warehouse technologies to provide accurate and reliable data on different aspects of warehouse operations. Below identified technologies are all incorporated with sensors.

3.4.1 Sensors

Sensors play a crucial role in modern warehouse technologies, as they capture data that can be used to optimize operations and increase efficiency. Some common sensors used in the warehouse industry that we found through searching different technology providers:

Table 3
Widely Used Sensor Types in the Industry

Sensor Type	Functionality	Application	Enhanced Capability
Barcode Scanners	Read barcodes on products or packages to identify, track, and record their movement within the warehouse	AGVs Picking Robots AS/RS	Efficiency Productivity
RFID Readers	Use radio waves to communicate with RFID tags attached to products, pallets, or containers	AS/RS	Producivity
Laser Scanners	Emit laser beams to measure distances and create maps of the warehouse environment	AGVs Picking Robots	Efficiency Productivity
Cameras	Capture visual data to monitor warehouse activities, identify products, and track their movement	AGVs Picking Robots	Efficiency Productivity
Ultrasonic Sensors	Measure distances by emitting sound waves and detecting the time it takes for the echo to return	AGVs Picking Robots	Efficiency Productivity
Inertial Measurement Units (IMUs)	Consist of accelerometers and gyroscopes that measure acceleration and rotation rates	AGVs Picking Robots	Efficiency Productivity
Force/Torque Sensors	Measure the force or torque applied to an object, enabling robots to handle items with appropriate force and prevent damage	Picking Robots	Producivity
Load Cells	Measure weight or force	AGVs Picking Robots AS/RS	Efficiency Productivity

These sensors are mature and widely used in various warehouse technologies, providing critical data for efficient and optimized operations in AS/RS systems, AGVs, and picking robots. By integrating data from these sensors with warehouse management systems, warehouse operators can improve inventory management, optimize routing and storage, and enhance overall operational efficiency.

3.4.2 Data Integration from Different Sensors and Systems

Data integration is a crucial aspect of data collection within a warehouse and plays a significant role in building warehouse workload forecast models. Data integration involves combining data from multiple sources to generate more comprehensive and accurate information, which can lead to better decision-making and improved warehouse operations. From our work with the company's warehouse, we see the importance of data integration in the following avenues that are presented in Table 4.

Table 4
Data Integration Benefits, How It Works & Results

Data Integration Benefit	How It Works	Result
Enhanced accuracy and reliability	Helps to minimize errors and inconsistencies resulting from individual data sources by combining and cross-referencing information from multiple sensors and systems	More reliable and accurate data, which is essential for building robust workload forecast models
Comprehensive view of warehouse operations	Merges data from various sources, such as WMS, LMS, and sensor technologies, data integration provides a more holistic view of the warehouse operations	Enables decision-makers to identify trends, patterns, and potential bottlenecks, thus facilitating better workload forecasting and resource allocation
Real-time decision-making	Allows for real-time data processing and analysis, which is crucial for making timely decisions in a dynamic warehouse environment	Enables warehouse managers to quickly respond to changing conditions and optimize resources, accordingly, leading to more efficient operations and better forecasting
Improved scalability	Enables warehouse managers to easily incorporate new data sources, sensors, or technologies into the existing infrastructure	Can handle larger data sets more effectively, ensuring that workload forecasting models remain relevant and useful even as the warehouse expands
Enhanced adaptability	Enables warehouse managers to easily incorporate new data sources, sensors, or technologies into the existing infrastructure	Keep up with technological advancements and ensuring that workload forecasting models are up-to-date and accurate
Cost savings	Helps warehouse managers optimize labor and equipment utilization by providing more accurate and reliable workload forecasts	Reduce operational costs and increasing overall efficiency

3.5 Data Analysis

Building on the technologies that can be integrated into the warehouse, another piece of the puzzle is the predictive model that can use the data generated by the technologies, as well as demand and labor information from the past to forecast the future. Before being able to forecast we needed to analyze the data provided by the warehouse team.

We started gathering the data and cleaning it, then we moved to taking a closer look to understand the different elements that could affect workforce variability. Looking at the history allowed us to see the trends and seasonality elements that affect the forecast.

3.5.1 Demand and Hours of Labor Investigation

We started our analysis by diving deeper into the historical demand and the daily hours and plotting them. To do that we needed to group the orders by day to see total order quantities received on a specific day (Drop Date column) and plotted them over the span of the 2 years being analyzed, as shown in Figure 7. We also followed the same methodology for the daily hours. We grouped the total hours on a certain day from all the employees, as shown in Figure 8. The plots for these 2 variables showed us that there are days with very low hours and orders, and in some cases zero hours and orders. The data led to the conclusion that there are days where the distribution center had low operations or no operation at all.

Figure 6
Daily Orders for 2021-2022

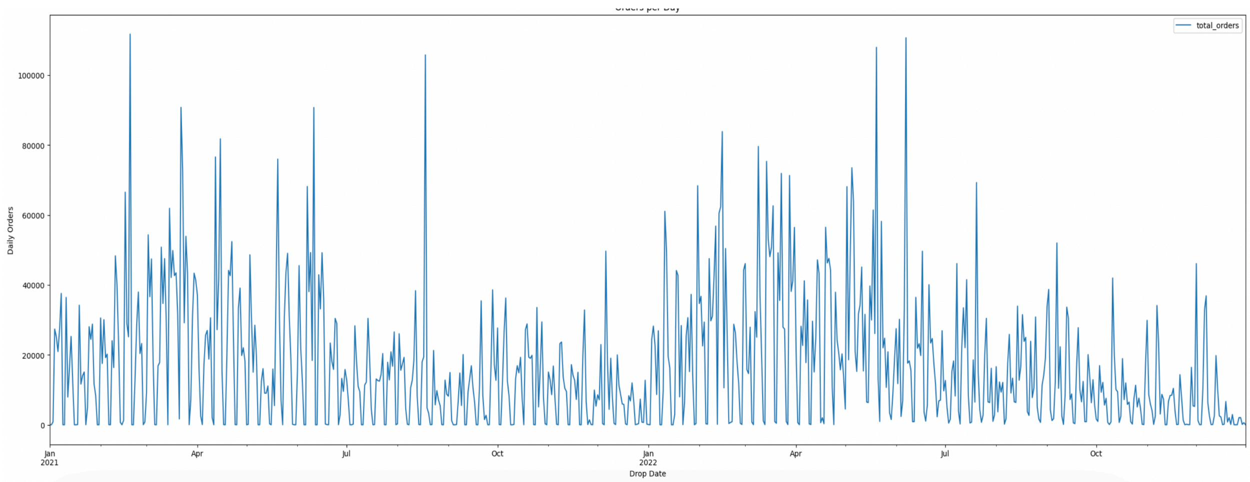
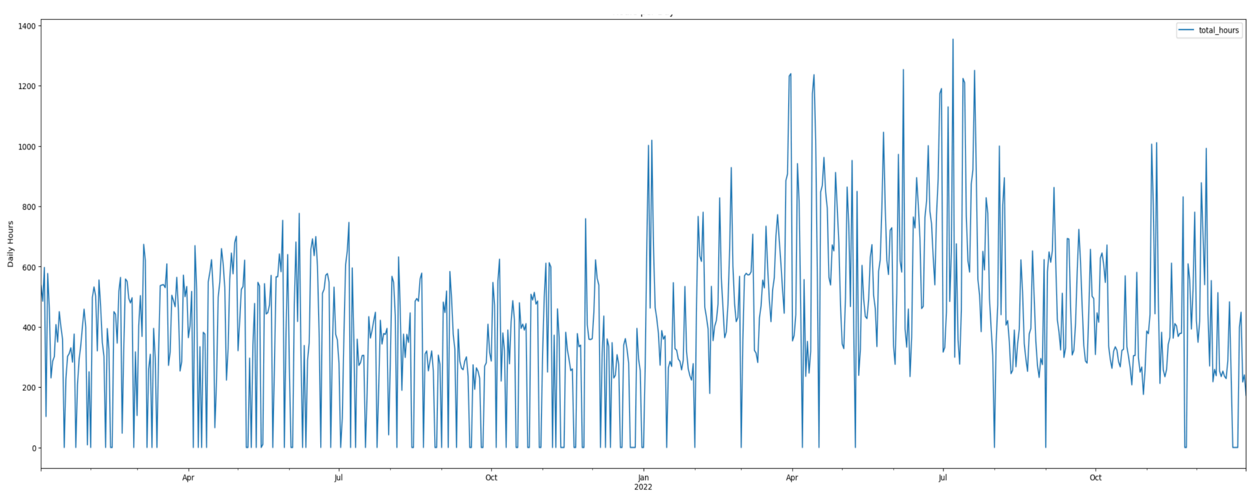


Figure 7
Daily Hours for 2021-2022



Furthermore, Figures 7 & 8 showed some high peaks in operations that needed to be investigated further. The peaks of daily labor hours were consistent with the periods of high demand. We needed to understand these peaks, as they affect the workforce forecasting dramatically.

Next, we obtained some statistics on the demand data and the orders behavior. Understanding the time window between orders reception to orders shipment and their quantities can help in segmenting the data in the right way to study its behavior. From the numbers shown in Table 5 we noticed that the median of the quantity of orders is 1 pair, and the median for the time window from order reception to order shipment is 3 days. The information led us to believe that the segregation of B2C and B2B orders could help us explain some of these peaks. The company confirmed that for B2C they have a strict KPI of 3 days from the day the order is received until it is shipped, which that matched this statistical data.

Table 5
Statistical Data for Orders and Duration

Attribute	count	mean	std	min	25%	50%	75%	max
Quantity	8,033,627	2.2	3.2	1	1	1	2	98
Order to Customer Duration	8,033,627	6 days	9 days	0 days	1 days	3 days	8 days	168 days

From Table 5 statistics it was obvious that B2B orders had a longer time window from the day the order is received until it is shipped, as B2B orders were mostly going to other distribution centers for retailers and their quantities were much bigger than B2C orders.

3.5.2 Trend and Seasonality Investigation

To go further into the details, we needed to plot the trend and seasonality of the demand and hours of labor. However, before doing so, we decided to add a feature for holidays. This feature is added to the total orders and total hours by day to see if the peaks of demand and hours can be explained by the holiday effect. Most retailers launch big sale campaigns before certain holidays, such as Black Friday and Christmas. The holidays feature was added in the forecasting methods used.

We continued our work by plotting the seasonality components of both the demand and the hours of labor to see the seasonal patterns and try to understand the behavior from that angle. The periodogram of the orders shows high biweekly frequency and an annual one (see Figure 9).

Figure 8
Periodogram for Daily Orders

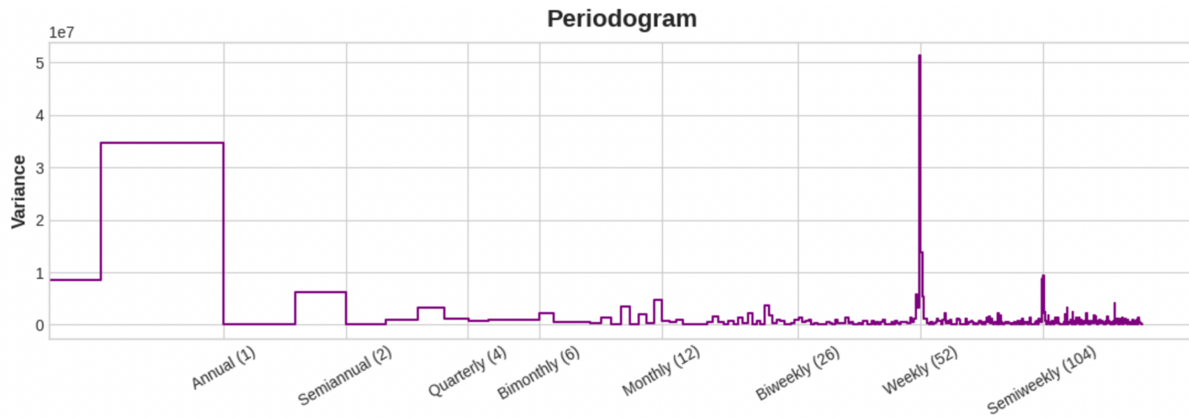
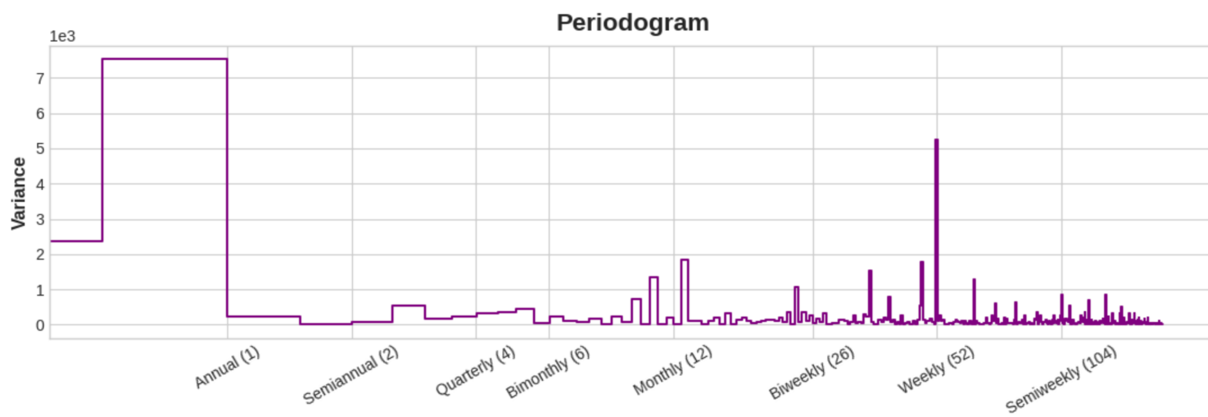


Figure 9
Periodogram for Daily Hours



However, the seasonality plot of the hours showed the same patterns, in addition to a bimonthly pattern, as shown in Figure 10.

