

Smart Container Solutions for Port Management Efficiency

**Senior Project II Final Report**

By

|  |  |
| --- | --- |
| Yousef Alnumani | 2140532 |
| Abdulrahman Alghamdi | 2142023 |
| Mishari Abdulla | 2141386 |
| Ghazi Makkawi | 2142524 |
| Abdulrahman Alorabi | 2142059 |

Supervised By

Dr. Mohammed Alghamdi

Department of Information Systems & Technology

College of Computer sciences & Engineering

University of Jeddah

Jeddah – Saudi Arabia

# **Abstract**

A brief summary outlining the project, which uses data science methods to improve the efficiency of port management by predicting container handling times, optimizing resource allocation, and enhancing logistical operations. The project uses a synthetic dataset based on real-world data features to offer insights into how port authorities can improve operations and make better strategic decisions. This approach aims to provide practical solutions that can be applied to various ports to enhance productivity and streamline workflows.

TABLE OF CONTENTS

[Abstract 1](#_Toc182296194)

[Chapter 1: Problem Understanding and Project Planning 5](#_Toc182296195)

[1.1 Introduction 5](#_Toc182296196)

[1.2 Problem Definition 5](#_Toc182296197)

[1.3 Proposed Solution 6](#_Toc182296198)

[1.4 Motivation and Objective 6](#_Toc182296199)

[1.5 Project Deliverables 6](#_Toc182296200)

[1.6 Project Methodology 7](#_Toc182296201)

[1.7 Project Framework 7](#_Toc182296202)

[1.8 Formulate Initial Hypotheses 7](#_Toc182296203)

[1.9 Identify Data Sources 7](#_Toc182296204)

[Chapter II: Background And Literature Review/ Related Work 8](#_Toc182296205)

[2.1 Introudction 8](#_Toc182296206)

[2.2 Related Work 8](#_Toc182296207)

[2.2.1 TIDALIS 8](#_Toc182296208)

[2.2.2 Integrated Port Management by Dubai Technologies 8](#_Toc182296209)

[2.2.3 OnePort 9](#_Toc182296210)

[2.2.4 Grieg Connect's Port Management System 9](#_Toc182296211)

[2.2.5 Dswi 9](#_Toc182296212)

[2.3 Comparison Between Proposed Systems and Literature 9](#_Toc182296213)

[Chapter III: Data Understanding and Exploratory Data Analysis 11](#_Toc182296214)

[3.1Overview of the Data Section 11](#_Toc182296215)

[3.2Data Source 11](#_Toc182296216)

[3.3Data Collection Methods 12](#_Toc182296217)

[3.3.1Technical Overview 12](#_Toc182296218)

[3.3.2Primary Data Simulation: 12](#_Toc182296219)

[3.3.3Secondary Data Integration: 12](#_Toc182296220)

[3.3.4Data Enrichment 12](#_Toc182296221)

[3.3.5Expert Validation 12](#_Toc182296222)

[3.4Data Description 13](#_Toc182296223)

[3.5 Database Schema 16](#_Toc182296224)

[3.6 Exploratory Data Analytics 17](#_Toc182296225)

[Chapter 4: Data Preparation and Preprocessing 19](#_Toc182296226)

[4.1 Introduction 19](#_Toc182296227)

[4.2 Problem with the Dataset 19](#_Toc182296228)

[4.3 Explanation of Data Generation Functions 19](#_Toc182296229)

[Chapter 5: Model Building 21](#_Toc182296230)

[5.1 Introduction 21](#_Toc182296231)

[5.2 Data Preparation 21](#_Toc182296232)

[5.2.1 Data Sources 21](#_Toc182296233)

[5.2.2 Data Cleaning and Preprocessing 22](#_Toc182296234)

[5.3 Model Initialization 22](#_Toc182296235)

[5.3.1 Daily Ship Arrivals Model 22](#_Toc182296236)

[5.3.2 Container Handling Time Model 23](#_Toc182296237)

[5.4 Model Application 23](#_Toc182296238)

[5.4.1 Daily Ship Arrivals Model Application 23](#_Toc182296239)

[5.4.2 Container Handling Time Model Application 23](#_Toc182296240)

[5.5 Model Evaluation 23](#_Toc182296241)

[5.5.1 Evaluation Metrics 23](#_Toc182296242)

[5.5.2 Results 24](#_Toc182296243)

[Chapter 6: Results and Discussions 26](#_Toc182296244)

[6.1 Presentation of Results 26](#_Toc182296245)

[6.2 Performance Evaluation Metrics 26](#_Toc182296246)

[6.3 Experiment Results 26](#_Toc182296247)

[6.4 Discussion 29](#_Toc182296248)

[` Chapter 7: Conclusion and Future work 30](#_Toc182296249)

[7.1 Conclusion 30](#_Toc182296250)

[7.1.1 Model Performance Table: 31](#_Toc182296251)

[7.3 Limitations 31](#_Toc182296252)

[7.4 Future work 32](#_Toc182296253)

[Reference 33](#_Toc182296254)

LIST OF TABLES

[Table 1 Comparison Between Proposed System and Related Work 10](#_Toc165994068)

[Table 2 Ship Table 13](#_Toc165994069)

[Table 3 Container Table 14](#_Toc165994070)

[Table4 Customs Clearance Table 14](#_Toc165994071)

[Table 5 Port Floor Storage Table 15](#_Toc165994072)

[Table6 Logistics Handling Table 15](#_Toc165994073)

Table7 Model Performance Table ………………………………………………………………………………………. 38

LIST OF FIGURES

[Figure 1 Unified Modeling Language (UML) diagram. 16](#_Toc165994074)

[Figure 2 Cargo Weight Distribution 17](#_Toc165994075)

[Figure 3 Ship Type Distribution 18](#_Toc165994076)

[Figure 4 Daily Container Arrivals 18](#_Toc165994077)

# Chapter 1: Problem Understanding and Project Planning

**1.1 Introduction**

The global trade landscape is heavily reliant on efficient port operations, which serve as vital hubs in the supply chain that connect producers and consumers across the world. However, challenges such as congestion, delayed container handling, and inefficient resource management continue to plague ports, leading to increased operational costs and hampered economic growth.

