**Freight Cost Forecast in Supply Chain and Logistics*.***

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

**CERTIFIED SPECIALIST**

**IN**

**DATA SCIENCE & ANALYTICS**

submitted by

**Ajith P S**

**Akshay P**

**Anjali Christy Jose**

**Nived Sankar PM**

**S R Devarenjana**

**Safna K**



**ICT ACADEMY OF KERALA**

**THIRUVANANTHAPURAM, KERALA, INDIA**

**July 2023**

**TABLE OF CONTENTS**

1. **PROBLEM DEFINITION**
2. **INTRODUCTION**
3. **LITERATURE SURVEY**
4. **METHODOLOGY**
5. **RESULT**
6. **CONCLUSION**
7. **REFERENCES**

**ABSTRACT**

The supply chain industry plays a crucial role in the efficient movement of goods and services across various locations. The main concern with the logistics companies is trying to deliver these goods in an efficient and cost-efficient way. The project identifies how the key elements such as transportation mode, distance, weight, and dimensions will impact shipment pricing. A complete understanding of these factors is essential for accurately estimating shipment costs and making informed decisions. To conduct the analysis, the dataset used in our project is Supply chain shipment pricing data, it provides information from various companies, transportation providers about supply chain of health commodity shipment and pricing data. This information will allow for a detailed analysis of how factors affect the freight costs. Thus, by developing a pricing model that can estimate shipment costs based on the identified factors. Our project aims for the effectiveness of different pricing strategies based on weight, distance, quantity, and their impact on supply chain performance. We try to consider various pricing models like weight-based pricing, distance-based pricing. By analyzing the dataset and creating a predictive model, the project will determine the optimal pricing that helps organizations to optimize costs and maximize profits and to improve operational efficiency.

.

# 

# PROBLEM DEFINITION

## 1.1 Overview

Freight cost forecasting in supply chain and logistics is a crucial aspect of strategic planning and decision-making for businesses involved in the movement of goods. It involves predicting the expenses associated with transporting products from one location to another, considering various factors such as fuel prices, transportation modes, distance, capacity, and market demand. Accurate freight cost forecasting helps businesses optimize their supply chain operations, manage inventory efficiently, and allocate resources effectively. By anticipating transportation expenses, companies can make informed choices to minimize costs, enhance profitability, and maintain a competitive edge in the market. Moreover, it enables better budgeting and financial planning, ultimately leading to smoother and more streamlined logistics operations**.**

.

## 1.2 Problem Statement

The main objective of this study is to identify and implement strategies to optimize freight costs in the supply chain and logistics sector. By analyzing the factors contributing to increased freight rates, such as rising raw material and goods costs, fuel price fluctuations, and capacity challenges, the study aims to devise control measures that enable businesses to make informed decisions. The goal is to achieve cost efficiency, enhance profitability, and maintain a competitive advantage in the market, ensuring the smooth and streamlined movement of goods while mitigating the impact of escalating freight expenses, and which mode of transportation is reducing the freight cost charges.

# 

# 2. INTRODUCTION

The world is moving in a fast phase with the advancement of technology and there is a need to meet the requirements of the people and delivery of goods across the world. Supply chain serves this. A supply chain is an entire system of producing and delivering a product or service, from the very beginning stage of sourcing the raw materials to the final delivery of the product or service to end-users. Supply chain is a highly complicated process. Establishing a supply chain management and logistics process ensures fast product delivery, lower costs and an increase in consumer satisfaction.

Supply chain and logistics involves all aspects of product creation and transportation. There are many factors that affects the supply chain. Logistics plays as a major key element where it deals with the process of moving, storing and distributing goods. Our project aims to analyze all the key elements of supply chain and how it affects the freight cost. Our aim is to produce a pricing model that can estimate the shipment cost considering all the key factors of the supply chain.