3.6 Workforce Forecasting Models

Since the problem at hand is to forecast workload/workforce for the warehouse operations with a focus on the picking process, we believe that a predictive model would be the best way to approach it. After looking into the different techniques available today, we investigated time series forecasting and machine learning as discussed in Chapter 2 of State of the Art section 2.5 Predictive Analytics.

3.6.1 Extreme Gradient Boosting (XGBOOST)

The time series model for this project used demand information as an input to predict the hours of labor. Additionally, we used the historical hours of labor information for the model to train and predict, which means that any irregularity in the demand and lack of accuracy in the hours of labor would contribute to a poor model behavior. Therefore, we needed to pay attention to those parameters to get solid prediction results.

We started the modeling process by using a simple Naïve forecast model, as it helped us set a baseline of how our other models should perform. We then moved to using a machine learning model using extreme gradient boosting (XGBOOST). XGBOOST is a popular and effective machine learning algorithm for regression and classification tasks. It is a type of ensemble learning that combines the outputs of multiple weak decision trees to create a stronger, more accurate model. XGBOOST uses gradient boosting, a technique that iteratively trains decision trees on the residuals of the previous iteration, with a focus on reducing the error of the model. XGBOOST is highly scalable, as it can handle large datasets and parallel computation, and is known for its ability to achieve high accuracy and efficiency in various domains such as finance, healthcare, and e-commerce. XGBOOST has become one of the most widely used machine learning libraries due to its accuracy, speed, and flexibility (Chen and Guestrin, 2016).

3.6.2 Fourier Series

After using XGBOOST we wanted to take a step back and look at the data from a different angle. Another method that can help in preprocessing the data is Fourier series, as it can help us identify the patterns and trends in the data. Then we moved to applying Linear Regression and XGBOOST to get better results from the de-seasonalized data.

Fourier series is a mathematical tool that can be used in machine learning time series forecasting to decompose a time series signal into its underlying frequencies and amplitudes (Box & Jenkins, 1976). This approach can be applied in warehouse operations to predict patterns and trends in inventory levels, demand, and shipping volume. By analyzing the frequency components of a time series, one can predict future values of the signal and optimize warehouse operations to meet customer demands and reduce costs.

In a study by Zhang et al. (2020), Fourier analysis was used to decompose warehouse demand signals into frequency components, and machine learning models were trained on these components to forecast future demand. The results showed that the Fourier-based approach outperformed traditional time series models in terms of forecasting accuracy, indicating the usefulness of Fourier analysis in machine learning time series forecasting for warehouse operations. This approach can be extended to other warehouse operations such as inventory control, transportation management, and labor scheduling to improve efficiency and reduce costs.

3.7 Project Framework & Real Time Integration

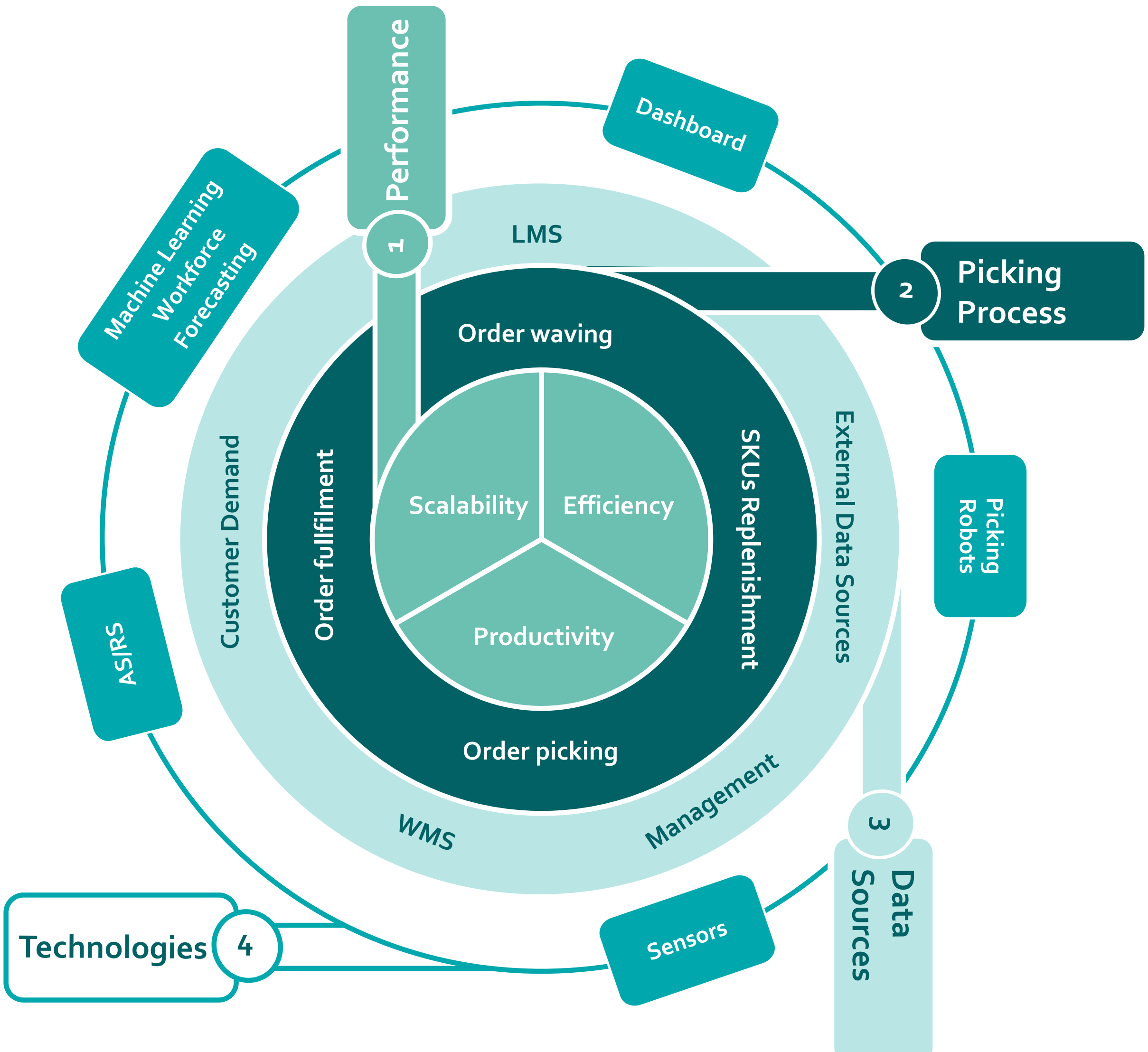
Technologies like AGVs, picking robots, AS/RS, and sensors generate a wealth of real-time data as they operate in the warehouse, such as their location, task completion rates, and performance metrics. By integrating this data into the machine learning forecasting model, the model can be updated in real-time to reflect the actual operational conditions of the warehouse. This allows for more accurate and dynamic predictions of workforce requirements, enabling efficient allocation of resources based on the current needs of the warehouse.

These technologies can be used to automate various tasks in the warehouse, and their performance can be monitored using sensors and other monitoring devices. For example, the system, through the data collected, can identify bottlenecks in the warehouse, such as areas with high demand but limited AGV or robot availability, and trigger adjustments in task allocation or AGV routes to address the issue. This helps to streamline warehouse operations, reduce labor costs, and improve overall efficiency and productivity.

Therefore, we saw the need to have a framework for the picking process digital twin prototype. The framework is generated based on the understanding of the picking process with the focus on the 3 main performance enhancement capabilities, efficiency, productivity, and scalability. In Figure 11 we can see four layers that represent the framework for the digital twin prototype. The four layers and the levers can be described as follows from the inside to the outside:

1. Layer 1: The core three performance enhancement capabilities that are going to be affected by the two outer layers.
2. Layer 2: Picking process, which is the process in the warehouse that we focused our work on.
3. Layer 3: Sources of data. The picking process is affected by the technologies existing in the warehouse and the management team's daily decisions.
4. Layer 4: Technologies proposed in this report.
5. Levers: The different levers can be changed to scale the framework to different performance metrics, processes, data sources, technologies, or other warehouses.

Figure 10
Picking Process Digital Twin Framework



Layer 3 includes data taken from WMS, LMS, customer demand, external data, and management's decisions on the daily operation. The data from Layer 3 is used in Layer 4 to improve the performance of the different technologies. Examples of these data sources and their details, along with their affected KPIs can be seen in Table 6.

Table 6
Sources of Data & KPIs affecting Performance Enhancement Capabilities

Data		
Sources	Details	KPI
WMS	Order lines	1. Throughput (qty/hr) (P) 2. Number of staff/day (E) 3. Experience level on the day (P) 4. Cost to serve (E) 5. Frequency of items ordered (Velocity) (E) 6. Picking cycle time (P) 7. Time to pick/picker (P) 8. Picking errors/complete correct orders rate (P)
LMS	Number of employees in every process	
Customer Demand	SKUs per day and demand trends	
WMS	Order lines	
LMS	Employees distribution over processes	
Customer Demand	B2B & B2C Split	
Management Changes	Focus of the shift/customer priorities/Order fill rate	
Turn-over Rate	Experience level of staff on the day in the assigned process	
Systems in the warehouse	Data flow between the different systems	1. Delta on workforce forecast accuracy (S) 2. Data quality (S) 3. Improvement of KPIs (S) 4. Key Learning Indicators (ex. Net Promoter Score (NPS)) (S)

(E) Efficiency KPI, (P) Productivity KPI, (S) Scalability KPI

Layers 3 and 4 are in constant communication with each other. The technologies proposed use data from the existing sources to create improvements in the operation. The improvements will be seen through the better workforce forecasting that will be achieved.

3.7.1 Example Scenario

To illustrate the framework and the interactions between the different elements the following example scenario can be used:

1. The machine learning workforce forecasting model generates a 1-week forecast based on the historical data for the orders and labor hours for the picking process labor hours with a 40% mean absolute percentage error (MAPE).
2. The AGVs and picking robots through their sensors are constantly measuring:
    a. AGVs: best replenishment routes to reduce replenishment time to the picking zones to improve efficiency.
    b. Picking Robots: best picking route for the list of orders assigned to its zone with its assigned picker to improve picking productivity.
3. The data from the AGVs and the picking robots are used as input features to the forecasting model for better productivity and efficiency measurement.

4. The external data such as experience level of the staff for the period is used as an input into the model.
5. The machine learning workforce forecasting model generates another forecast with an accuracy of 20% and reports it on the dashboard.
6. The dashboard shows the model output among other KPIs and those contribute to the actionable recommendations that this framework generates.
7. Based on the management's input and operational factors, scenario planning is done for future cases to better predict workforce changes and equipment utilization.

This scenario is one of many scenarios where the warehouse team can use the technologies to improve the efficiency and productivity. Once the data quality from the different systems reaches a mature level, this prototype can be scaled up to other processes in the warehouse and potentially other warehouses.

In Table 6 we see many KPIs that exist today (KPIs 1, 2, 5-8) in the warehouse as performance metrics, however, the addition here is the improvement of these KPIs and also the improvement in the accuracy of measuring them. We also added some new KPIs that can be tracked as they will be beneficial with this framework, such as KPI 3 in Table 11 for productivity, and all the scalability KPIs. The framework in Figure 11 can then be scaled up once the data quality reaches a mature level, and the different systems and technologies are working together in harmony to improve. Scaling up this framework can be done in many ways, whether through adding more technologies to the same process, or expanding to other processes and eventually other warehouses.