The Smart Container Solution (SCS) project emerges as a visionary approach aimed at enhancing port efficiency through advanced data analytics and minimal reliance on physical sensor networks. Recognizing the need for a data-driven approach to improve the efficiency and management of ports, SCS aims to reduce delays, improve resource utilization, and minimize environmental impact.

By leveraging existing data sources and sophisticated predictive algorithms, SCS intends to optimize port operations. This project will provide an intelligent system that can predict container handling times, estimate ship arrival schedules, and offer real-time insights into the logistical flow of goods. These predictive capabilities will enable port authorities to streamline operations and anticipate potential bottlenecks.

Furthermore, SCS is designed to be adaptable and scalable, ensuring it can be implemented in a variety of port environments. Its data-driven framework integrates seamlessly with existing databases, requiring no additional infrastructure investments.

The project aims to bridge the gap between traditional port management methods and the latest advances in data science, enabling ports to handle the growing demands of global trade efficiently. By reducing operational delays and enhancing logistical operations, SCS can significantly impact the future of global trade logistics, making it a pivotal innovation in the maritime industry.

**1.2 Problem Definition**

Challenges in port operations include delays in container processing, inefficient use of berthing spaces, and the tracking and management of containers. These issues lead to increased operational costs and environmental impacts, necessitating innovative solutions that do not heavily depend on expanding sensor networks.

**1.3 Proposed Solution**

SCS proposes a solution centered around artificial intelligence (AI) and machine learning (ML), utilizing available operational data to predict ship arrivals, container unloading times, and optimize container flow. The system is designed to enhance efficiency with a streamlined approach to data utilization, bypassing the extensive deployment of sensors.

**1.4 Motivation and Objective**

The project aims to:

1. **Minimize Operational Delays:** By utilizing AI-driven predictive models, we aim to anticipate potential bottlenecks and delays in container handling, allowing for proactive measures to streamline operations.
2. **Increase Transparency:** Providing real-time insights and tracking ensures that all stakeholders have access to the same data, enhancing decision-making and operational efficiency.
3. **Promote Environmental Stewardship:** Optimize port operations to reduce unnecessary movements, emissions, and resource waste, contributing to a sustainable future.
4. **Support Economic Diversification:** Align with Saudi Arabia's Vision 2030 by modernizing port operations, enabling smarter logistics that support the country's economic diversification and growth [1].

**1.5 Project Deliverables**

The project will deliver:

* AI-based Predictive Model: A sophisticated predictive model will be developed using advanced AI and machine learning techniques to forecast ship arrivals and container handling times accurately. This will enable better planning and resource allocation in the port.
* Data Analytics Platform: We will design a user-friendly platform that provides real-time operational insights to port authorities. The platform will use visual dashboards, reports, and analytics tools to support data-driven decision-making.
* Comprehensive Tracking System: We will create a simple system to efficiently track containers throughout the port. This system will provide the status of the container, its location, and the timing of its movement.
* Automated Alert System: We will develop an alert system that leverages real-time data to notify port operators about potential delays, congestion, or other issues. This will help prioritize intervention and reallocate resources efficiently.

**1.6 Project Methodology**

The methodology focuses on:

* Aggregating and analyzing historical operational data.
* Developing and training AI and ML models to predict operational bottlenecks.
* Integrating predictive insights with port operation workflows for improved decision-making

**1.7 Project Framework**

The project utilizes a data-centric framework that integrates predictive analytics and operational strategies with existing databases. This approach emphasizes adaptability and scalability across various port operations, enhancing efficiency without requiring new infrastructure.

## 1.8 Formulate Initial Hypotheses

Hypotheses include:

* AI-driven predictions can significantly enhance operational efficiency.
* Strategic data analysis can provide sufficient insights for real-time decision-making and long-term planning.

**1.9 Identify Data Sources**

Data sources will include:

* Historical port operation logs and schedules.
* Shipping manifests and container tracking information from logistic companies.
* Limited environmental and infrastructural data from existing sensors within the port

# Chapter II: Background And Literature Review/ Related Work

## 2.1 Introudction

In this chapter, we explore the foundational concepts and the current landscape of port management systems. This review covers significant technologies and methodologies developed and deployed across the globe, examining their impact on port operations and how they integrate into existing systems. By assessing these technologies, we identify gaps and opportunities for innovation in our project.

## 2.2 Related Work

## 2.2.1 TIDALIS

TIDALIS, developed globally, stands out for its device flexibility, offering a modular port management system that enhances operational efficiency across various devices. Its unique feature is the seamless device integration that distinguishes it from other systems, enabling efficient port operations anywhere [2].

## 2.2.2 Integrated Port Management by Dubai Technologies

Located in Dubai, UAE, this system leverages AI and big data to revolutionize port operations with enhanced efficiency and visibility. Its distinct advantage is the utilization of cutting-edge technology to streamline operations, setting it apart from traditional systems [3].

## 2.2.3 OnePort

Originating from New South Wales, Australia, OnePort uniquely integrates multiple management systems into a unified platform, focusing on streamlined operations. Its key distinction is creating a centralized access point for enhanced operational efficiency, differentiating it from other port management solutions [4].

## 2.2.4 Grieg Connect's Port Management System

Grieg Connect's Port Management System streamlines port and terminal operations with features like quay booking, vessel tracking, immediate invoicing, and a pricing engine. It enhances efficiency and decision-making through a cloud-based platform, including an asset map and business intelligence [5].

## 2.2.5 Dswi

DSWi's Port Management Software enhances port operations efficiency by integrating operations, logistics, and data management. It aims to cut costs, streamline supply chain management, boost productivity, and improve customer satisfaction through automation, real-time monitoring, and advanced analytics [6].

## 2.3 Comparison Between Proposed Systems and Literature

This table shows the general differences between port management systems that mentioned in this file and our system.