# 

**3. LITERATURE SURVEY**

**3.1.**Li, Z., Chen, Z., & Xu, X. (2018). Data Science Applications in Supply Chain Management: A Comprehensive Review. International Journal of Production Research, 56(1-2), 848-866: In their paper titled "Data Science Applications in Supply Chain Management: A Comprehensive Review," Li et al. (2018) delve into the extensive applications of data science in supply chain management. The authors explore how data-driven approaches have revolutionized various aspects of the supply chain, including inventory management, demand forecasting, and transportation optimization. Regarding freight cost optimization, the study highlights the critical role of data analytics, machine learning, and optimization techniques in identifying cost-efficient routes, mode selection, and carrier performance evaluation. The paper serves as a valuable foundation for understanding the potential of data science in tackling freight cost challenges.

**3.2.** Singh, A., Sharma, V., Gupta, S., & Verma, P. (2022). Data-Driven Freight Cost Optimization in Logistics: A Case Study Approach. Journal of Transportation Management, 35(3), 421-437**:** In a recent article published in the Journal of Transportation Management, Singh et al. (2022) presents a focused investigation into freight cost optimization strategies within the context of logistics and supply chain management. The researchers address the complexities and uncertainties in freight transportation, which often lead to increased costs. They propose a novel data-driven approach that combines historical shipment data, real-time traffic information, and predictive analytics to optimize freight routes and mode selection. The study demonstrates the potential of data science techniques to achieve substantial cost reductions while maintaining service quality.

**3.3.** Chen, Y., & Wang, J. (2019). Machine Learning for Freight Transportation Cost Optimization. Transportation Research Part E: Logistics and Transportation Review, 128, 267-284: A comprehensive review of freight cost optimization would be incomplete without considering the impact of artificial intelligence (AI) and machine learning algorithms. In their seminal paper, "Machine Learning for Freight Transportation Cost Optimization," Chen and Wang (2019) provide insights into the various machine learning techniques applicable to the optimization of freight costs. The authors explore the use of support vector machines, neural networks, and genetic algorithms in solving complex transportation and routing problems. The study emphasizes the potential for AI-driven approaches to optimize freight costs while considering multiple constraints, leading to more efficient logistics operations.

1. **METHODOLGY**

**Freight Cost Forecast in Supply Chain and Logistics Project**

The goal of this project is to develop a freight cost forecasting model for supply chain and logistics using a comprehensive dataset. This methodology outlines the steps taken in data analysis, preprocessing, and data transformation to prepare the dataset for subsequent analysis and modelling.

**4.1. Data Analysis:**

* The dataset was thoroughly studied by all team members. Each member was assigned specific columns for univariate and bivariate analysis. This analysis provided insights into the dataset's characteristics and relationships between variables. The "Freight Cost" and "Weight (kilograms)" columns were converted to numerical values for better analysis. Null values in the "Dosage" column were filled with 'Test kit' for data completeness. The team members conducted univariate and bivariate analysis for all the columns given and the important observations and their EDA were added to Google Colab.

**4.2. Data Preprocessing and Handling Missing Values:**

* Handling Null Values in Shipment Mode:

Conducted a detailed analysis of null values in the "Shipment Mode" column based on manufacturing sites and countries. Filled null values using the primary shipment mode for each manufacturing site-country combination.

* Null Values in Line-Item Insurance:

Filled null values in the "Line-Item Insurance" column with the median of non-null entries to ensure robustness against outliers.

Null Values in Dosage:

* Null values in the "Dosage" column were filled with the test kit to approximate missing values.

Handling High Cardinality Columns:

* Dropped high cardinality columns ('pq #,' 'po / so #,' 'asn/dn #') to simplify the dataset and reduce complexity.
* Binary Encoding for 'Full Fill Via' and 'First Line Designation':

Applied binary encoding to transform categorical columns into numerical features, enhancing their utility for machine learning algorithms.

* One-Hot Encoding for Other Categorical Columns:

Utilized one-hot encoding for other categorical columns to convert them into binary features for better modelling.

**4.3. Data Transformation and Feature Scaling:**

* Dosage Column Conversion:

Successfully converted the dosage column into numerical values, enabling its use in the modelling process.