4 Results

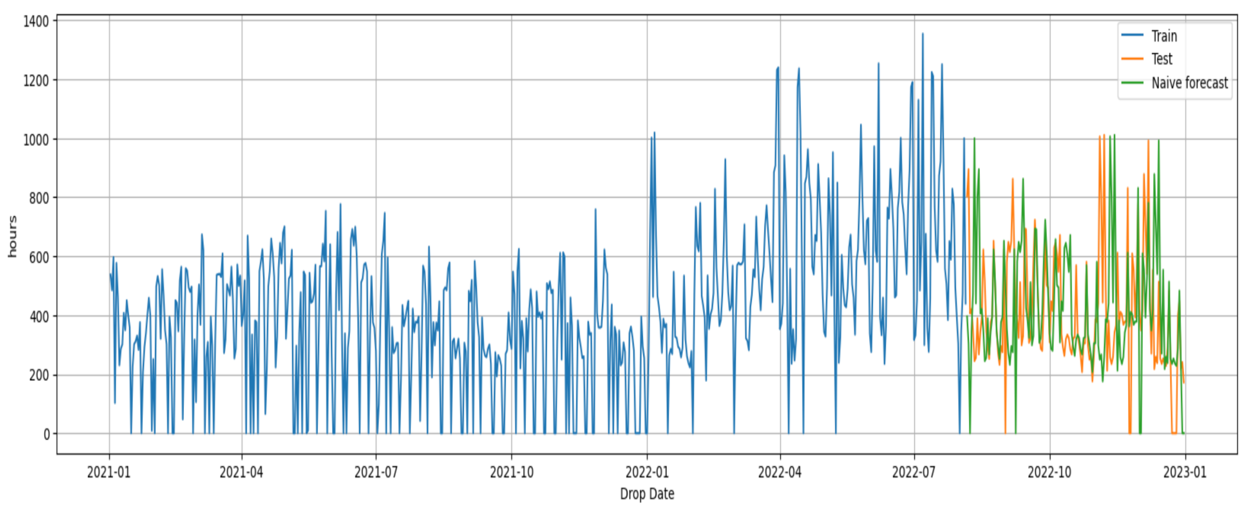
The different forecasting models showed varying results and helped in getting a much deeper understanding, as well as raised many questions. In this chapter we show the details of the results of the machine learning techniques for the forecasting problem, as well as the dashboard that connects the workforce forecasting model and creates a digital twin prototype.

4.1 Machine Learning Workforce Forecasting Results

4.1.1 Naïve Forecast

The main use of a naïve forecast in machine learning is to establish a baseline performance for comparison with more complex models. By using a simple and easy-to-implement approach, such as a naïve forecast, as a baseline, we can evaluate the performance of more advanced models and assess whether their predictions are better than the basic naïve forecast. Figure 12 below shows a naïve forecast with an 80/20 test/train split and maintaining a full week in both splits.

Figure 11
Naïve Forecast with 80/20 Split



	Method	SMAPE	RMSE	MAPE
0	Naive method	51.71	268.54	inf

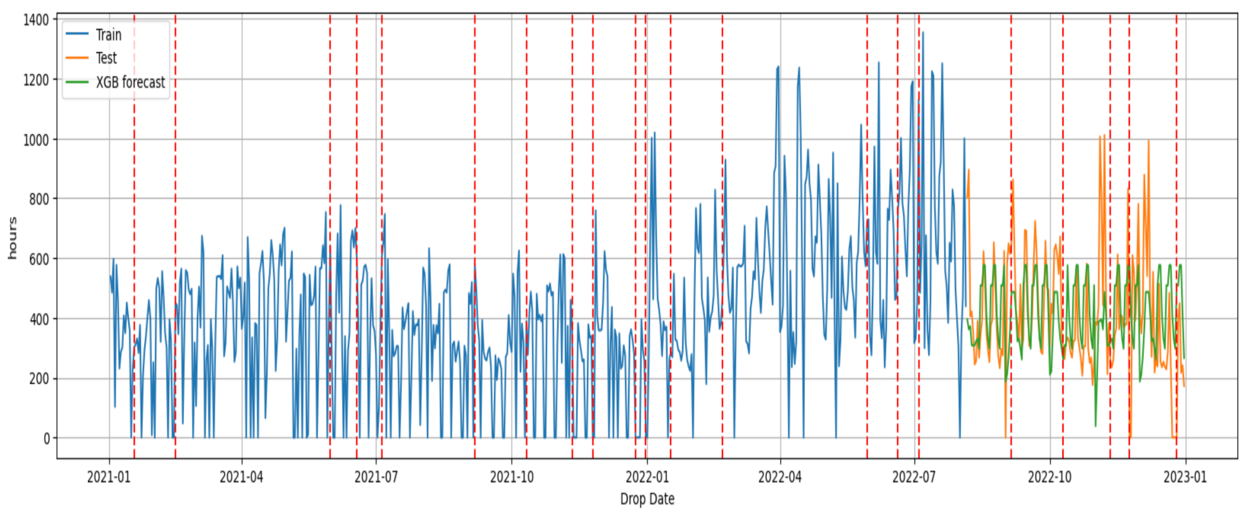
The high root mean squared error (RMSE) of 268.54 hours indicates that the model had a low forecast accuracy even with a naïve forecast. If we take a look at the testing data, we see that the highest daily hours are around 1000 hours/day, and they do not happen often, therefore, an RMSE value of 268.54

hours shows a high fluctuation in the forecast, which is not a good sign before we proceeded with XGBOOST. In addition, the "inf" mean absolute percentage error (MAPE) confirmed that as well.

4.1.2 XGBOOST

Before running XGBOOST we wanted to add the holidays to the graph to understand whether the peaks are happening around the same time as the holidays. We used XGBOOST with the same train/test split used for the naïve forecast. Figure 13 demonstrates poor performance for the model. As can be seen in the plot, the model is not able to capture the peaks in the data, which can also be confirmed by the high RMSE and MAPE. In addition to that, the holidays did not help in explaining any of the peaks, as some of the highest peaks did not happen around any holidays, such as the peaks in April 2022.

Figure 12
XGBOOST with Holidays (MAPE in %)



	Method	RMSE	MAPE
0	XGBOOST	18961.39	94.0

4.1.3 Fourier Series with Linear Regression and XGBOOST

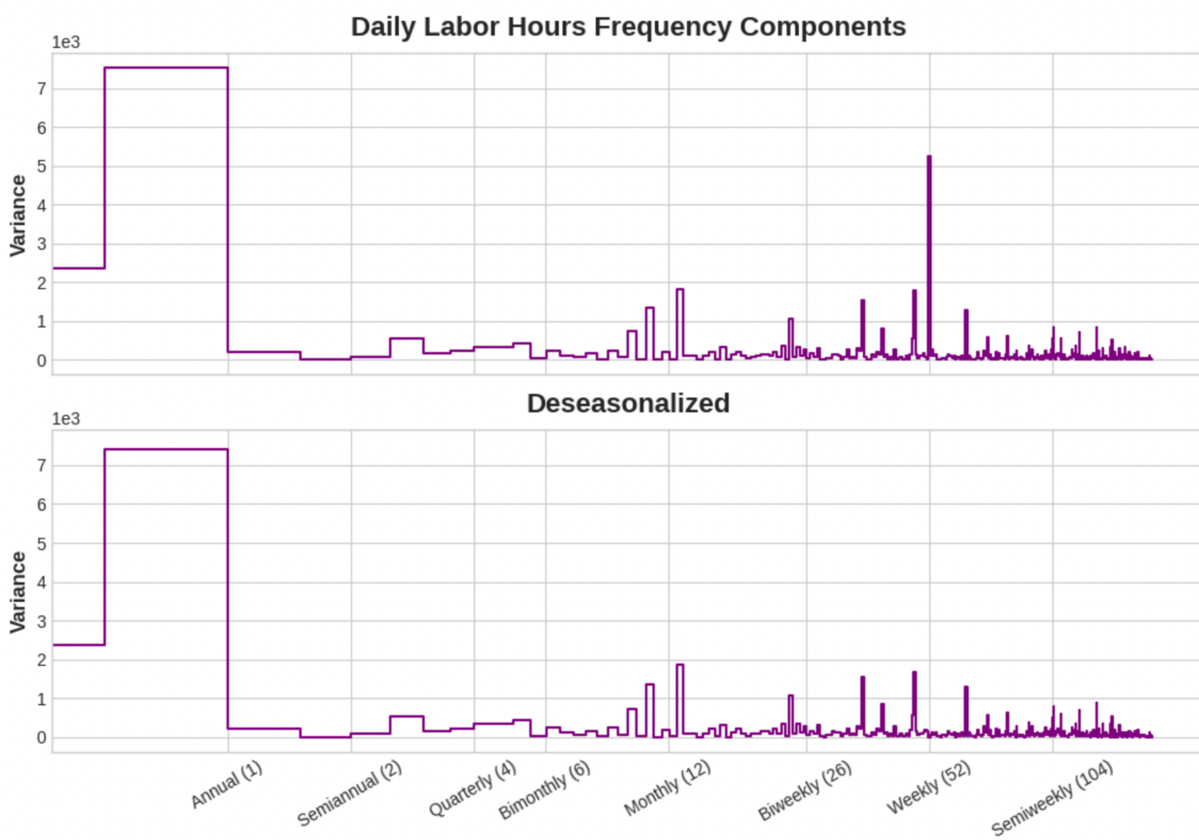
We used Fourier series in two ways, with daily labor hours and with weekly labor hours. We ran the data through the Fourier function and plotted the periodogram and partial auto-correlation function to see the effect. Then moved to seeing how a simple Linear Regression model would perform, and lastly used

XGBOOST. However, within every model we tried different iterations of lagging features, such as orders or hours. We also changed the test/train split to see how the model is affected. In this section we demonstrate some of these iterations.

4.1.3.1 Daily Labor Hours

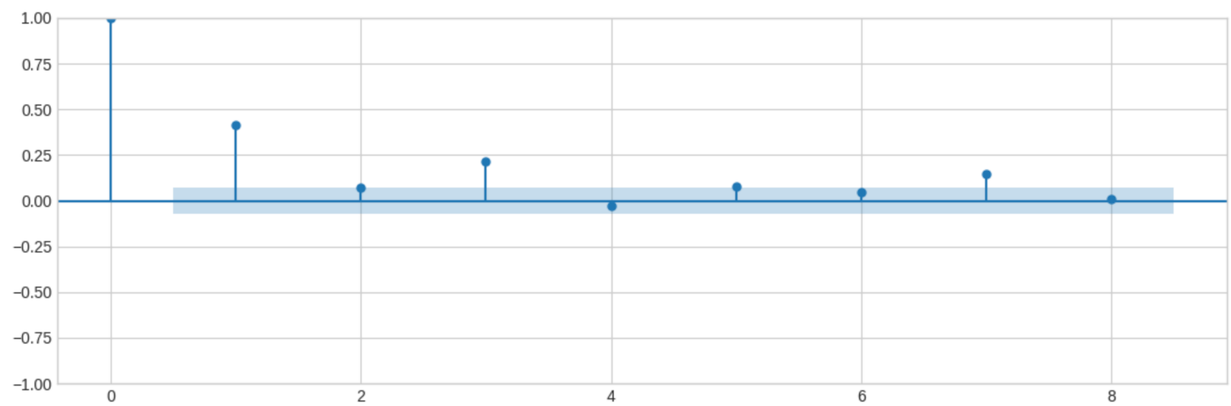
We started with a Fourier series function with a weekly frequency and an order of 7. We ran the periodogram again to see what a scenario without seasonality would look like. As seen in Figure 14, Fourier series was able to decompose the strong weekly seasonality in the data.

Figure 13
Periodogram Before and After Using Fourier Series



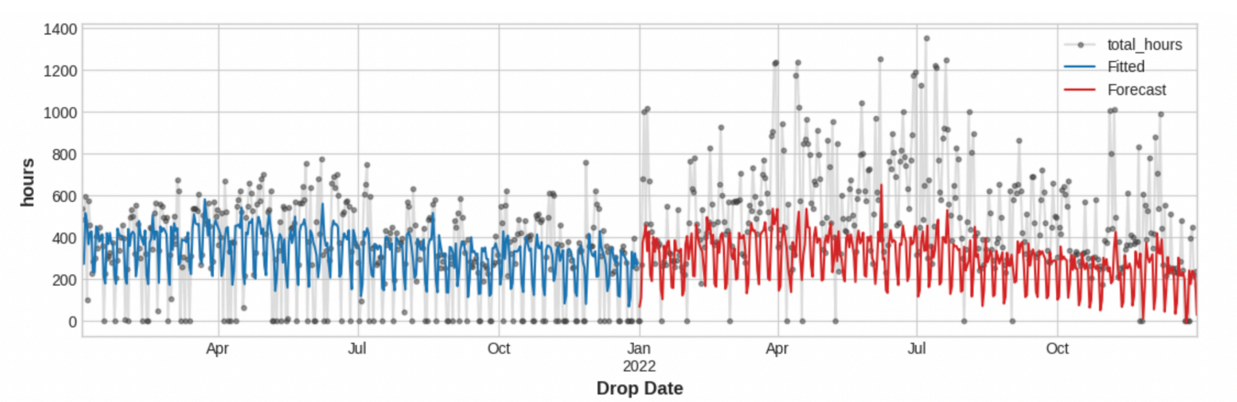
When examining serial dependence through partial auto-correlation function plot (Figure 15) shows a significance on the previous day forecast in affecting today's forecast, as well as some significance for the third day and seventh day. This means that the previous day's performance had a strong effect on today's performance, which makes sense. If the workload on the previous day was high, there is a good possibility that it will be high today, especially when it is a busy season.