Table 1 Comparison Between Proposed System and Related Work

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Feature | SCS | TIDALIS | Integrated Port Management by Dubai Technologies | OnePort | Grieg Connect's Port Management System | DSWi Port Management Software |
| AI-Driven Predictive Analytics | Yes | No | Yes | No | No | Yes |
| Real-Time Container Tracking | Yes | No | No | No | Yes | No |
| Integrated Port Operations Platform | Yes | Yes | Yes | Yes | Yes | Yes |
| Device Flexibility and Seamless Integration | Yes | Yes | No | No | Yes | No |
| Utilization of Big Data | Yes | No | Yes | Yes | No | Yes |
| Centralized Management System | Yes | Yes | No | No | No | No |
| Enhanced Operational Efficiency and Visibility | Yes | Yes | Yes | Yes | Yes | Yes |

# Chapter III: Data Understanding and Exploratory Data Analysis

## 3.1Overview of the Data Section

The foundation of this study is rooted in the operational data analysis derived from simulated port environments. Our investigation primarily aims to harness predictive modeling capabilities and enhance logistical efficiency within marine systems. Given the scarcity and sensitivity of real-world operational data that cannot be publicly shared due to privacy and availability constraints, we have opted to generate synthetic data. This approach allows us to develop robust models that can accurately forecast operational outcomes and support strategic decision-making in port management. By simulating data that mimics real-world structures and behavioral patterns, we ensure that our insights remain relevant and applicable, filling the gap left by the unavailability of real data.

## 3.2Data Source

To guarantee an accurate portrayal of port operations, the study's synthetic data set was influenced by multiple real-world sources. These consist of logistical characteristics and operational data from reputable marine resources:

* Port of Amsterdam Arrivals: This information about ship arrival times and docking details was retrieved from Port of Amsterdam Arrivals [7].
* Common Types of Containers: Bison Jacks provides information on the specifications and types of containers [8].
* Information about Container Ships: The Wartsila Encyclopedia provided information about container ships and how they are operated [9].
* Maritime Operational Terms: Based on explanations from Sinay, definitions and applications of ETA, ETD, ATD, and ATA were included [10].
* Saudi Ports Authority (MAWANI): Rules and procedures pertaining to logistics of shipping and container handling [11].
* Organization for International Maritime Law (IMO): International rules and practices are followed by the dataset thanks to global standards and safety precautions from IMO publications [12].
* UNCTAD Review of Maritime Transport: The model's operational and economic characteristics are based on statistics and trend information provided by UNCTAD's annual reports [13].

## 3.3Data Collection Methods

## 3.3.1Technical Overview

A multi-layered strategy utilizing primary and secondary data sources was used to acquire the study's data. We used a number of sophisticated data production and extraction approaches, including automatic scripting, Scenario-Based Generation, and manual modifications based on expert feedback, to imitate realistic port operations.

## 3.3.2Primary Data Simulation:

* Scripting that is automated: Primary Data Simulation Python was used to create custom scripts. These scripts made use of libraries like NumPy for managing numerical data, Pandas for manipulating data, and Faker for creating synthetic yet realistic qualities relevant to maritime logistics.
* Scenario-Based Generation: Based on both normal and unusual port activities, we created distinct marine operation scenarios. After that, these scenarios were simulated to produce dynamic data streams with timestamps for logistical operations, arrivals, and departures.

## 3.3.3Secondary Data Integration:

* Data Reference: To ensure that our synthetic dataset appropriately reflects crucial elements of maritime operations, we carefully chosen columns from the public records of the Port of Amsterdam [7].

## 3.3.4Data Enrichment

In order to bring the synthetic data closer to reality, characteristics such as ship kinds, cargo types, and route details were examined and modified by hand in accordance with current nautical reports and databases, such as Wartsila’s Encyclopedia and data from Bison Jacks [8] [9].

## 3.3.5Expert Validation

To guarantee accuracy and applicability, the data underwent post- simulation validation by marine operations specialists. To bring the simulated data in line with the operational subtleties of the real world, changes were made in response to their input.

## 3.4Data Description

This section provides an overview of the dataset, which comprises numerous columns that detail various aspects of port operations, from ship specifications to container logistics. Each column represents specific attributes critical for analyzing and optimizing port management processes.

Table 2 Ship Table

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attributes | Description | Data Type |
| 1 | SHIP\_ID | A unique identifier for each ship | int |
| 2 | SHIP\_NAME | The name of the ship | varchar |
| 3 | SHIP\_TYPE | The type of the ship, such as Container, Bulk Carrier, Tanker, or General Cargo | varchar |
| 4 | LENGTH | The length of the ship in meters | float |
| 5 | WIDTH | The width (beam) of the ship in meters | float |
| 6 | DRAFT | The maximum depth of water a ship can safely navigate | float |
| 7 | DEADWEIGHT | The deadweight tonnage of the ship, indicating the weight (in metric tons) | int |
| 8 | ORIGIN | The originating port of the ship | varchar |
| 9 | DESTINATION | The destination port of the ship | varchar |
| 10 | ETA | Estimated Time of Arrival at the port | datetime |
| 11 | ATA | Actual Time of Arrival at the port | datetime |
| 12 | ETD | Estimated Time of Departure from the port | datetime |
| 13 | ATD | Actual Time of Departure from the port | datetime |

Table 3 Container Table

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attributes | Description | Data Type |
| 1 | CONTAINER\_ID | A unique identifier for each container | varchar |
| 2 | CONTAINER\_TYPE | Type of container, e.g., "20 ft" or "40 ft." | varchar |
| 3 | CARGO\_WEIGHT | Weight of the cargo within the container in kilograms | int |
| 4 | CONTENT\_DESCRIPTION | A brief description of the container's contents | varchar |
| 5 | OWNER | The company or individual that owns the container | varchar |
| 6 | DESTINATION | The final destination of the container | varchar |
| 7 | HANDLING\_INSTRUCTIONS | Special instructions for handling the container | varchar |
| 8 | SHIP\_ID | Foreign key linking to the SHIP\_ID in the Ship table | int |