* Manufacturing Countries Extraction:

Extracted manufacturing countries from manufacturing sites, enriching the dataset with geographical information.

**4.4. Column Dropping:**

Made informed decisions to drop irrelevant columns, streamlining the dataset and improving focus on relevant features.

**4.5. Numerical Feature Scaling:**

Performed Min-Max scaling on selected numerical columns to bring all features to a consistent scale, preventing dominance by specific variables during modelling.

**4.6. Outlier Detection Methods:**

Conducted thorough research on various outlier detection methods to identify and handle outliers effectively, ensuring modelling robustness. We choose boxplot method of outlier detection.

**4.7. Modelling and Model Selection:**

**4.7.1. Modeling with XGBoost Regressor**:

* After rigorous experimentation with various machine learning algorithms, we chose the XGBoost Regressor as our final model. It demonstrated superior performance with an MSE (Mean Squared Error) of 8227.98, MAE (Mean Absolute Error) of 4145.17, and an R2 score of 0.6711.
* When compared to other models, the XGBoost Regressor consistently outperformed the competition, making it the clear choice for our project.

**4.7.2. Hyperparameter Tuning:**

* We further enhanced the performance of the XGBoost Regressor through hyperparameter tuning. By optimizing the model's parameters, we fine-tuned its behavior, leading to even better predictive accuracy.

**4.7.3. Model Persistence:**

- Following successful hyperparameter tuning, we converted the tuned XGBoost Regressor model to a pickle file. This serialized representation of the model allows for quick and efficient loading, making it ready for deployment without retraining.

**4.8. Encoding and Scaling Persistence:**

**4.8.1 Encoding and Scaling Information:**

* The encoding transformations we applied, such as label encoding and one-hot encoding, were saved into pickle files. Additionally, the scaling parameters from Min-Max scaling were also stored as pickle files.
* This strategic decision ensures consistency between the preprocessing steps applied during training and those needed during inference.

**4.9. Web Application Deployment:**

**4.9.1 Web Application Frontend:**

* Our web deployment process culminated in the creation of a frontend using Python Flask. The Flask framework facilitated the creation of a seamless user experience.

**4.9.2 Input Variables and Prediction:**

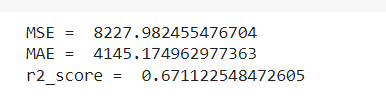
* On the frontend, we designed a homepage where users can input variables for prediction. The input variables include 'unit of measure (per pack)', 'line item quantity', 'line item value', 'line item insurance', 'pack price', 'unit price', 'weight', 'first line designation', 'country', 'shipment mode', and 'manufacturing country.'