Figure 14
Partial Autocorrelation Function for Serial Dependence



Considering this information, we introduced a one-day lag in the orders and the hours, and we ran a Linear Regression model with 50/50 train/test time series split. The reason for choosing a 50/50 split is that we wanted to see if the performance of 2021 is enough for the model to predict 2022's performance. Figure 16 shows the result; however, a Linear Regression model is not able to capture the behavior of the training data properly. Therefore, its performance on the testing data is also poor. That can also be seen from the high symmetric mean absolute percentage error (SMAPE) values. SMAPE is used in this case because of the zero values in the daily orders that prevented us from using MAPE.

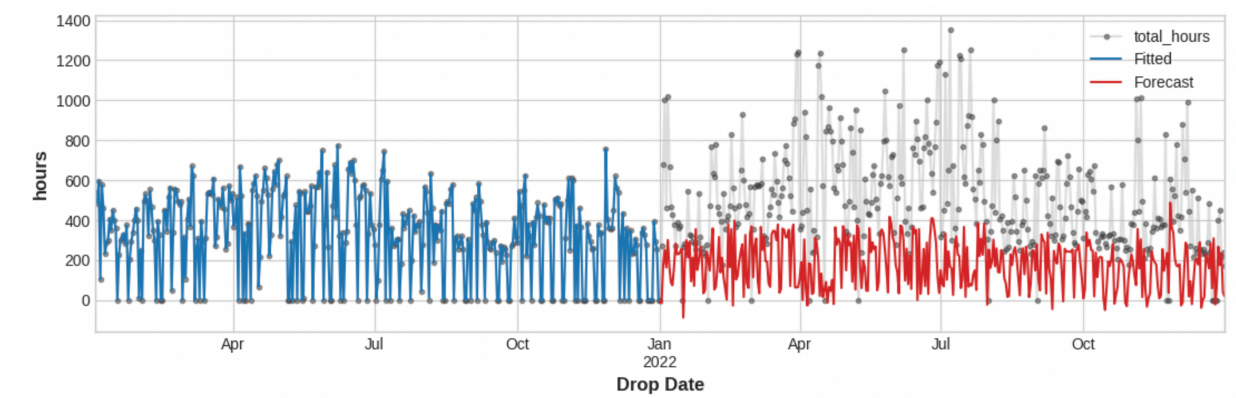
Figure 15
Linear Regression with 50/50 Split for Daily Hours (SMAPE in %)



	Method	Validation SMAPE	Validation RMSE
0	Linear Regression	63.78	318.35

We moved to an XGBOOST model with the same parameters used for the Linear Regression model to see how it will perform. From Figure 17, XGBOOST improved on the training data. That can also be seen from its SMAPE, however, on the testing data it still performed poorly and was still not able to capture the upward trend in 2022.

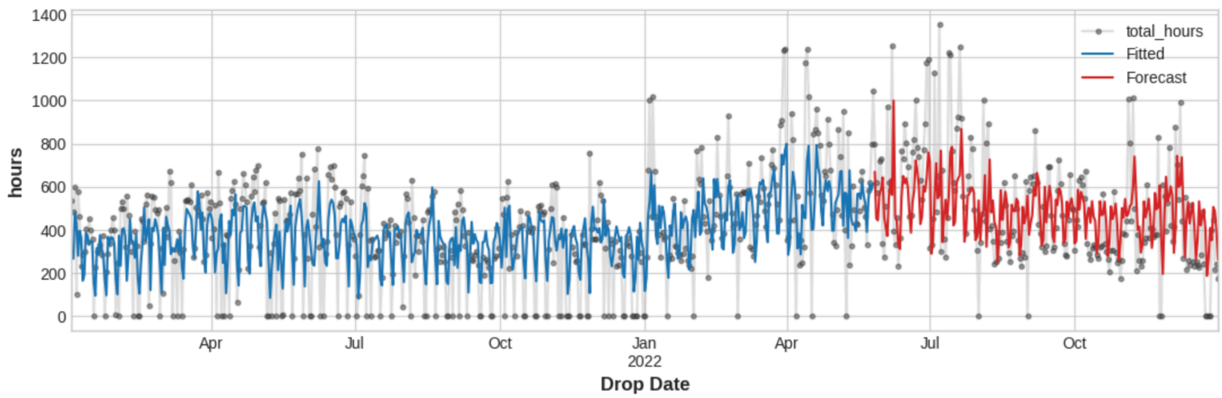
Figure 16
XGBOOST with 50/50 Split for Daily Hours (SMAPE in %)



	Method	Validation SMAPE	Validation RMSE
0	Linear Regression	104.0	423.01

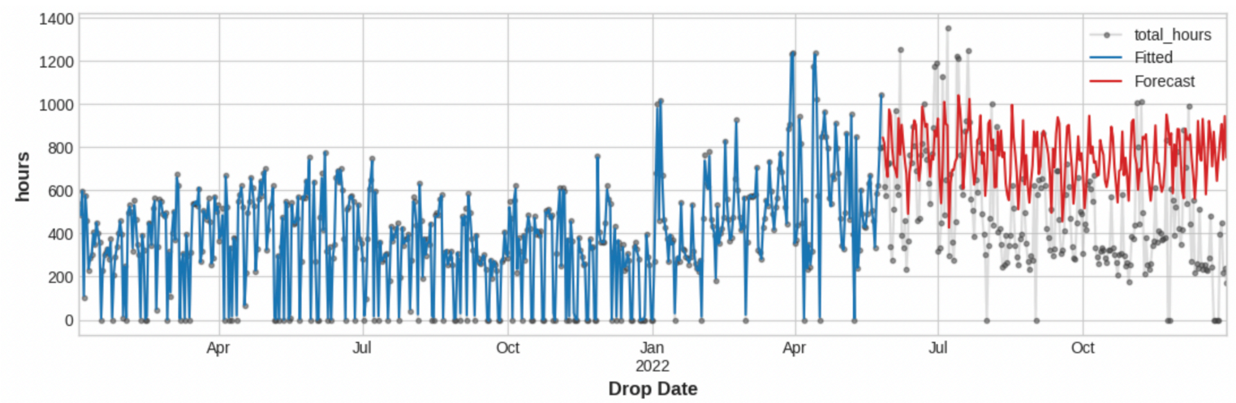
With these results we wanted to change the train/test split to include some data from 2022 to see how the models will perform. We changed the split from 50/50 to 70/30 and ran both models. Figures 18 and 19 below show the results of these iterations.

Figure 17
Linear Regression with 70/30 Split for Daily Hours (SMAPE in %)



	Method	Validation SMAPE	Validation RMSE
0	Linear Regression	39.38	235.47

Figure 18
XGBOOST with 70/30 Split for Daily Hours (SMAPE in %)



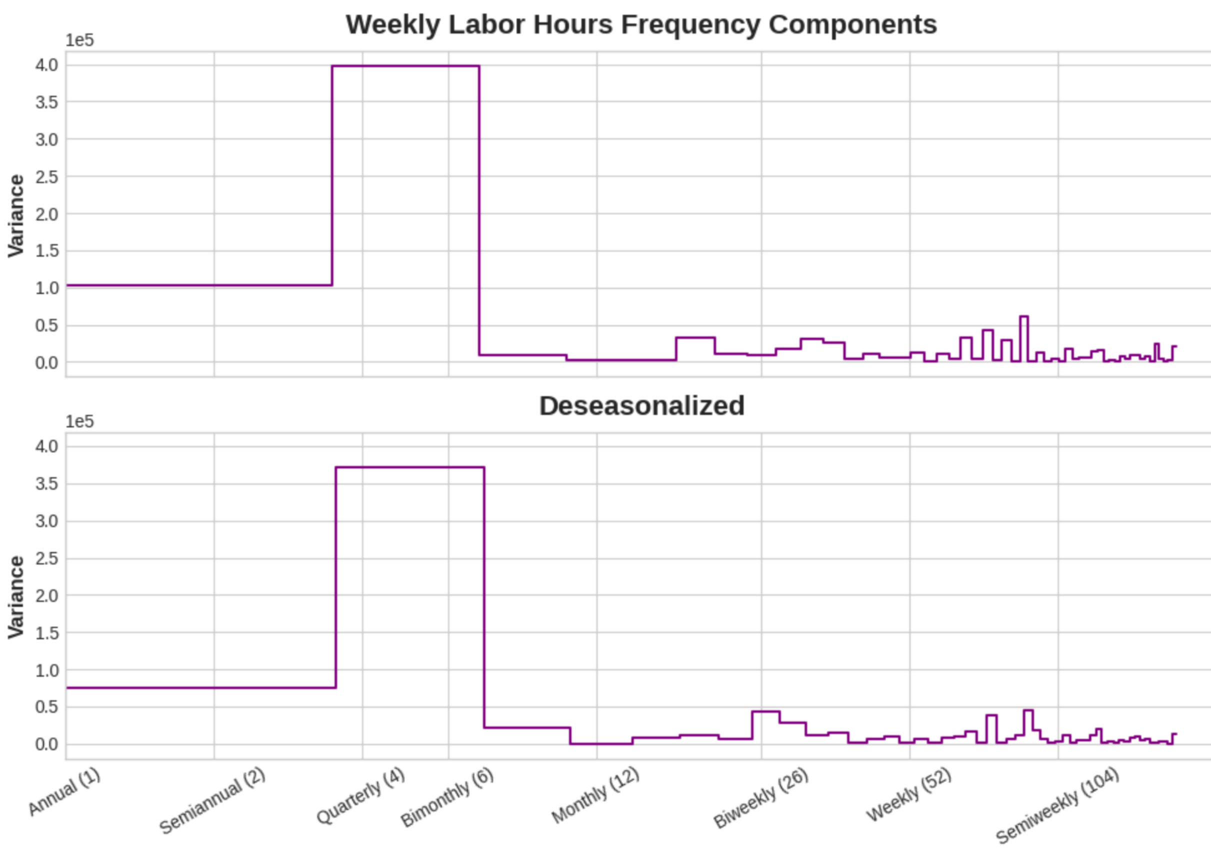
	Method	Validation SMAPE	Validation RMSE
0	Linear Regression	58.0	365.77

As seen in Figures 18 and 19 both models improved in performance due to having training data from 2022, however, both are still performing poorly when it comes to predicting the high and low seasons, which is extremely important for such a predictive model.

4.1.3.2 Weekly Labor Hours

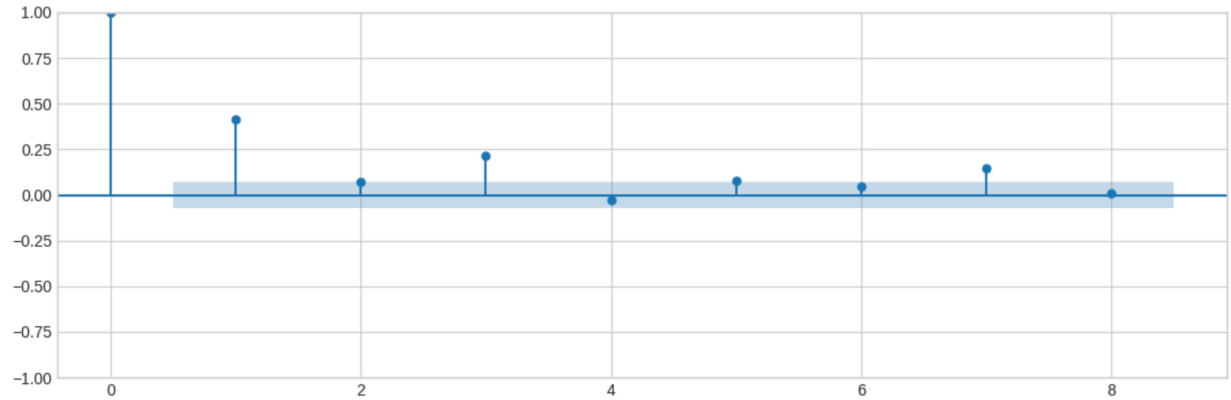
We started with a Fourier series function with a monthly frequency and an order of 4. We ran the periodogram again to see how the de-seasonal one would look like. As seen from Figure 20, the Fourier series was able to decompose some of the signal, but some seasonal components can still be seen.