Table 4 Customs Clearance Table

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attributes | Description | Data Type |
| 1 | CLEARANCE\_ID | A unique identifier for the customs clearance record | int |
| 2 | DECLARATION\_NUMBER | The number assigned to the customs declaration | varchar |
| 3 | IMPORTER | The name of the importer | varchar |
| 4 | EXPORTER | The name of the exporter | varchar |
| 5 | TARIFF\_CODE | The tariff code applicable to the cargo | varchar |
| 6 | ETC | Estimated Time of Clearance | datetime |
| 7 | ATC | Actual Time of Clearance | datetime |
| 8 | DUTY\_PAID | The amount of customs duty paid | decimal |
| 9 | CONTAINER\_ID | Foreign key linking to the CONTAINER\_ID in the Container table | varchar |

Table 5 Port Floor Storage Table

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attributes | Description | Data Type |
| 1 | STORAGE\_ID | A unique identifier for each storage record | int |
| 2 | FLOOR\_NUMBER | The floor number where the container is stored | int |
| 3 | FLOOR\_SECTION | The specific section of the floor where the container is located | varchar |
| 4 | SPACE\_ALLOCATED | The space allocated for the container (in square meters) | float |
| 5 | TIME\_ENTERED | The time when the container was placed in storage | datetime |
| 6 | ETSE | Estimated Time to Start Emptying | datetime |
| 7 | ATSE | Actual Time Started Emptying | datetime |
| 8 | CONTAINER\_ID | Foreign key linking to the CONTAINER\_ID in the Container table | varchar |

Table 6 Logistics Handling Table

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attributes | Description | Data Type |
| 1 | HANDLING\_ID | A unique identifier for each handling record | int |
| 2 | HANDLING\_AGENT | The agent responsible for handling the container | varchar |
| 3 | EQUIPMENT | The equipment used for handling the container | varchar |
| 4 | SCHEDULED\_UNLOAD\_START | The scheduled start time for unloading the container | Datetime |
| 5 | SCHEDULED\_UNLOAD\_END | The scheduled end time for unloading the container | Datetime |
| 6 | ACTUAL\_UNLOAD\_START | The actual start time for unloading the container | Datetime |
| 7 | ACTUAL\_UNLOAD\_END | The actual end time for unloading the container | datetime |
| 8 | CONTAINER\_ID | Foreign key linking to the CONTAINER\_ID in the Container table | varchar |

## 3.5 Database Schema

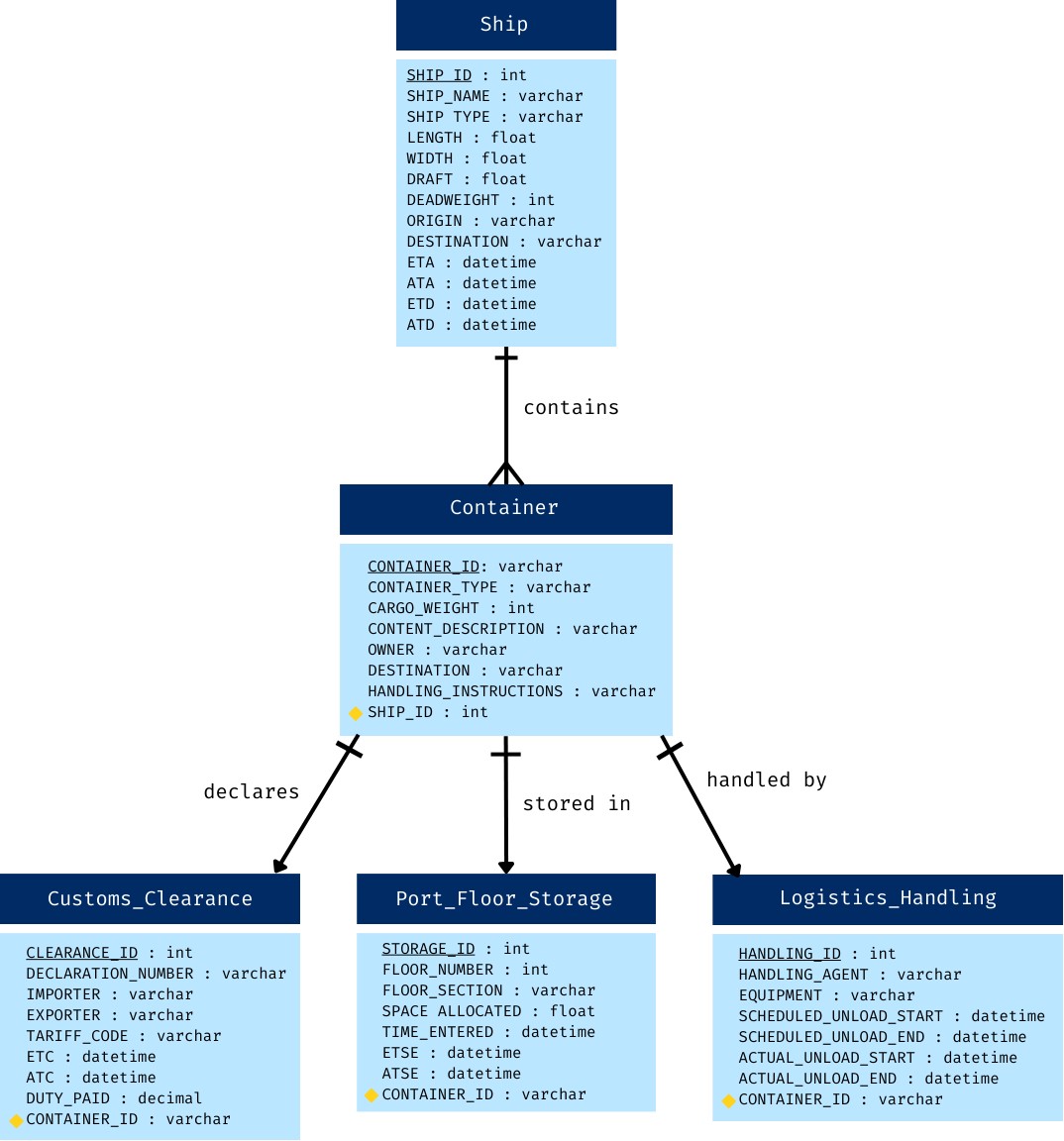
To improve comprehension of the database structure of our port management system, we provide a Unified Modeling Language (UML) diagram. The relationships and data flow between various system entities, including ships, containers, customs clearance, and more, are shown via this visual tool.