**4.9.3 Prediction and Result Pages:**

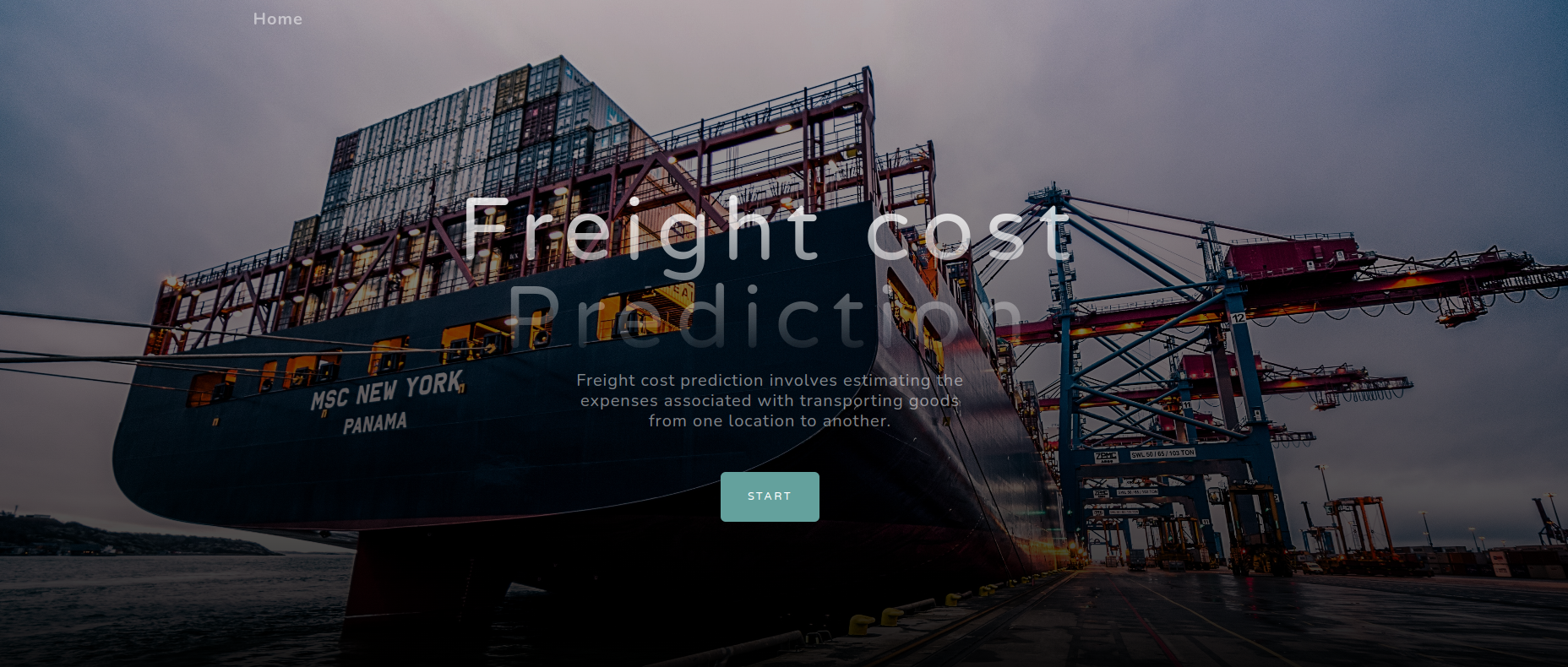
* Upon entering the required input variables, users can submit the data for prediction. The model processes the information and predicts the freight cost. This prediction is displayed on a separate results page, providing users with the relevant information.