Figure 19
Periodogram Before and After Fourier Series for Weekly Hours



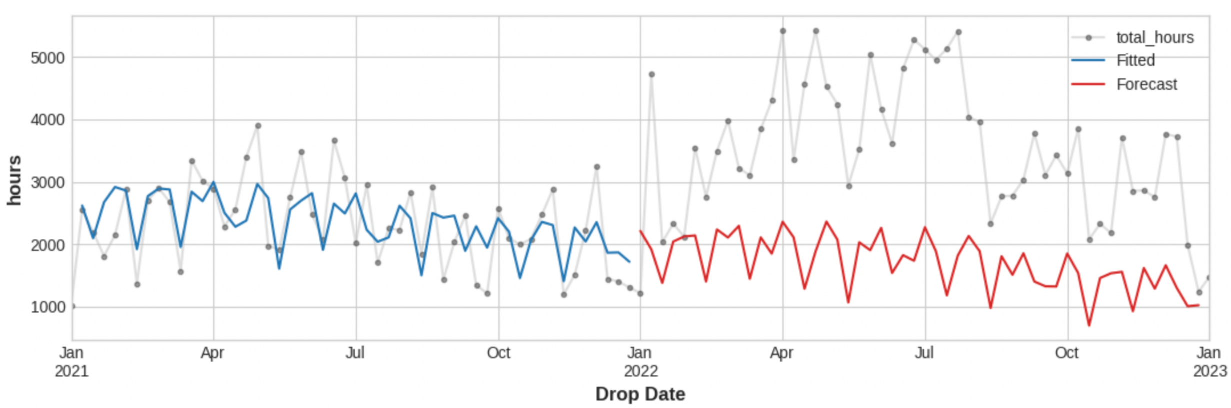
When examining serial dependence through partial auto-correlation function plot, the plot, Figure 21, shows similar results to Figure 15 of the daily hours.

Figure 20
Partial Autocorrelation Function for Weekly Hours



We introduced a one-week lag in the orders and the hours, and we ran a Linear Regression model with 50/50 train/test time series split, as we did for the daily labor hours. Figure 22 shows the result, and it seems we have a similar behavior to the daily labor hours with the same parameters. We also see that 2022's upward trend is not captured in the model. However, this time we have lower MAPE for both training and testing data.

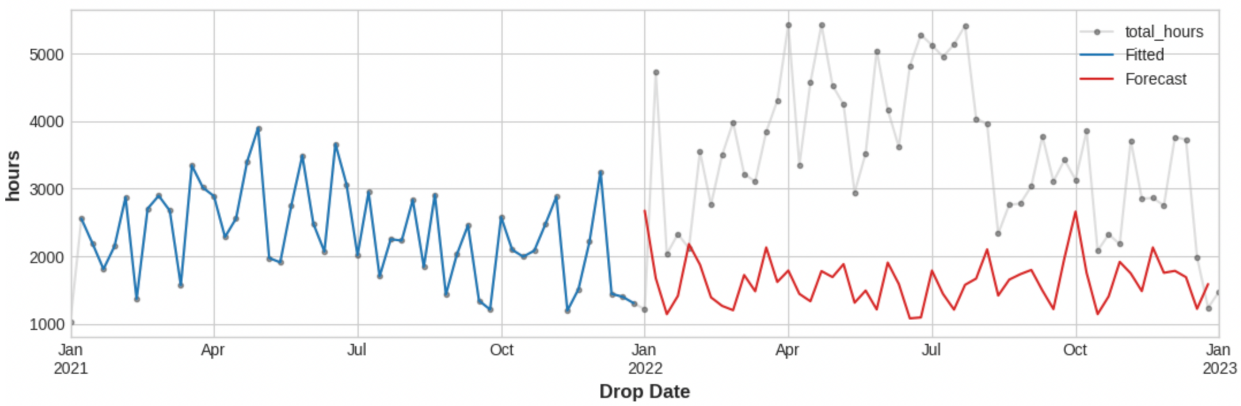
Figure 21
Linear Regression with 50/50 Split for Weekly Hours (MAPE in %)



	Method	Validation SMAPE	Validation MAPE	Validation RMSE
0	Linear Regression	67.99	50.0	2065.58

We moved to an XGBOOST model with the same parameters for the Linear Regression model to see how it will perform. From Figure 23, XGBOOST improved significantly on the training data and that can also be seen from its very low MAPE, however, on the testing data it did not improve in comparison to the Linear Regression model and was still not able to capture the upward trend in 2022.

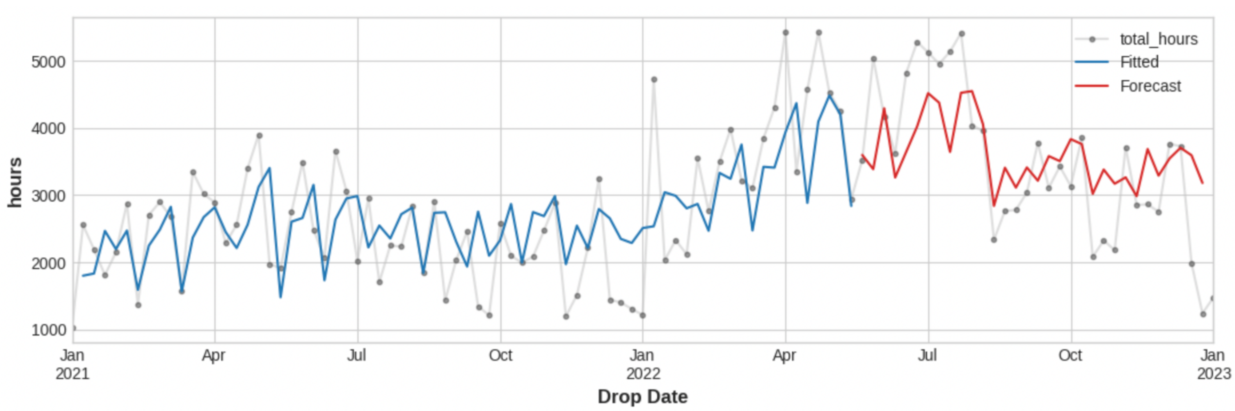
Figure 22
XGBOOST with 50/50 Split for Weekly Hours (MAPE in %)



	Method	Validation SMAPE	Validation MAPE	Validation RMSE
0	Linear Regression	71.86	53.0	2225.31

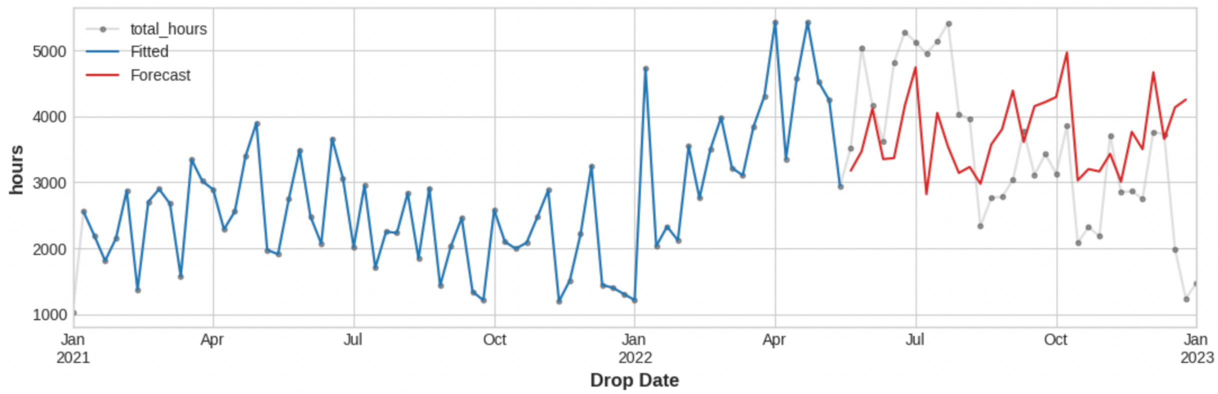
Similar to what we did for the daily labor hours, we changed the train/test split to 70/30 to see the effects. Figures 24 and 25 below show the results of these iterations.

Figure 23
Linear Regression for 70/30 Split for Weekly Hours (MAPE in %)



	Method	Validation SMAPE	Validation MAPE	Validation RMSE
0	Linear Regression	22.03	26.0	873.89

Figure 24
XGBOOST with 70/30 Split for Weekly Hours (MAPE in %)



	Method	Validation SMAPE	Validation MAPE	Validation RMSE
0	Linear Regression	27.11	34.0	1079.22

From Figures 24 & 25 we can see that this iteration demonstrated the highest improvement to the model. In Linear Regression the model improved on the validation data by 50% in comparison to the weekly 50/50 split iteration. With XGBOOST the validation MAPE improved by 40% in comparison to the weekly 50/50 split iteration.

4.1.4 Forecasting Models Summary

Table 7 shows a summary of all the results of the different models that used. As seen from Table 7, the models' performance improved when we moved from daily data to weekly data using Fourier Series with Linear Regression and XGBOOST.

Table 7
Summary of All Forecasting Models Iterations

Iteration	Model	SMAPE	MAPE	RMSE
1	Naïve Forecast	51.71%	Inf	268.54
2	XGBOOST	46.13%	94.00%	18961.39
3	Linear Regression on Daily Hours after Fourier Series with 50/50 Split	63.78%	inf	318.35
4	XGBOOST on Daily Hours after Fourier Series with 50/50 Split	104.00%	inf	423.01
5	Linear Regression on Daily Hours after Fourier Series with 70/30 Split	39.38%	inf	235.47
6	XGBOOST on Daily Hours after Fourier Series with 70/30 Split	58.00%	inf	365.77
7	Linear Regression on Weekly Hours after Fourier Series with 50/50 Split	67.99%	50.00%	2065.58
8	XGBOOST on Weekly Hours after Fourier Series with 50/50 Split	71.86%	53.00%	2225.31
9	Linear Regression on Weekly Hours after Fourier Series with 70/30 Split	22.03%	26.00%	873.89
10	XGBOOST on Weekly Hours after Fourier Series with 70/30 Split	27.11%	34.00%	1079.22

4.2 Workforce Forecasting Model Limitations

The results from the different models we tried showed improvements, however, to have a robust model with a higher level of accuracy, we need more improvement than we obtained. The accuracy and reliability of a machine learning model heavily depend on the quality and completeness of the data used for training and prediction. In the context of our project, if the data used to train the model is inaccurate, inconsistent, or incomplete, it can result in inaccurate predictions. For example, if the historical data used to train the model contains errors in workforce records, misses accurate reporting of working hours by employees or the system, or lacks important variables, such as seasonality or shift patterns, the model may produce inaccurate predictions, and this what we are experiencing in this case.

In addition to that, availability of data can also be a limitation. Workforce forecasting models often require a large amount of data to train and generate accurate predictions. However, obtaining sufficient and relevant data may be challenging, especially in cases where data collection processes are not well-established, or data is not consistently recorded. For example, data on specific workforce variables, such as absenteeism or employee skill levels, may be limited or unavailable, which can affect the accuracy of the model's predictions. In our case we only had data for 2 years of records, with no indication of experience levels or turnover rates.

Workforce demand in a warehouse can be influenced by various factors, such as seasonality, promotions, or changes in business operations. If the data used to train the model does not capture these changes or if the underlying patterns in the data are unstable, it can impact the accuracy of the model's predictions. For example, sudden changes in demand patterns or operational processes that are not reflected in the historical data may lead to inaccurate predictions, as the model may not be able to adapt to such changes. This was clear to us in many instances where the demand spiked without a clear reason in the data. When we tried to look at it from a holidays point of view it did not provide us convincing conclusions. It is important to look in granular detail at such spikes in demand and working hours to understand their underlying causes and to be better prepared when they happen again.