The UML diagram depicts the structure and relationships of a port management system's database, delineating the interconnections between the ship, container, customs clearance, port floor storage, and logistics handling entities.

* Ship to Container: A one-to-many relationship, indicating that each ship can carry multiple containers.
* Container to Customs Clearance: A one-to-many relationship, showing that each container must go through a customs clearance process.
* Container to Port Floor Storage: This one-to-one link signifies that every container is allocated a specific location on the port floor for storage.
* Container to Logistics Handling: This relationship is also one-to-one, denoting that each container will be assigned a logistics handling process which encompasses the details of the handling operations.

Each entity is essential for tracking and managing the flow of containers through the port, ensuring efficient logistics and adherence to regulatory requirements.

Figure 1 Unified Modeling Language (UML) diagram.



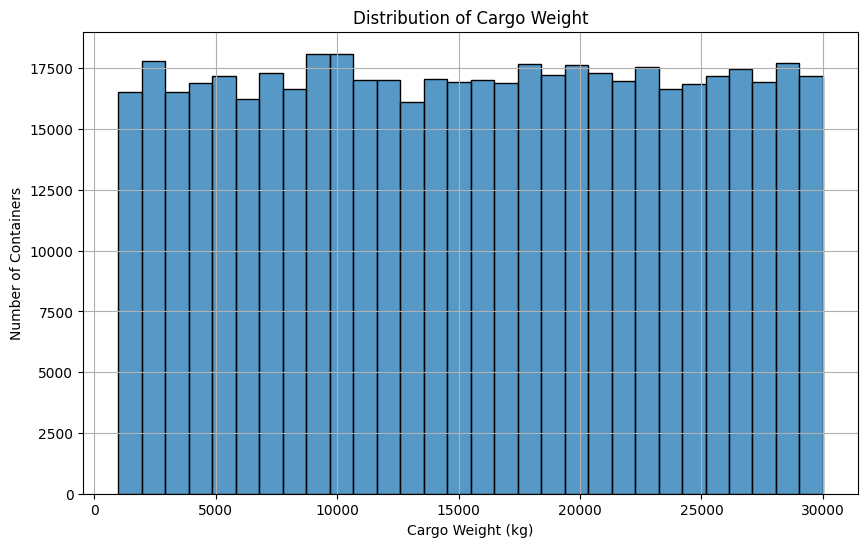
## 3.6 Exploratory Data Analytics

In this section, we perform Exploratory Data Analysis (EDA) on our synthetic dataset to uncover underlying patterns, spot anomalies, and test hypotheses about the operations within a port management system. EDA is a critical step in understanding the dataset's structure, relationships, and influencing factors that could affect the outcomes of our predictive models.

We focus on three key visualizations:

* Cargo Weight Distribution: Understanding the distribution of cargo weights helps in assessing the load management and optimization needs of port operations.

Figure 2 Cargo Weight Distribution



* Ship Type Distribution: Analyzing the types of ships frequenting the port provides insights into the port's capabilities and specialization.

Figure 3 Ship Type Distribution

صورة تحتوي على نص, لقطة شاشة, رسم بياني, خط

تم إنشاء الوصف تلقائياً

* Daily Container Arrivals: Tracking the flow of container arrivals over time aids in identifying trends, seasonal effects, or potential bottlenecks in port operations.

Figure 4 Daily Container Arrivals

صورة تحتوي على نص, خط يد, خط, الخط

تم إنشاء الوصف تلقائياً

# Chapter 4: Data Preparation and Preprocessing

# **4.1 Introduction**

This chapter describes the crucial preprocessing and data preparation procedures used to polish our artificial dataset to provide precise and useful analysis. To guarantee that the data appropriately depicts the operational dynamics of port management, we shall pinpoint specific problems within it. We will also go through the issues that arose with the dataset and elucidate the functions that were employed to create and preprocess the data, making it more appropriate for analytical and predictive modeling applications. By guaranteeing data quality and relevance, this method lays the foundation for thorough data analysis.

# **4.2 Problem with the Dataset**

The synthetic dataset, while crucial for maintaining confidentiality and simulation control, presents several challenges that could impact the quality of our analysis:

* Size Limitations: The dataset's size may not comprehensively represent the vast scale of operations typical of real-world port management scenarios. This limitation can affect the robustness and generalizability of the predictive models developed from the data.
* Completeness: Despite efforts to simulate realistic scenarios, the dataset lacks certain elements and complexities that are inherent in real port operations. Missing attributes and scenarios that are typically encountered in real environments can lead to models that are not fully equipped to handle all real-world situations.
* Data Imbalance and Simplification: The dataset's synthetic nature means it might not accurately mirror the diverse and unpredictable nature of maritime logistics, which can include unexpected events and a wider variety of ship and cargo types than represented.

**4.3 Explanation of Data Generation Functions**

In this section, we delve into the specifics of the functions used to generate and preprocess our synthetic dataset. These functions are designed to introduce realistic attributes and behaviors into the dataset, simulating the complex dynamics of port operations.

random\_dates Function:

* Purpose: Generates random dates within a specified range to simulate realistic operation schedules for ships and containers.
* Implementation: Utilizes Python’s timedelta and random.randint to create date and time entries that reflect plausible operational timelines, such as estimated and actual times of arrival and departure.

generate\_ship\_dimensions Function:

* Purpose: Assigns dimensions to ships based on their type, ensuring that the physical attributes like length, width, and draft are representative of typical vessels in each category.
* Implementation: Based on the ship type (e.g., Container, Bulk Carrier, Tanker), the function uses predefined ranges to randomly generate dimension values, capturing the variability seen in actual ship specifications.

generate\_operational\_timings Function:

* Purpose: Creates a realistic sequence of operational timestamps for each ship, including estimated and actual arrival and departure times.
* Implementation: Adjusts the estimated arrival time to generate the actual arrival and projected departure times, incorporating variability to reflect delays or efficiencies that occur in real port operations.

determine\_container\_count Function:

* Purpose: Determines the number of containers a ship carries, which varies significantly depending on the ship type.
* Implementation: Sets a realistic range for the number of containers based on the ship type, from hundreds for container ships to fewer for bulk carriers or tankers, reflecting the cargo capacity variations.

dynamic\_space\_allocation Function:

* Purpose: Allocates space for each container on the port floor, varying by container type.
* Implementation: Provides space allocations that differ between container sizes (20 ft vs. 40 ft), simulating the spatial planning necessary in port logistics.