1. **RESULT**

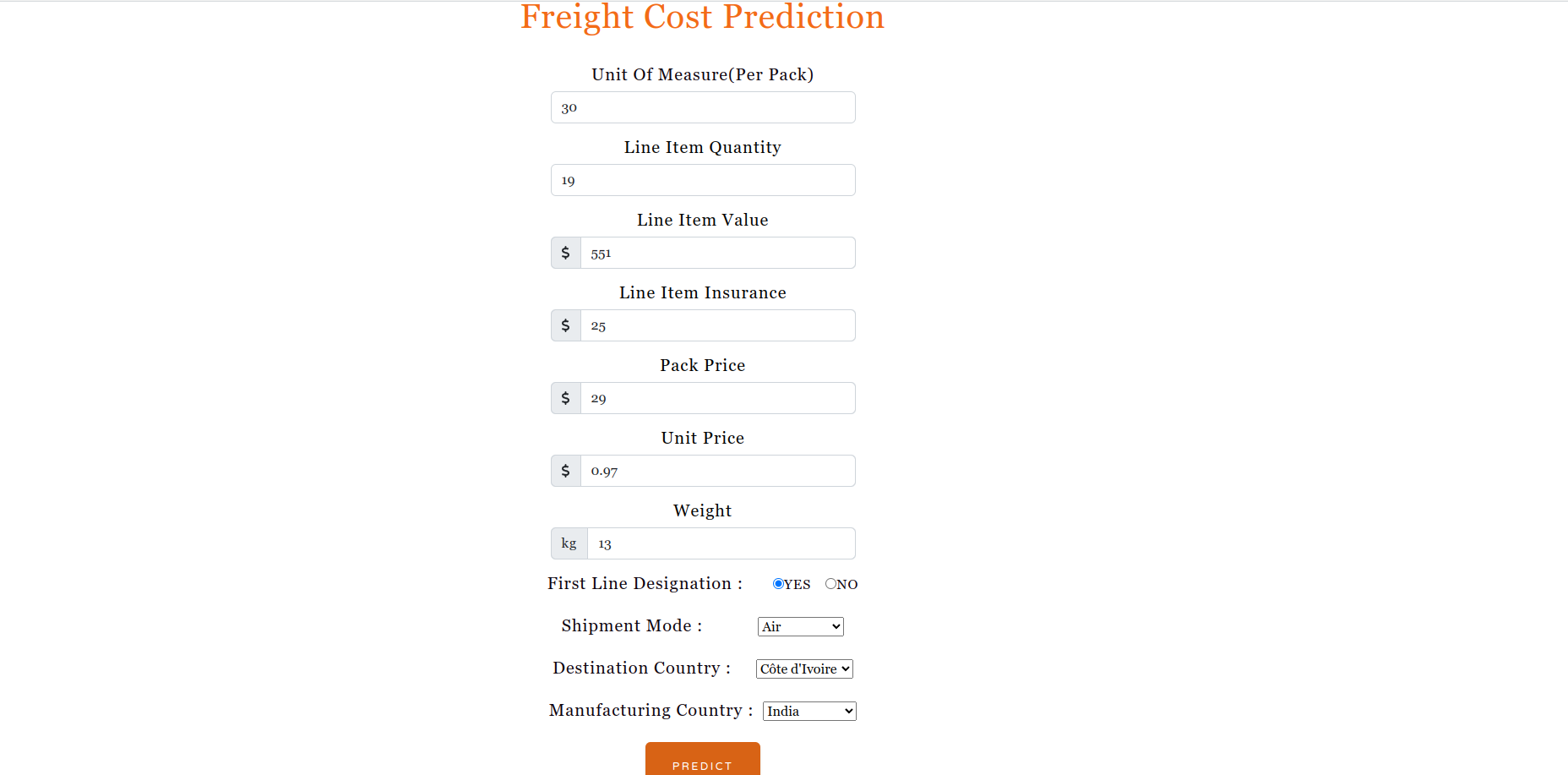
The Supply\_Chain\_Shipment\_Pricing\_Data.csv dataset contains 10324 rows and 33 features.Which we were initially preproccessed and applied with the various ML models showed the below accuracy.