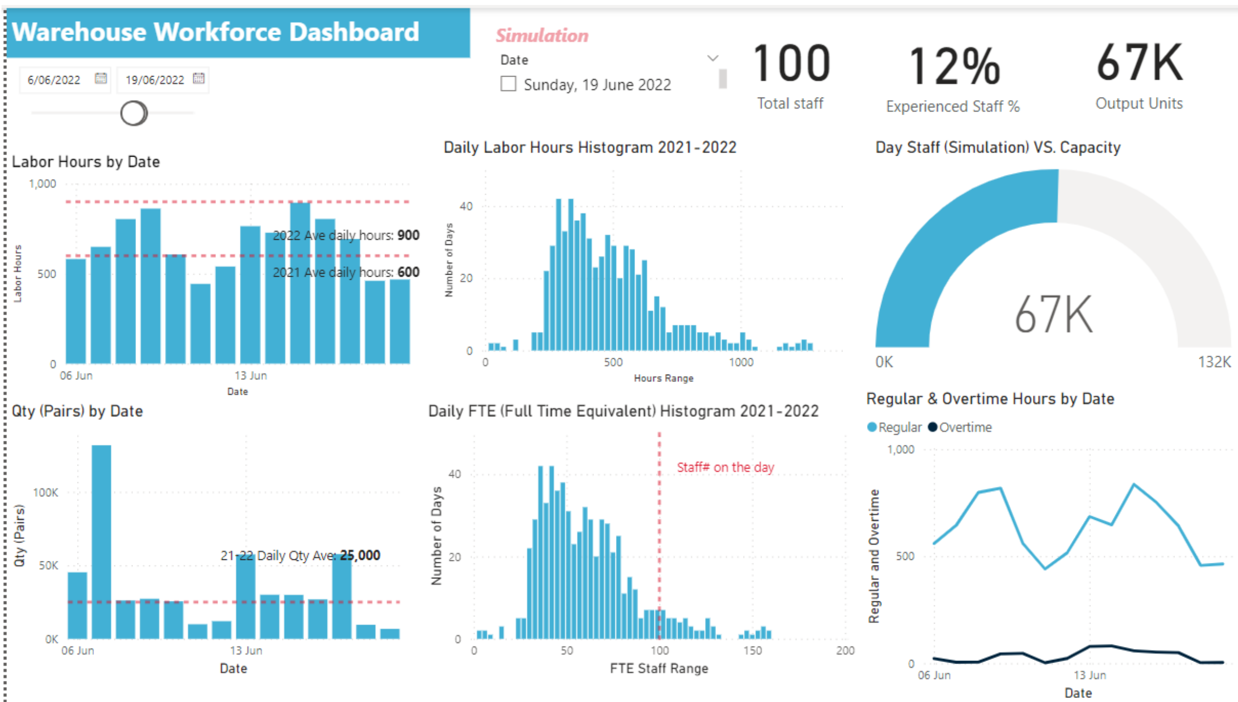
Bias in the data used for training the model can also be a limitation. If the historical data used to train the model is biased, such as different hiring practices or inaccurate recording of workforce data, it can result in biased predictions. For instance, if the historical data used to train the model contains biased information about certain operational behaviors, it can lead to biased predictions, resulting in inaccurate workforce planning decisions. This point will be discussed in more detail in Chapter 5 Discussion.

4.3 Dashboard

A well-designed performance dashboard is essential for monitoring warehouse workforce and efficiency improving workforce forecasting, as it enables real-time tracking of key performance indicators (KPIs), facilitates informed decision-making, and promotes continuous optimization of warehouse operations. We developed two dashboards' prototypes showcasing the interaction of digital twins with various technologies, assisting warehouse managers in making data-driven decisions.

The first dashboard, as shown in figure 25, presents a visual representation of key performance indicators (KPIs) that offer insights into the efficiency and productivity of warehouse employees, including daily labor hours and same-day order quantities—specifically, pairs of shoes in this case. Daily labor hours can be further broken down into regular and overtime hours. A simulation, displayed in the top corner, uses dummy data to demonstrate staff experience within the warehouse. The gauge visual indicates warehouse capacity and the simulated daily expected output, calculated based on experienced staff productivity, new staff productivity, the total number of staff members on a given day, and the percentage of experienced staff.

Figure 25
Workforce Forecasting Model Dashboard - Part 1

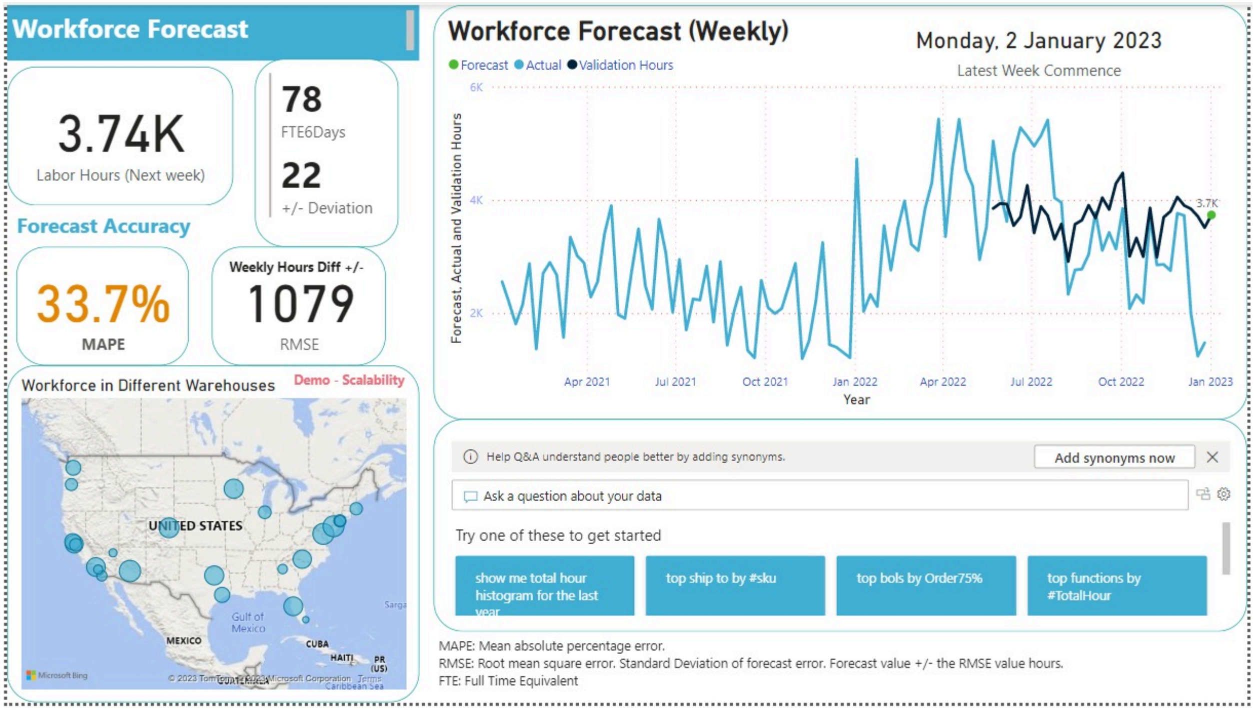


To optimize the use of this data, warehouses should categorize employees as experienced or new in their labor management systems. In other words, the warehouse team can have a database to flag the onboarding of completely new employees and build on it to record their experience levels. As data is collected, it can be integrated as a feature in the workforce forecasting model, enabling the model to learn from the new information and continuously improve its forecasting accuracy.

The second dashboard not only illustrates the warehouse workforce situation in correlation with daily order quantities—pairs of shoes in this case—but also presents the machine learning forecasting prototype results to the warehouse team, as depicted in Figure 26. The dashboard serves as a convenient tool for the warehouse management team, assisting them in the decision-making process. Figure 26 shows more contributions:

1. Weekly forecasts generated by the model.
2. Model accuracy measurements: Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).
    a. MAPE, the percentage error display will change color to indicate different levels of accuracy: green signifies less than 20% error, amber indicates an error between 20% and 50%, and red represents errors exceeding 50%, which triggers alerts.
    b. RMSE, or Root Mean Square Error, represents the standard deviation of forecast errors. In our case, it measures the weekly hours' difference from the forecast. On average our weekly forecast was off by 1,079 hours compared to the actual value that is based on a 6-day/8-hour week (4,800 hours) for 100 employees.

Figure 26
Workforce Forecasting Model Dashboard – Part 2



To ensure ease of understanding and usability by warehouse managers, all hourly data are converted into Full Time Equivalent (FTE) for the number of staff required, assuming an 8-hour workday and a 6-day work week. This conversion is based on the understanding that warehouses typically prefer to avoid weekend work. However, observations reveal that weekend hours are still recorded, albeit at a lower rate compared to Monday through Friday.

3. Full Time Equivalent (FTE) for the number of employees involved in the picking process, based on an 8-hour per day, 6-day work week.
4. A map highlighting the potential for replicating this solution in other warehouses, creating a network that displays the number of employees at any given moment.
5. An AI-generated list of questions, produced by Power BI from the underlying data, to address frequently asked queries.

5 Discussion

In this chapter we discuss the insights from the workforce forecasting model and the three technologies that we recommend to be used in the warehouse.

5.1 Workforce Prediction Model Insights

Considering the results reported in Chapter 4 and how they did not improve dramatically even with more complex forecasting techniques, we took another look at the data of the orders and hours to try to understand the behavior. We plotted the daily orders from 2021 and 2022 to see how they changed from one year to another. Then we created the same plot for the daily hours. Figures 26 & 27 represent these two plots, and we can see the following:

- Daily orders: There was no major increase in the number of orders between the two years, as shown in Figure 26. The total orders increased by 9% in 2022 when compared to 2021.
- Daily hours: There was a significant increase in the number of hours between 2021 and 2022, as shown in Figure 27 (red line). The total hours increased by 52% in 2022 when compared to 2021.

Figure 27
Daily Orders for 2021-2022

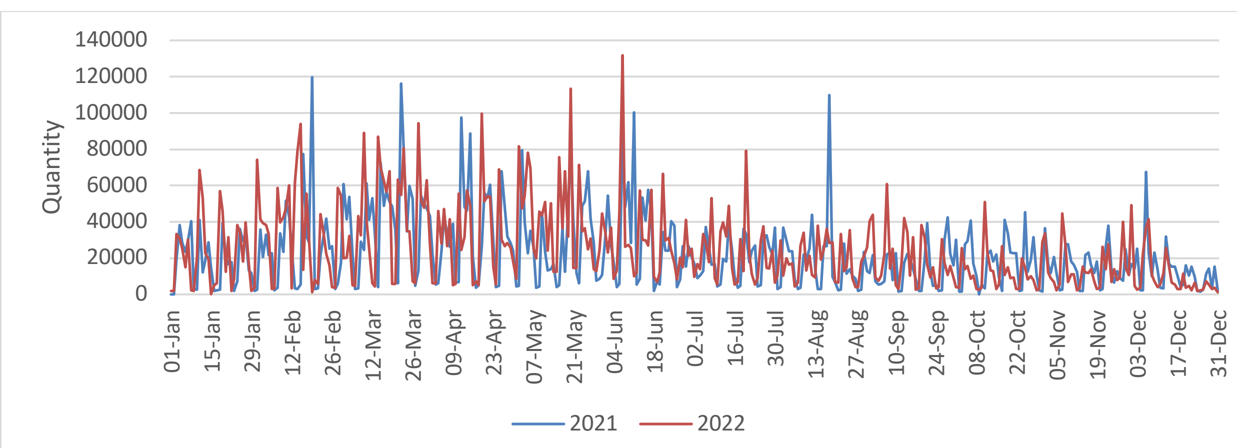
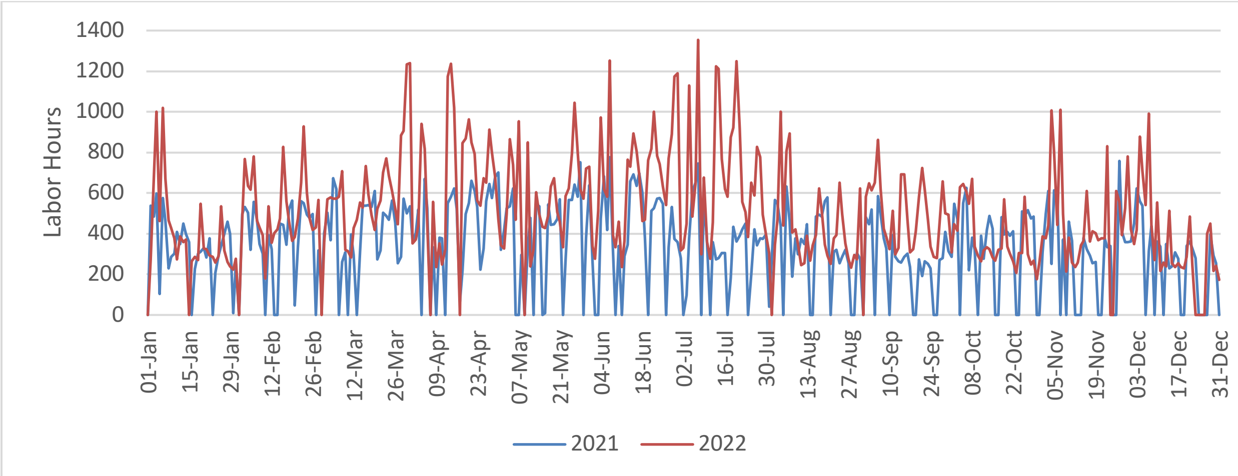


Figure 28
Daily Hours for 2021-2022



The number of additional orders does not justify the increase in hours of operation. These increases can be attributed to several factors, such as:

A. Labor availability: If there are constraints in labor availability, such as reduced staffing levels due to labor shortages or turnover, it may result in longer picking hours to complete the same volume of orders. Limited staffing experience levels could result in reduced picking efficiency and productivity, leading to extended hours of operation.

B. Labor agencies: if the agencies used regularly are changed or ended up sending labor with different experience levels than the warehouse needs, it may result in a longer learning curve, which can extend the number of hours of operation.

C. Operational changes: Changes in operational processes, such as new picking methods, revised order prioritization, or changes in order batching, may impact picking efficiency and require longer hours of operation to maintain the same level of order fulfillment. For example, if a warehouse shifts from batch picking to zone picking, it may require longer hours of operation to cover all zones.