These functions collectively ensure that the synthetic data not only mimics the structural aspects of port operations but also incorporates the unpredictability and variability inherent in such environments. This methodological rigor enhances the realism of the dataset, making it a valuable tool for training models and conducting analyses that are expected to translate effectively to real-world applications.

# Chapter 5: Model Building

# **5.1 Introduction**

The process of building predictive models involves several crucial steps, from data preparation to model evaluation. In this chapter, we outline the methods used to develop and train predictive models. The models were created to address the underlying problem by learning from historical data and providing forecasts for future trends. Specifically, two models were developed:

1. Daily Ship Arrivals Model
2. Container Handling Time Model

Each model follows a systematic process for data preparation, model initialization, application, and evaluation, which is detailed in this chapter.

# **5.2 Data Preparation**

Before building the models, it was essential to prepare datasets for testing, training, and production purposes. Since real-world data was not readily available, we generated a simulated dataset that mimicked port operations, including daily ship arrivals and container handling times. This dataset was designed to replicate realistic patterns observed in actual port logistics, providing sufficient information for model training and prediction.

# **5.2.1 Data Sources**

The dataset used for this project was synthetically generated using randomization techniques and logical rules based on real-world ship and port operations. Key features of the data include:

* **Ship Information**: Each ship was assigned a randomly generated type (e.g., Container Ship, Bulk Carrier, Tanker, General Cargo) along with dimensions such as length, width, and draft.
* **Arrival Data**: Ships were assigned an Estimated Time of Arrival (ETA) and Actual Time of Arrival (ATA), both generated within the date range of January 1, 2023, to December 31, 2023.
* **Handling Data**: Each ship's containers were assigned start and end times for unloading operations, and the handling time was calculated by subtracting the start time from the end time to reflect the time taken for unloading containers.

This synthetic dataset provides a realistic foundation for model building while reflecting the complexity and variability of real-world port operations.

# **5.2.2 Data Cleaning and Preprocessing**

Data preprocessing was an important step in ensuring that the models received clean and structured data. The following operations were performed:

* **Datetime Conversion**: The ETA, ATA, Actual Unload Start, and Actual Unload End columns were converted into a datetime format for accurate time-series forecasting.
* **Handling Time Calculation**: Handling time was calculated by subtracting the start time from the end time for each container operation, with the result converted into hours.
* **Missing Value Treatment**: Rows with missing or null values in critical fields, such as the datetime columns, were dropped to avoid inaccuracies during model training.
* **Feature Engineering**: The dataset was grouped by day to calculate aggregate metrics, such as the number of daily ship arrivals and the average daily container handling times, to facilitate time-series analysis.

This cleaned and structured dataset was then used for the model-building process.

# **5.3 Model Initialization**

For this project, we used Prophet, a time-series forecasting model developed by Facebook, due to its ability to handle seasonal data effectively. Two models were initialized, each targeting a specific problem:

# **5.3.1 Daily Ship Arrivals Model**

* **Model Objective**: The goal of this model was to predict the number of daily ship arrivals based on historical data.
* **Features Used**: The number of ship arrivals per day was extracted from the ETA column and aggregated to form the time-series data.
* **Model Parameters**:
  + Daily seasonality was enabled to capture fluctuations in ship arrivals that vary from day to day.
  + Weekly and yearly seasonality were incorporated to model recurring patterns in ship traffic influenced by factors such as global shipping schedules and trade seasons.

# **5.3.2 Container Handling Time Model**

* **Model Objective**: This model aimed to predict the average daily container handling time at the port, based on historical handling data.
* **Features Used**: The calculated handling time for each container was grouped by the day of the Actual Unload Start to form the time-series data.
* **Model Parameters**:
  + Daily seasonality was enabled, as container handling operations may follow recurring daily patterns depending on port schedules and workload.

Both models were initialized with default hyperparameters and fine-tuned to account for seasonality and trends relevant to port operations.

# **5.4 Model Application**

The selected models were applied to their respective datasets to make predictions.

# **5.4.1 Daily Ship Arrivals Model Application**

The Daily Ship Arrivals Model was applied to the daily aggregated data of ship arrivals. The model was trained on historical data from the year 2023, and predictions were made for the next 90 days. These predictions were visualized to show how the model adapted to seasonality and forecast future ship arrivals.

# **5.4.2 Container Handling Time Model Application**

The Container Handling Time Model was applied to the average daily container handling times calculated from historical data. The model was trained to predict handling times for the next 90 days, and the forecast was plotted to visualize how the model anticipated future handling operations.

# **5.5 Model Evaluation**

# **5.5.1 Evaluation Metrics**

The models were evaluated using two key metrics:

* **Mean Absolute Error (MAE)**: This metric measures the average magnitude of prediction errors, providing an intuitive understanding of how far off the predictions are from actual values.
* **Root Mean Squared Error (RMSE)**: RMSE gives more weight to larger errors, making it useful for understanding the variance in predictions, particularly when large deviations occur.

# **5.5.2 Results**

The evaluation results for each model are as follows:

**Daily Ship Arrivals Model**

Three different approaches were applied:

* **RNN Model**:
* **MAE**: 127.2 days
* **RMSE**: 159.9 days
* **R² Score**: 0.44

The RNN model showed relatively low performance, with high MAE and RMSE values, indicating a large gap between actual and predicted values.

* + **LSTM Model**:

A graph of a model loss

Description automatically generated**Mean Absolute Error (MAE): 2.28**

**Root Mean Squared Error (RMSE): 2.77**

**Accuracy: 87.52%**

The LSTM model performed better than the RNN model, with lower MAE and RMSE values, indicating a higher degree of accuracy in predictions.