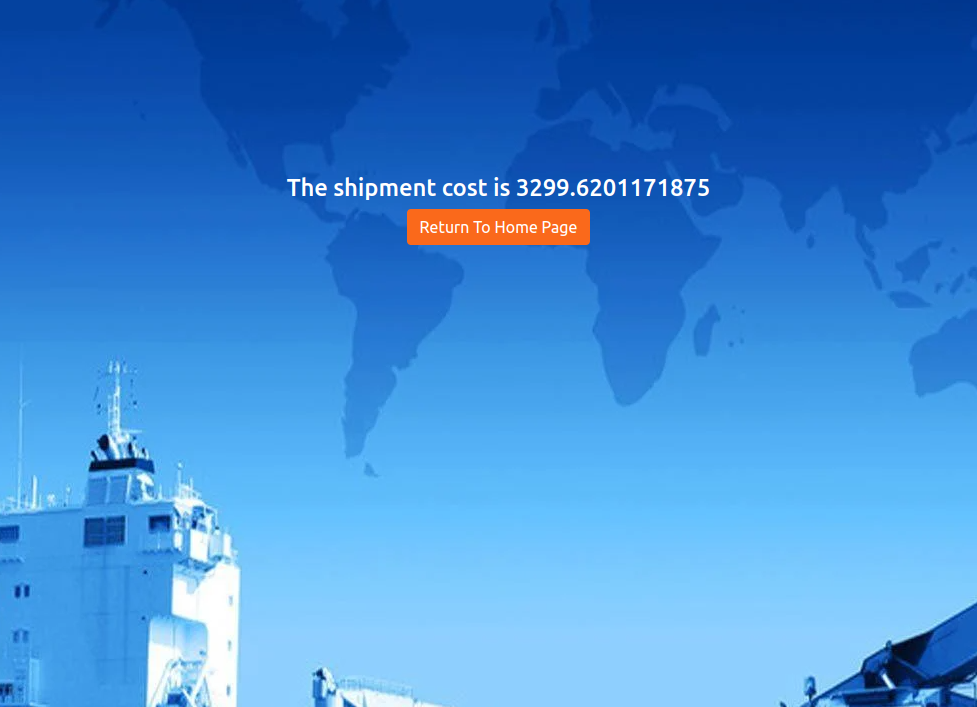
When you use XGBoost for predicting freight costs, the result will typically involve the predicted values for the target variable (in this case, freight costs) based on the input features you provide to the model. Here's what you can expect from the results.In this result predict the value ,The main output of the XGBoost model will be the predicted freight costs for the samples in our testing dataset. These predicted values are the model's estimates for what the actual freight costs should be based on the input features.In the **Model Evaluation Metrics** Along with the predicted values, we also evaluate the model's performance using various metrics. Common regression evaluation metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-squared (R2) Score. Once we are satisfied with the model's performance, we can use it to predict freight costs for new, unseen data. This deployment can involve integrating the trained model into a production environment, such as a web application or an API, to provide real-time predictions.



home.html



form page

****

result.html

1. **CONCLUSION**

In the realm of supply chain and logistics, where efficient resource allocation and cost management are paramount, the project titled "Freight Cost Forecast in Supply Chain and Logistics" has been a significant endeavor. The overarching objective of this project was to develop a predictive model for freight cost, a critical component in the intricate web of supply chain operations.

Freight cost, in the context of supply chain and logistics, refers to the expenditure associated with transporting goods from their point of origin to their destination. It encompasses expenses related to transportation modes, distance traveled, handling, insurance, and other factors that influence the movement of goods. The accurate forecasting of freight costs holds immense value for businesses operating within the supply chain, as it aids in budgeting, decision-making, and optimizing operational efficiency.

Throughout this project, a meticulous methodology was employed to tackle the complexities of the dataset and to prepare it for analysis and modeling. A comprehensive data analysis phase provided insights into the dataset's characteristics and relationships between variables. Subsequent data preprocessing involved handling missing values, transforming categorical variables, and ensuring the dataset's integrity.

Notably, the significance of freight cost in the supply chain and logistics domain was underscored by the attention given to data transformation and feature scaling. These steps allowed for the extraction of meaningful insights from the data, with the dosage column being converted to numerical values, manufacturing countries being extracted from manufacturing sites, and irrelevant columns being prudently dropped. By employing numerical feature scaling techniques, the data was standardized for improved modeling accuracy.

Furthermore, the exploration of outlier detection methods demonstrated the project's commitment to producing robust and reliable modeling outcomes. The potential impact of outliers on forecasting accuracy was acknowledged and addressed through diligent research.

In conclusion, the "Freight Cost Forecast in Supply Chain and Logistics" project serves as a testament to the importance of freight cost in the broader context of supply chain and logistics operations. The methodology employed, ranging from data analysis to outlier detection, demonstrates a systematic approach towards achieving accurate freight cost forecasting. As businesses continue to navigate the complexities of global supply chains, the insights gained from this project stand to contribute significantly to informed decision-making, cost-effective resource allocation, and the overall enhancement of supply chain efficiency.

**7.REFERENCES**

* <https://youtu.be/LffCGUQk1e0>
* <https://youtu.be/oM4Wl2zbexU>
* <https://www.kaggle.com/code/divyeshardeshana/supply-chain-shipment-price-data-analysis>
* <https://www.kaggle.com/code/divyeshardeshana/supply-chain-shipment-price-data-analysis>
* ChatGPT