D. Equipment limitations: If the warehouse relies on specific equipment, such as forklifts or conveyor belts, and there are limitations in equipment availability, capacity, or even longer downtimes of these equipment, it may result in longer picking times.

5.2 Technologies Insights

AGVs

Given that the Warehouse primarily deals with shoes, we recommend focusing on the following AGV types to enhance warehouse operations. Forklift AGVs and pallet AGVs can efficiently move pallets of shoes from one location to another, accommodating various load sizes and handling different order types with ease. Additionally, Conveyor AGVs can complement the existing conveyor systems, further streamlining the movement of goods. Once these AGVs are deployed in the warehouse, they can capture data on product movement, velocity, and location, among other parameters. This data can then be integrated with the WMS for real-time monitoring and analysis, optimizing warehouse operations and increasing overall efficiency.

Picking Robots

Because the selected Warehouse primarily deals with shoes, we recommend focusing on the following picking robot types:

1. Piece Picking Robots: These robots can handle individual items, making them ideal for picking various shoe sizes during the e-commerce order fulfillment process.
2. Robotic Arm Pickers: With their versatile robotic arms, these pickers can easily pick and place shoes, streamlining the picking process.
3. Goods-to-Person Robots: By bringing items or shelves directly to the picker. More precisely, time spent walking to, through, and back from the storage areas.

Once picking robots are deployed in the warehouse, they can capture data on product dimensions, weight, and location, among other parameters. This data can be integrated with the WMS for real-time monitoring and analysis. To make the captured data more accessible for workload forecasting, the implementation steps are like those for AGVs. Ensuring that picking robots are equipped with the necessary sensors, integrating their data with the WMS, using data analytics tools for processing, analyzing the data, and utilizing developed models to improve workload forecasting accuracy are all key components for a successful implementation of picking robots in the warehouse.

AS/RS systems

Considering that the warehouse primarily deals with shoes, we recommend focusing on the following AS/RS types:

1. Mini Load AS/RS: Since shoes are relatively small and lightweight, Mini Load AS/RS can provide efficient storage and retrieval solutions tailored to the warehouse's product mix.
2. Vertical Lift Modules (VLMs): These systems can offer compact storage for shoeboxes while maximizing vertical space and ensuring quick access to stored items.
3. Horizontal Carousels: For high turnover shoe models or sizes, Horizontal Carousels can provide fast and efficient storage and retrieval, streamlining the picking process.

Once AS/RS are deployed in the warehouse, they can capture data on product location, quantity, and movement, among other parameters. This data can be collected and integrated with the warehouse management system (WMS) for real-time monitoring and analysis. To ensure the successful integration of new data into the WMS, warehouse managers should consider data compatibility, quality, integration, and security. By addressing these issues systematically, managers can guarantee that the new data is accurate, reliable, and accessible for effective workload forecasting and optimization.

Sensors

Sensors play a big role in a digital twin solution, as they are an integral part of the three technologies recommended, and facilitate the real-time data gathering, as well as the communication between the different parts of the process. The provide the means for getting accurate data from the technologies and storing it in the WMS & LMS, which can help analyzing workflows, labor requirements, and getting business insights.

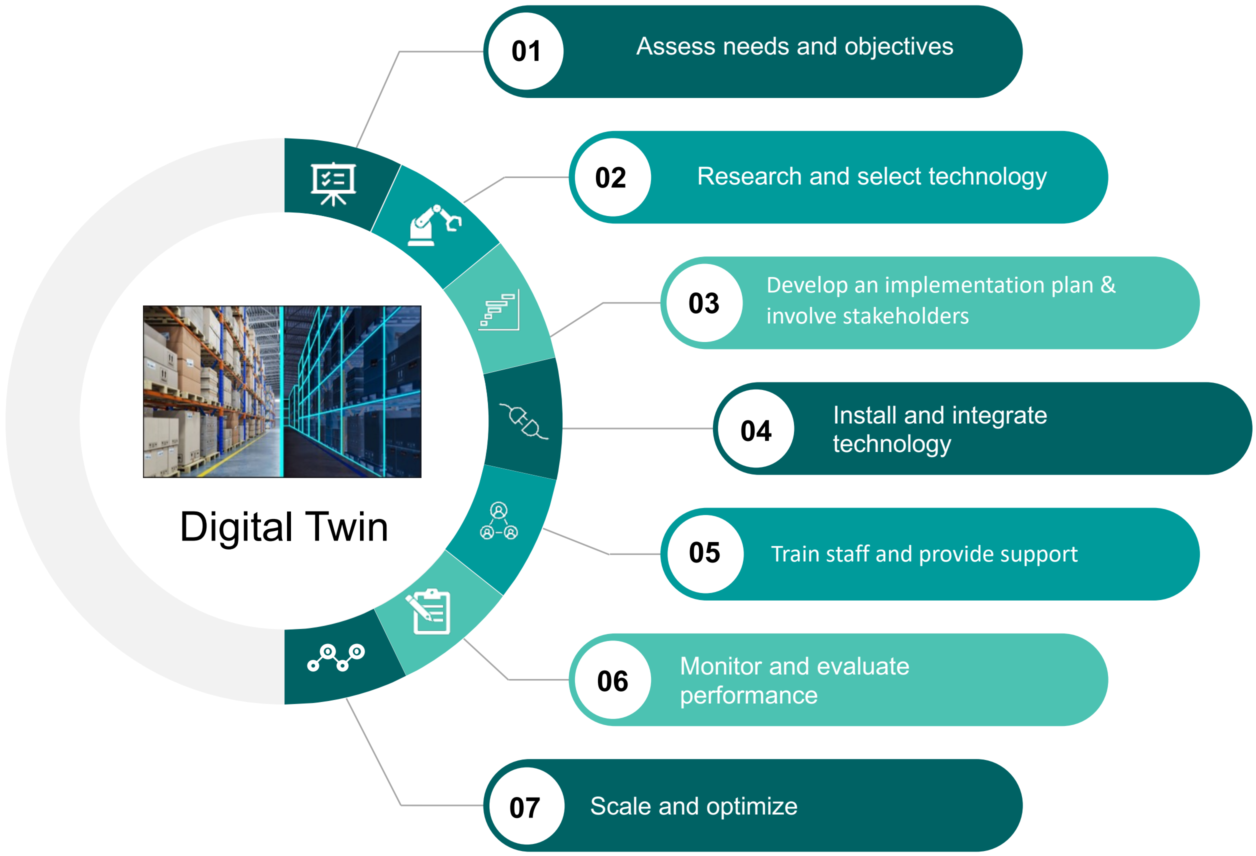
6 Actionable Recommendations

In this chapter we will discuss the roadmap that the company can follow to achieve a digital twin solution for the picking process and the possibility of expanding it to other warehouses processes. This roadmap can also serve as a guide for scaling this solution to other warehouses.

6.1 Digital Twin Implementation Roadmap

Implementing new technologies in a warehouse involves several key steps to ensure successful integration and adoption. A guide of the essential steps that we believe will be applicable to our specific case is presented in Figure 28, and below it is the detailed explanation of the roadmap.

Figure 29
Digital Twin Implementation Roadmap



Small image inside the circle: Tynan, D. (2022). *How digital twins can help unsnarl global supply chains* [Image]. Workflow.

1. Assess needs and objectives (Done): Identify the specific challenges, requirements, and goals that the new technology aims to address. Determine the desired outcomes and improvements the technology should provide.
   - Need: better workforce forecasting model and technologies recommendation to improve the picking process.
   - Objective: enhance performance by using digital twins to improve efficiency, productivity, and scalability.

2. Research and select technology (Done): Investigate the various available technologies that can meet the identified needs and objectives. Evaluate their features, benefits, costs, and compatibility with existing systems. Select the most suitable technology based on these factors.
    - Technologies: AGVs (as an example for the first implementation)
    - Evaluation: improved efficiency and productivity through the KPIs measurement from Chapter 3 section 3.7.
3. Develop an implementation plan and involve stakeholders (1-2 months): Create a detailed plan outlining the steps, timeline, resources, and responsibilities involved in the implementation process. This plan should include staff training, equipment installation, system integration, and testing. Then engage with all relevant stakeholders, including employees, management, and external partners, to ensure they are informed about the new technology and its benefits. Obtain their input and address any concerns they may have.
    - Implementation plan: share the warehouse layout with the AGVs supplier, in addition to the number of movements done in different workload seasons. This will allow the possibility of determining how many AGVs are needed for the facility and the location for the charging stations.
    - Resources: gather a project management taskforce that will be responsible for all matters related to the full implementation.
    - Stakeholders' alignment: warehouse management, picking staff and supervisors, and any external parties that are involved in the warehouse operation or movements. Inputs from all levels are crucial for successful implementation.
4. Install and integrate technology (1-2 months): Implement the new technology according to the plan, integrating it with existing systems and processes. This may involve hardware installation, software setup, and data migration or integration. Then conduct thorough testing and validation

of the new technology to ensure it is functioning correctly and meeting the desired objectives. Identify and resolve any issues that may arise during this stage.

- Hardware setup: decide on the suitable AGVs type for the picking process and look at the relevant movement paths in the warehouse to ensure the right environment for those movements.
- Software setup: setup AGVs system through an interface with the ERP and WMS in the warehouse to allow its integration.
- System requirements and changes: in some cases, the WMS will need an additional configuration to enable the AGV interface, which needs to be checked.
- Test and validate start with trial runs without any real material movement, then monitor data inputs, outputs, and performance. Then the process can move to trial runs with material on low workload days or an assigned zone to study its movement in an isolated environment with minimum interruption to the normal operation.

5. Train staff and provide support (1 month): Train employees on how to use the new technology effectively and safely. Offer ongoing support to ensure they can troubleshoot issues and maximize the benefits of the new technology.
    - Train: workshops to be delivered by the technology provider, along with on-the-job trainings, and shadowing more experienced staff with such technologies.
    - Technical support: from technology provider and technology/IT department within the warehouse team.
6. Monitor and evaluate performance (1 month): Continuously monitor the performance of the new technology, gathering data and feedback to assess its effectiveness in meeting the identified needs and objectives. Use this information to adjust and make improvements as needed.

- Monitor: downtime of the technology, charging time, interaction with staff, and related operational issues that arise from the use of the AGVs
- Evaluate performance: KPIs improvement compared to before AGVs implementation and staff feedback and concerns from the new way of working is crucial to successful implementation.

7. Scale and optimize (2-3 months): Once the new technology has proven effective, consider scaling its use to other areas of the warehouse or organization. Continuously seek opportunities to optimize its performance and maximize the benefits it provides.
   - Scale: AGV adoption in other processes in the warehouse, or other warehouses.
   - Optimize: continuous feedback loop to ensure optimum performance.

6.2 Warehouse Digital Twin Solution

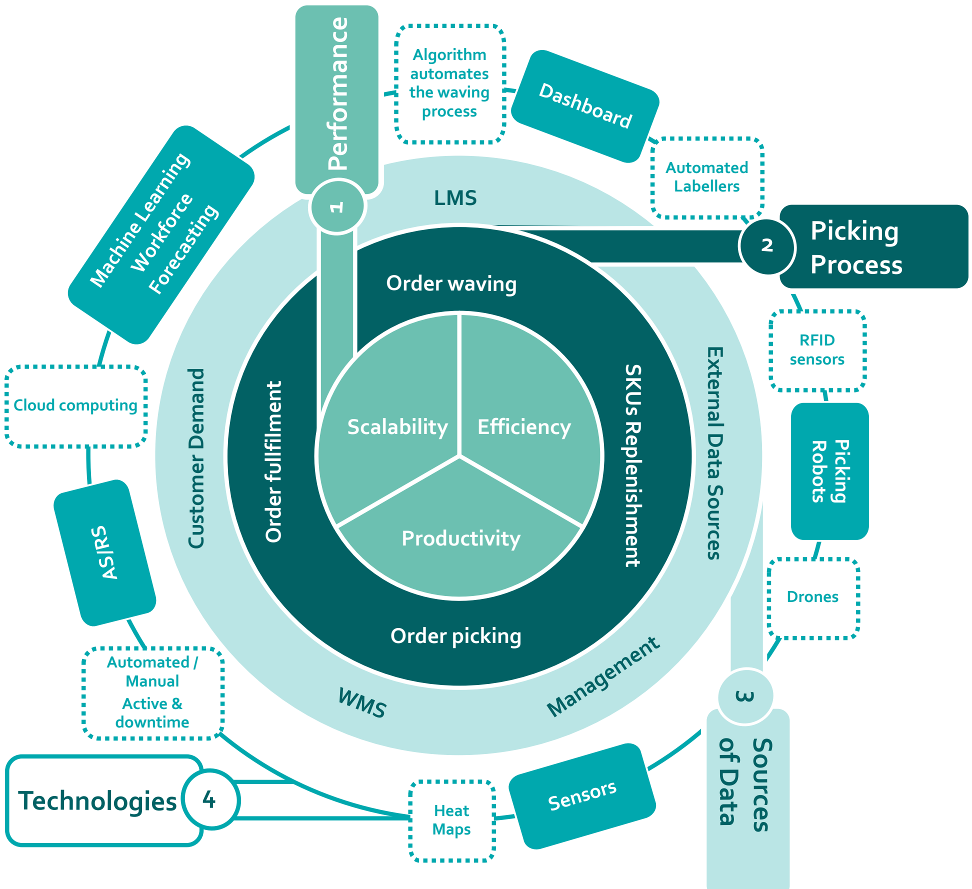
Finally, we presented a comprehensive digital twin solution embodying the four-layer structure delineated in Chapter 3, Section 3.7 (Figure 10), with more and better technology adoption and integration. The framework presented in Figure 30 within this section illustrates the expansive potential of the digital twin prototype and its intrinsic flexibility, making it a tool which can be used in numerous ways.

