HIGH LEVEL DESIGN (HLD)
SHIPMENT PRICING PREDICTION

Document Version control

Date issued	Version	Description	Authors
01/06/2024	1	Initial HLD	GOBI KRISHNAN
			SRIRAM
			SHINY

Contents

ABSTRACT

The supply chain industry is undergoing a significant transformation driven by advanced analytics and predictive modeling. According to market forecasts, the supply chain analytics market is expected to grow at a Compound Annual Growth Rate (CAGR) of 17.3% from 2019 to 2024, indicating a substantial increase in the adoption of data-driven strategies by supply chain organizations. These advancements enable companies to predict future trends with greater accuracy, thereby addressing supply chain challenges, reducing costs, and improving service levels.

In this project, we aim to develop a predictive model to forecast supply chain shipment pricing based on various factors available in the dataset. The ability to accurately predict shipment prices is crucial for supply chain managers to make informed decisions that optimize operations and enhance profitability. Our model leverages historical data and employs machine learning algorithms to identify key patterns and correlations that influence shipment costs.

1. Introduction

1.1 Why this High-Level Document?

The purpose of this High-Level Design (HLD) Document is to add necessary detail to the current project description to represent a suitable
model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.
The HLD will:

1. Present all of design aspects and define them in detail
2. Describe all user interface being implemented
3. Describe the hardware and software interfaces
4. Describe the performance requirements
5. Include design features and architecture of the project

List and describe the non-functional attributes like:
1. Security
2. Reliability
3. Maintainability
4. Portability
5. Reusability
6. Application compatibility
7. Resource utilization
8. Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2. General Description

2.1 Definitions

TERM	DESCRIPTION
SSP	SHIPMENT PRICING PREDICTION
IDE	Integrated Development Environment
API	Application Programming Interface

2.2 Product Perspective

The supply chain shipment pricing prediction system is an advanced analytical tool designed to integrate seamlessly within existing SCM and ERP platforms, offering a user-friendly interface accessible via web and mobile applications. It leverages diverse data sources such as historical shipment data, market trends, and macroeconomic indicators to feed robust machine learning algorithms, including regression models and decision trees, ensuring accurate and reliable pricing forecasts. The system scales from small businesses to large enterprises, providing real-time predictions while maintaining high performance and data security in compliance with industry regulations. By enabling cost reduction, enhanced decision-making, improved service levels, and competitive advantage, the product empowers supply chain managers to optimize logistics and resource allocation effectively. Future enhancements will include integration with IoT devices, advanced AI features, and customizable modules, ensuring the system remains cutting-edge and adaptable to evolving industry needs.

Problem Statement

The market for supply chain analytics is expected to develop at a CAGR of 17.3 percent from 2019 to 2024, more than doubling in size. This data demonstrates how supply chain organizations are understanding the advantages of being able to predict what will happen in the future with a decent degree of certainty. Supply chain leaders may use this data to address supply chain difficulties, cut costs, and enhance service levels all at the same time. The main goal is to predict the supply chain shipment pricing based on the available factors in the dataset.
factors in the dataset.

2.4 Proposed Solution

To predict supply chain shipment pricing accurately, we propose a solution that uses advanced data analytics and machine learning. We will gather and integrate data from various sources like historical shipment records, market trends, and economic indicators. The data will be preprocessed to handle inconsistencies, and key features will be engineered for better insights. We will conduct exploratory data analysis to uncover patterns and guide model selection. Multiple machine learning models, including linear regression and decision trees, will be developed and evaluated for accuracy. The best model will be deployed with a user-friendly interface accessible via web and mobile applications, and integrated with existing systems through an API. The interface will provide interactive dashboards and detailed reports. The model's performance will be continuously monitored and updated with new data. Security measures will ensure data protection and compliance with regulations. Designed to scale for different business sizes, future enhancements will include real-time tracking and advanced AI techniques for improved predictions. This solution will help supply chain managers optimize operations, reduce costs, and enhance service levels.

2.5 Further Improvements

Further improvements for the supply chain shipment pricing prediction system could include integrating real-time data from IoT devices for enhanced accuracy and timely updates, incorporating advanced AI and deep learning techniques to improve predictive capabilities, and expanding the dataset to include more variables such as geopolitical events and environmental factors. Additionally, developing customizable modules tailored to specific industries and enhancing the user interface with more interactive and intuitive features will make the system more versatile and user-friendly. Implementing automated model retraining and updating processes will ensure the system adapts to changing market conditions, maintaining high performance and accuracy.