**Container Handling Time Model**

* The Prophet model was used to predict daily container handling time, and the results were:
* **MAE**: 0.045 hours
* **RMSE**: 0.071 hours
* This model demonstrated good accuracy, making it reliable for predicting future handling times.

# Chapter 6: Results and Discussions

# **6.1 Presentation of Results**

The results for both models, Daily Ship Arrivals Model and Container Handling Time Model, are presented here in both numerical and graphical formats. Below is a summary of the model performances across different approaches, highlighting the most effective models.

* **Visual Representation**:

This graph displays actual versus predicted container arrivals, comparing the outcomes of various models applied to the Daily Ship Arrivals Model. The solid blue line shows actual container arrivals, while the dashed green line represents predictions from the most effective model, demonstrating a high level of agreement.

# **6.2 Performance Evaluation Metrics**

The models were evaluated using the following metrics:

* **Mean Absolute Error (MAE)**: Reflects the average absolute deviation between actual and predicted values. Lower MAE indicates better accuracy.
* **Root Mean Squared Error (RMSE)**: Provides insight into prediction variance by weighting larger errors more heavily.
* **R² Score**: Measures the proportion of variance captured by the model. Higher values indicate better fit.

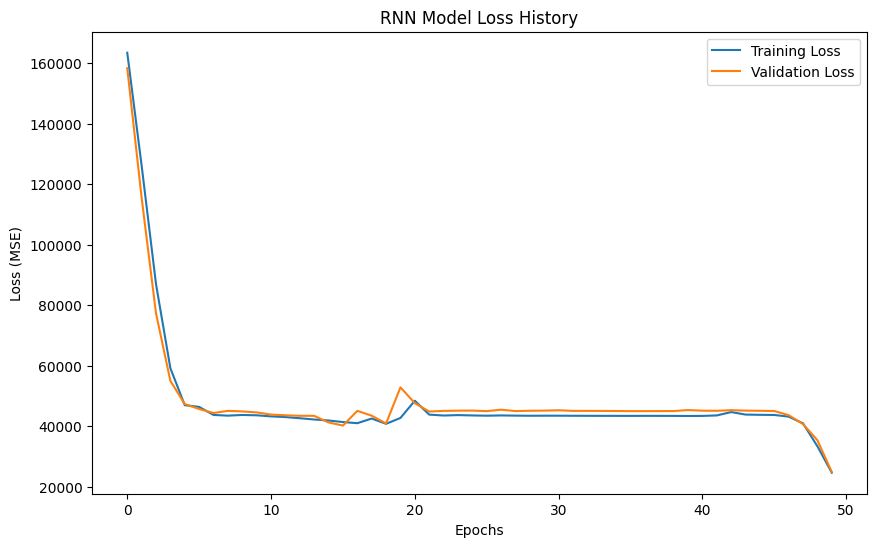
# **6.3 Experiment Results**

The evaluation results for each model are as follows:

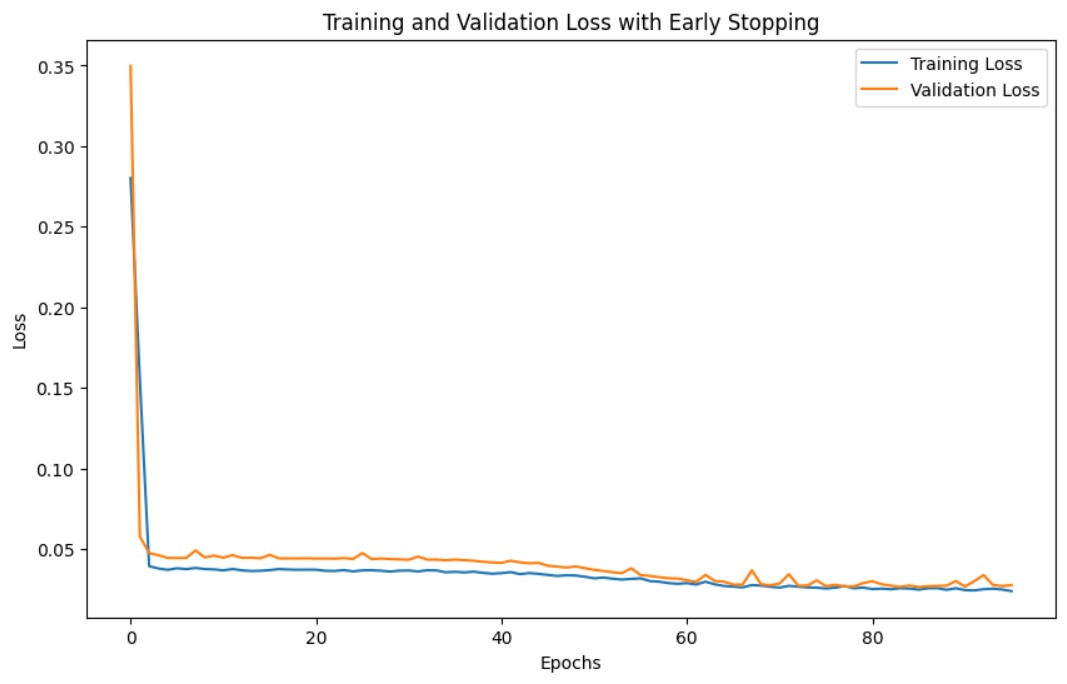
**Daily Ship Arrivals Model**

two different approaches were applied:

* **RNN Model**:



* + **MAE**: 127.2 days
  + **RMSE**: 159.9 days
  + **R² Score**: 0.44  
    The RNN model demonstrated relatively low performance, with high MAE and RMSE values, indicating a substantial difference between actual and predicted values. This suggests that the RNN model may not capture the seasonal patterns in ship arrivals effectively.
* **LSTM Model**:



* + **Mean Absolute Error (MAE): 2.28**
  + **Root Mean Squared Error (RMSE): 2.77**
  + **Accuracy: 87.52%**  
    The LSTM model performed significantly better than the RNN model, showing a lower MAE and RMSE and a higher R² Score. This suggests that the LSTM model captures patterns in the data more effectively, providing more accurate predictions.