The lever within Layer 1 can be altered to gauge diverse performance metrics, such as inventory KPIs, and order fulfillment rates, contingent upon the specific targets set by the team. The lever in Layer 2 can facilitate navigation among different warehouse processes. Digital twin technology can enhance processes such as inventory management or packing and can be extended to encompass the entire warehouse, or even multiple warehouses. Layer 3's lever holds potential for extension to incorporate additional systems, such as the Yard Management System (YMS), as data sources. This expansion would enable the warehouse to extend the application of digital twin technologies beyond its own confines to other connected areas outside the warehouse. Finally, Layer 4, as depicted in Figure 30, underscores the potential for advanced technological adoption in the picking process. With digital twin solutions and sensor integration, each component of the picking process can be enlivened. This advancement can also extend to other warehouse processes.

The adoption of the digital twin solution necessitates an enhanced level of integration and data quality improvement across various processes and systems. The integration of sensors and technologies can streamline this process by providing vital data, thus fostering effective interconnectivity among systems and processes. Warehouse teams can begin to test various possibilities for integration derived from performance metrics needing improvement. Real-time data acquisition from the various elements promotes the learning and operational improvement of this solution. Moreover, it provides a crucial foundation for the team to simulate scenarios involving operational changes, swiftly adapt to evolving customer requirements, and prepare for peak demand periods, thereby optimizing resource utilization.

This expanded framework serves as a foundation for future larger-scale adoption of digital twin solutions. It outlines a clear path to address varied requirements while maintaining the flexibility to meet the company's diverse needs and objectives.

Figure 30
Complete Warehouse Digital Twin Solution



6.3 Conclusion

In conclusion, our research for the implementation of digital twin technology in the warehouse picking process provides promising insights and potential for improving efficiency and productivity. A digital twin prototype represents a digital replica of the physical entities. It includes technologies that gather real time data, learn dynamically, and is able to facilitate decision making.

By leveraging machine learning forecasting models and integrating technologies such as automated guided vehicles (AGVs), picking robots, and automated storage and retrieval systems (AS/RS), along with the sensors in these technologies, we encapsulate the core components of a digital twin prototype. The sensors enable real-time interaction with the different systems and facilitate continuous learning and operational improvements. Taking into account the data quality recommendations, the workforce forecasting model, combined with the recommended technologies, will enable the warehouse team to make informed decisions and better anticipate labor requirements, resulting in optimized operations.

6.3.1 Limitations and Future Avenues

Our research highlighted the limitations of the forecasting model due to the data availability and data quality issues. Machine learning forecasting models require large datasets, and more features to provide better predictions. Features such as experience level, turnover rates, accurate measurement of throughput, promotions, and downtime will help improve the prediction of the forecasting model. As the model will keep learning from these patterns to enhance the forecast accuracy. These features will enable the model to understand underlying causes behind labor hours fluctuations.

In the case of having these features as part our model and longer time horizon datasets, we would have explored the possibility of seeing a better model performance and provided a deeper understanding of the behavior changes in the labor requirements from one year to another. This can still be done in future work by capitalizing on the foundation of this project and using it as a starting point.

Furthermore, there is significant potential to expand this work beyond the picking process. The insights gained from this project can be applied to other processes within the warehouse, such as inventory management, order fulfillment, and shipping. In addition, it helped to uncover the importance of capturing important metrics for the operation, such as staff experience levels and turnover rates that contribute to the workforce forecasting. By adopting other innovative technologies and integrating them into the digital twin framework, we can further streamline operations and maximize efficiency across the entire warehouse ecosystem.

Moreover, this digital twin solution can be scaled and implemented in other warehouses, offering a standardized approach to improving performance and meeting the evolving demands of the industry.

There are great possibilities for the expansion of the framework and the roadmap developed in this research in different dimensions, which can generate greater value for the company’s warehouse operations. Digital twins continue to prove to be a relevant tool for taking warehouse operations to new frontiers and pushing the boundaries of innovation in a segment that consistently grows in importance in the supply chain world.

REFERENCES

Bormann, R., de Brito, B. F., Lindermayr, J., Omainska, M., & Patel, M. (2019). Towards automated order picking robots for warehouses and retail. In *Computer Vision Systems: 12th International Conference, ICVS 2019, Thessaloniki, Greece, September 23–25, 2019, Proceedings 12* (pp. 185-198). Springer International Publishing. https://doi.org/10.1007/978-3-030-34995-0_18

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Broo, D. G., Bravo-Haro, M., & Schooling, J. (2022). Design and implementation of a smart infrastructure digital twin. *Automation in Construction*, *136*, 104171. https://doi.org/10.1016/j.autcon.2022.104171

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining (pp. 785-794).*

Cupek, R., Drewniak, M., Fojocik, M., Kyrkjebo, E., Lin, J. C. W., Mrozek, D., Ovsthus, K., & Ziebinski, A. (2020). Autonomous Guided Vehicles for Smart Industries – The State-of-the-Art and Research Challenges. *International Conference on Computational Science - ICCS 2002*, 330-343. https://doi.org/10.1007/978-3-030-50426-7_25

Dallari, F., Marchet, G., & Melacini, M. (2008). Design of order picking system. *The International Journal of Advanced Manufacturing Technology*. https://link.springer.com/content/pdf/10.1007/s00170-008-1571-9.pdf?pdf=button

Faloutsos, C., Gasthaus, J., Januschowski, T., & Wang, Y. (2018). Forecasting big time series: Old and new. *Proceedings of the VLDB Endowment*. https://doi.org/10.14778/3229863.3229878

Ferrari, A., Zenezini, G., Rafele, C., & Carlin, A. (2022). A Roadmap towards an Automated Warehouse Digital Twin: current implementations and future developments. *IFAC-PapersOnLine*, *55*(10), 1899-1905. https://doi.org/10.1016/j.ifacol.2022.09.676

Gerlach, B., Zarnitz, S., Nitsche, B., & Straube, F. (2021). Digital Supply Chain Twins—Conceptual Clarification, Use Cases and Benefits. Logistics, 5(4), 86. https://doi.org/10.3390/logistics5040086

Ghaouta, A., Riad, M., & Okar, C. (2021). Machine Learning for Warehouse Management: A conceptual framework. *2021 Third International Conference on Transportation and Smart Technologies (TST)*. https://doi.org/https://ieeexplore.ieee.org/abstract/document/9516095

Guo, J., & Lv, Z. (2022). Application of Digital Twins in multiple fields. *Multimedia tools and applications*, *81*(19), 26941-26967. https://doi.org/10.1007/s11042-022-12536-5

Ho, G. T. S., Choy, S. K., Tong, P. H., & Tang, V. (2022). A forecasting analytics model for assessing forecast error in e-fulfilment performance. *Industrial Management & Data Systems*, (ahead-of-print). https://doi.org/10.1108/IMDS-01-2022-0056

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36-52. https://doi.org/10.1016/j.cirpj.2020.02.002

Kahraman, C., Öztayşi, B., Onar, S.C. (2020). Warehouse Location Design Using AS/RS Technologies: An Interval Valued Intuitionistic Fuzzy AHP Approach. In: Kahraman, C., Cebi, S. (eds) *Customer Oriented Product Design. Studies in Systems, Decision and Control* (pp 379-397), vol 279. Springer, Cham. https://doi.org/10.1007/978-3-030-42188-5_19

Kumar, R., Haleem, A., Garg, S., & Singh, R. (2015). Automated guided vehicle configurations in flexible manufacturing systems: A comparative study. *International Journal of Industrial & Systems Engineering*, *Vol 21, No. 2*. https://doi.org/10.1504/IJISE.2015.071510

Lambrechts, W., Klaver, J. S., Koudijzer, L., & Semeijn, J. (2021). Human factors influencing the implementation of Cobots in high volume distribution centres. *Logistics*, *5*(2), 32. https://doi.org/10.3390/logistics5020032

Manzini, R., Gamberi, M., & Regiattieri, A. (2006). Design and control of an AS/RS. *The International Journal of Advanced Manufacturing Technology*, 766-774. https://doi.org/10.1007/s00170-004-2427-6

Moshayedi, A. J., Jinsong, L., & Liao, L. (2019). AGV (automated guided vehicle) robot: Mission and obstacles in design and performance. *Journal of Simulation and Analysis of Novel Technologies in Mechanical Engineering*, *12*(4), 5-18.

Mugrage, K. (2022, November 2). Better developer platforms are the key to better digital products. *MIT Technology Review*.

Nguyen, T., Jump, A., & Casey, D. (2023, January 19). Emerging Tech Impact Radar: 2023. *Emerging Tech Impact Radar*.

Saenz, M. J., & Tozanli, O. (2022). Unlocking the Potential of Digital Twins in Supply Chains. *Sloan Management Review*.https://sloanreview.mit.edu/article/unlocking-the-potential-of-digital-twins-in-supply-chains/

Smith, D. (2018). *Birth of the Technology Radar*. Thoughtworks. https://www.thoughtworks.com/insights/blog/birth-technology-radar#:~:text=It's%20been%20almost%20a%20decade,and%20evolving%20trends%20in%20tech.

Stączek, P., Pizoń, J., Danilczuk, W., & Gola, A. (2021). A digital twin approach for the improvement of an autonomous mobile robots (AMR's) operating environment—A case study. *Sensors*, *21*(23), 7830. https://doi.org/10.3390/s21237830

Tang, Y. M., Ho, G. T. S., Lau, Y. Y., & Tsui, S. Y. (2022). Integrated smart warehouse and manufacturing management with demand forecasting in small-scale cyclical industries. *Machines*, *10*(6), 472. https://doi.org/10.3390/machines10060472

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72(1), 37-45*.

Tekinerdogan, B., & Catal, C. (2021). Design of a reference architecture for developing smart warehouses in industry 4.0. *Computers in Industry*. https://doi.org/10.1016/j.compind.2020.103343

Tutam, M. (2022). Warehousing 4.0 in Logistics 4.0. In: İyigün, İ., Görçün, Ö.F. (eds) *Logistics 4.0 and Future of Supply Chains. Accounting, Finance, Sustainability, Governance & Fraud: Theory and Application* (pp 95-118). Singapore, Springer. https://doi.org/10.1007/978-981-16-5644-6_7

Tynan, D. (2022). *How digital twins can help unsnarl global supply chains* [Image]. Workflow. https://www.servicenow.com/workflow/it-transformation/how-digital-twins-help-mitigate-supply-chain-risk.html

Van Gils, T., Ramaekers, K., Caris, A., & Cools, M. (2016). The use of time series forecasting in zone order picking systems to predict order pickers' workload. *International Journal of Production Research*, *55*(21). https://doi.org/10.1080/00207543.2016.1216659

Vrabič, R., Erkoyuncu, J. A., Butala, P., & Roy, R. (2018). Digital twins: Understanding the added value of integrated models for through-life engineering services. Procedia Manufacturing, 16, 1-8. https://doi.org/10.1016/j.promfg.2018.10.167

Winkelhaus, S. (2022). Chapter 3 - Smart warehouses—a sociotechnical perspective Author links open overlay panel. In E. H. Grosse (Ed.), *The Digital Supply Chain* (pp. 47–60). essay, Elsevier Inc. https://doi.org/10.1016/B978-0-323-91614-1.00003-4

Yao, F., Keller, A., Ahmad, M., Ahmad, B., Harrison, R., & Colombo, A. W. (2018). Optimizing the Scheduling of Autonomous Guided Vehicle in a Manufacturing Process. *IEEE 16th International Conference on Industrial Informatics*, 264-269. https://doi.org/ 10.1109/INDIN.2018.8471979

Zhang, M., Tao, F., Huang, B., Liu, A., Wang, L., Anwer, N., & Nee, A. Y. C. (2022). Digital twin data: methods and key technologies. *Digital Twin, 1, 2*. https://doi.org/10.12688/digitaltwin.17467.2

Zhao, Z., Shen, L., Yang, C., Wu, W., Zhang, M., & Huang, G. Q. (2021). IoT and digital twin enabled smart tracking for safety management. *Computers & Operations Research*, *128*, 105183. https://doi.org/10.1016/j.cor.2020.105183