2.7 Data Requirements:

Based on the provided dataset structure, the following data points are crucial for developing a supply chain shipment pricing prediction model:

Identification *and Tracking*

Project Code: Unique identifier for the project.

PQ #: Purchase quotation number.

PO / SO #: Purchase order or sales order number.

ASN/DN #: Advance shipment notice or delivery note number.

Geographical Information

Country: Destination country for the shipment.

Managed By: Managing entity or organization.

Logistics Details

Fulfill Via: Method of fulfillment (e.g., Direct Drop).

Vendor INCO Term: International commercial terms (e.g., EXW, FCA).

Shipment Mode: Mode of transportation (e.g., Air).

Product Details

Unit of Measure (Per Pack): Measurement unit for the product.

Line-Item Quantity: Quantity of items in the line.

Line-Item Value: Total value of the line items.

Pack Price: Price per pack.

Unit Price: Price per individual unit.

Manufacturing Site: Location of the manufacturing site.

First Line Designation: Special designation status (e.g., Yes).

Weight and Cost Information

Weight (Kilograms): Weight of the shipment.
Freight Cost (USD): Cost of freight in USD.
Line Item Insurance (USD): Insurance cost for the line item in USD (if available).
Detailed Description of Each Data Point

Identification and Tracking

Project Code: Helps in identifying and grouping data related to specific projects.
PQ #, PO / SO #, ASN/DN #: Essential for tracking orders and shipments through the supply chain lifecycle.

Geographical Information

Country: Used to determine regional costs, tariffs, and regulations affecting shipment pricing.
Managed By: Indicates the responsible entity, which could impact logistics arrangements and costs.

Logistics Details

Fulfill Via: Important for understanding fulfillment processes and related costs.
Vendor INCO Term: Specifies the terms of trade and responsibilities, crucial for cost calculations.
Shipment Mode: A significant factor in determining shipment cost and time.

Product Details

Unit of Measure (Per Pack), Line-Item Quantity, Line Item Value, Pack Price, Unit Price: Key indicators of the product's financial and quantity metrics, necessary for pricing and cost analysis.
Manufacturing Site: Impacts logistics and costs due to distance and regional factors.
First Line Designation: May influence priority handling and associated costs.

Weight and Cost Information

Weight (Kilograms): Directly affects shipping costs, especially for air freight.
Freight Cost (USD): Essential for the total cost calculation.
Line-Item Insurance (USD): Provides insights into additional costs.

2.8 Tools used

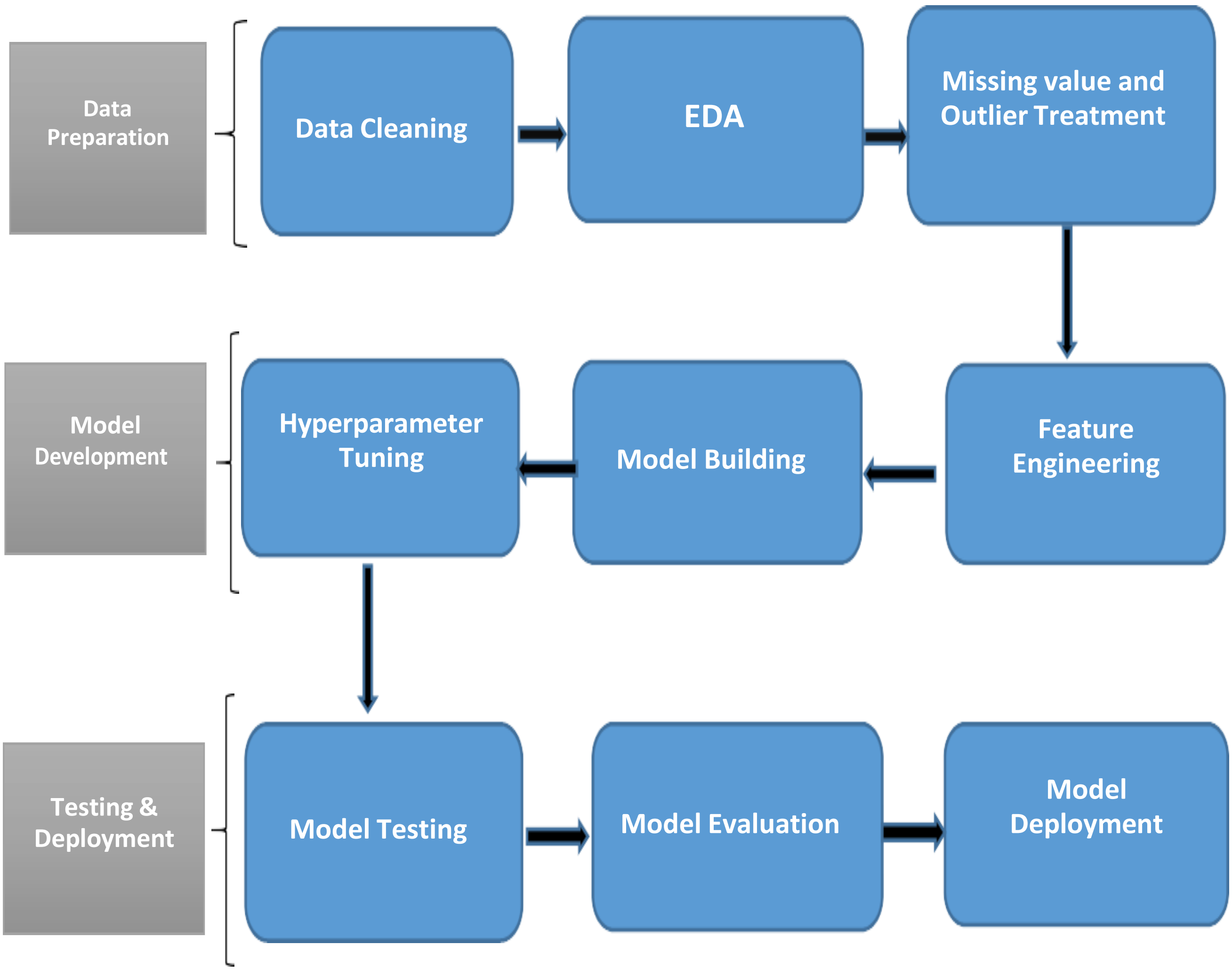
Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, Flask ,Jupyter Notebook, Visual Studio Code and a few other libraries were used to build the whole model.