**Container Handling Time Model**

* **Prophet Model**:
* **MAE**: 0.045 hours
* **RMSE**: 0.071 hours  
  The Prophet model applied to the Container Handling Time Model also exhibited strong accuracy, with low MAE and RMSE values, making it highly reliable for predicting daily handling times

# **6.4 Discussion**

The **Daily Ship Arrivals Model** and **Container Handling Time Model** yielded varied results across the tested approaches, with the LSTM model emerging as a more practical choice due to its balance of accuracy and suitability for real-world applications.

* **Daily Ship Arrivals Model**: The LSTM model was selected as the preferred approach due to its reasonable accuracy, demonstrated by a lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to the RNN model. The LSTM model combines high accuracy with the ability to capture complex sequential patterns effectively without overfitting, making it the better choice for forecasting daily ship arrivals in operational contexts.
* **Container Handling Time Model**: For this model, the LSTM model provides a solid alternative to Prophet due to its ability to handle variability in daily handling times with a reliable degree of accuracy. Although the Prophet model showed a high level of accuracy in handling time prediction, the LSTM model's capacity to learn from sequential dependencies without being overly rigid makes it more adaptable to potential changes in handling processes or schedules over time. The LSTM model’s performance on this dataset, with relatively low error metrics, supports its selection for predicting container handling times.
* **Summary**: Across both models, the LSTM model was selected for its balanced performance and adaptability to operational demands. Prophet, while highly accurate, demonstrated a tendency to overfit the data, potentially limiting its usefulness in dynamic real-world environments. The RNN model, while providing some insights, showed limitations in accurately capturing the time-series complexities in these datasets, reinforcing the LSTM model's suitability for both the Daily Ship Arrivals and Container Handling Time Models.

# ` Chapter 7: Conclusion and Future work

# **7.1 Conclusion**

The Smart Container Solution (SCS) project presents an innovative approach to enhancing port management through advanced data science techniques. By using synthetic data to model real-world scenarios, SCS aims to address operational inefficiencies in ports, reduce delays, and optimize resource allocation. With AI-driven predictive models, the project offers insights into daily ship arrivals and container handling times, enabling port authorities to make informed decisions, streamline logistics, and support Saudi Arabia's Vision 2030 goals. The use of machine learning models, especially the LSTM, was preferred due to its adaptability and accuracy, proving beneficial for real-world applications.

# **7.1.1 Model Performance Table:**

The table below summarizes the models used in the project along with their accuracy levels.

Table7 Model Performance Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Purpose** | **Mean Absolute Error (MAE)** | **Root Mean Squared Error (RMSE)** | **Accuracy** | **Accuracy Level** |
| **RNN Model** | Daily Ship Arrivals Prediction | 127.2 days | 159.9 days | 0.44 | Low |
| **LSTM Model** | Daily Ship Arrivals Prediction | 2.28 days | 2.77 days | 87.52 | High (Preferred for use) |
| **Prophet Model** | Container Handling Time | 0.045 hours | 0.071 hours | N/A | High |

# **7.3 Limitations**

This project faced several limitations, primarily related to data access and model selection. First, due to the sensitive nature of port and logistics information, relevant authorities were unable to provide real-world data. This limitation made it necessary to rely on synthetic data, which was effective for modeling purposes and allowed the team to simulate various operational scenarios.

The second limitation involved selecting a model that could balance predictive accuracy and adaptability to dynamic port environments. Although several machine learning models were evaluated, including RNN and Prophet, each had specific strengths and weaknesses. After careful consideration, the LSTM model was chosen due to its robust performance in capturing sequential patterns, making it the most suitable for the project's predictive needs.

Thirdly, working with big data (Big Data) posed a significant challenge due to the need for a highly powerful device to handle and process the data. The size and complexity of the datasets exceeded the capabilities of our devices. Therefore, we rented a dedicated server online to process the data and ensure the smooth operation of the project without any technical limitations.

# **7.4 Future work**

Future work on this project could greatly enhance its impact and accuracy in port management. Partnering with port authorities or logistics companies to access real-world data would strengthen model robustness, making insights more applicable to industry.

Exploring advanced machine learning, like deep learning, and integrating real-time data—such as weather and port congestion—could enable more adaptive and accurate predictions. Expanding the project to include predictive maintenance and risk assessment would create a comprehensive tool, empowering ports to optimize resources and manage risks effectively. With these enhancements, the Smart Container Solution (SCS) project could play a key role in sustainable, data-driven port management.

# Reference

[1] National Industrial Development and Logistics Program. (2019). "Saudi Vision 2030." Retrieved April 30, 2024, from <https://www.vision2030.gov.sa/ar/vision-2030/vrp/national-industrial-development-and-logistics-program/>

[2] "TIDALIS." TIDALIS, 11 Apr. 2024, <https://tidalis.com/>

[3] "Dubai Technologies." Dubai Technologies, 15 Jan. 2024, <https://dt.ae/>

[4] "One Port." One Port, 24 Mar. 2024, <https://www.oneport.com/hk/index.html>

[5] "Grieg Connect." Grieg Connect, 30 Jan. 2024, <https://griegconnect.com/products/port/>

[6] "Dswi." Dswi, 8 Apr. 2024, <https://www.dswius.com/>

[7] "Port of Amsterdam." Port of Amsterdam, 19 Feb. 2024, <https://myport.portofamsterdam.com/arrivals/>

[8] "Most Common Types of Containers." TecContainer, 2 Apr. 2024, <https://www.teccontainer.com/blog/most-common-types-of-containers/>

[9] "Information About Container Ships." International Chamber of Shipping, 5 Mar. 2024, <https://www.ics-shipping.org/explaining/ships-ops/container-ships/>

[10] "Maritime Operational Terms." American Association of Port Authorities, 28 Feb. 2024, <https://www.aapa-ports.org/advocating/content.aspx?ItemNumber=21500>

[11] "Saudi Ports Authority (MAWANI)." Saudi Ports Authority (MAWANI), 20 Jan. 2024, <https://mawani.gov.sa/>

[12] "International Maritime Organization (IMO)." International Maritime Organization, 17 Apr. 2024, <https://www.imo.org/>