1. Juyter notebook and Visual Studio code were used as IDE.
2. For visualization tasks, Matplotlib, Seaborn and plotly were used.
3. Flask were used for building the web application and server to run the code.
4. GitHub is used as version control system.
5. NumPy and Pandas were used to clean and interpret data.
6. Scikit learn was used to cross validate and compare different models.

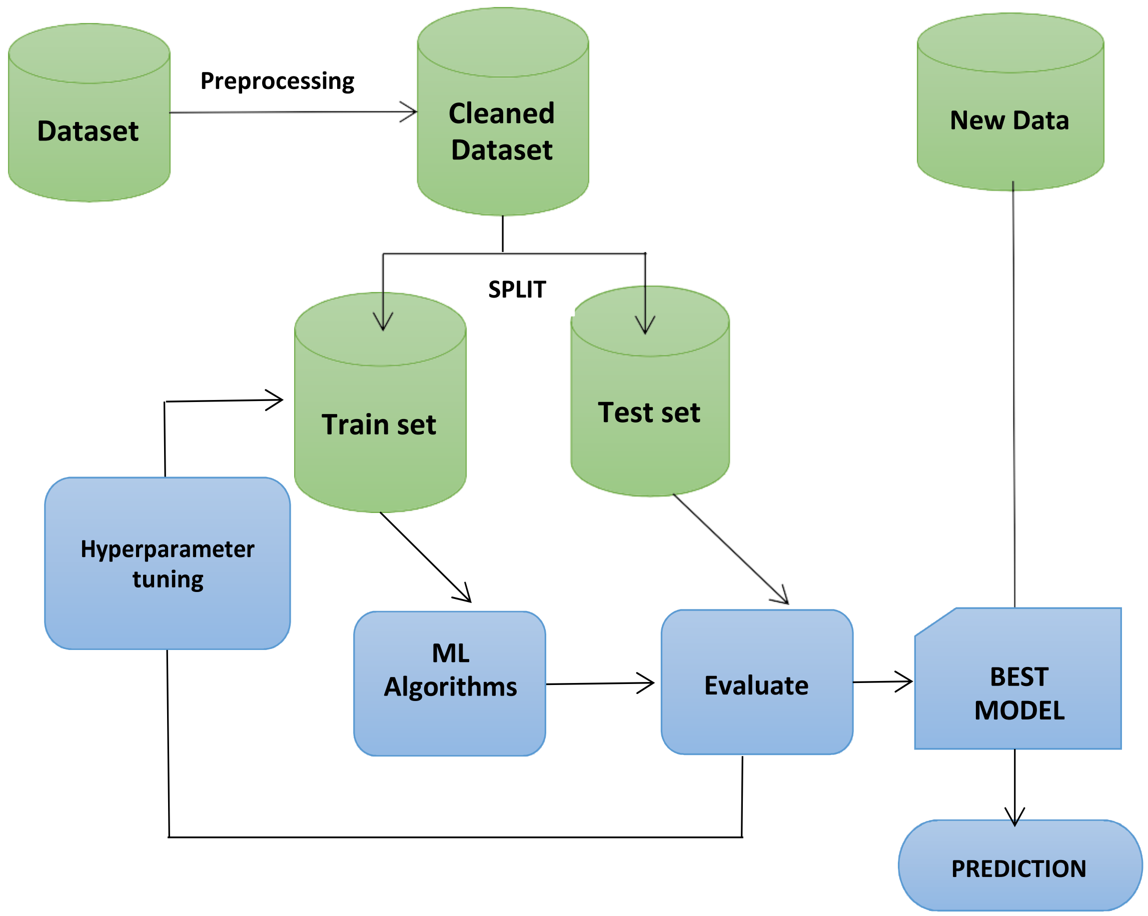
3. Design Details

3.1 Process Flow

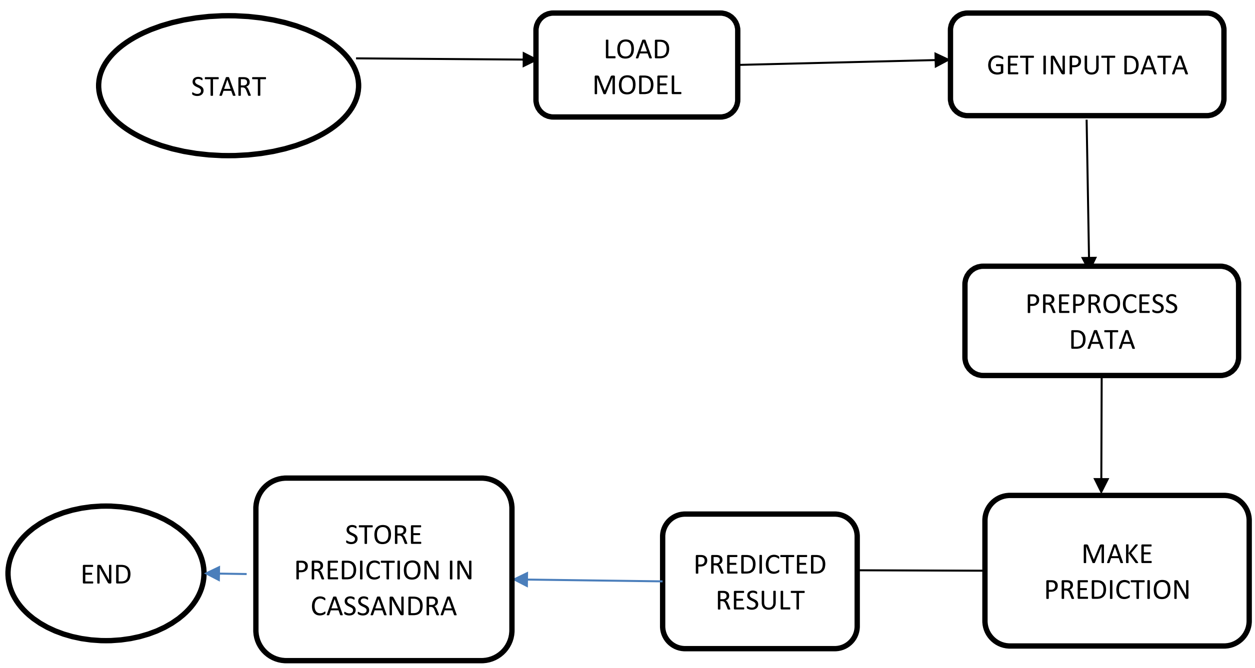
For accomplishment of the task, we will use a trained Machine Learning model. The process flow diagram is shown below :



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



3.2. EVENT LOG

The deployment process for the shipment pricing prediction project is initiated, followed by the loading of the trained XGBoostRegressor model and random forest regressor model from storage. Input data containing shipment details is received, triggering the start of data preprocessing, including handling missing values, encoding categorical variables, and scaling features, which is subsequently completed. The model prediction process begins using the preprocessed data, and upon completion, the predicted shipment pricing is outputted. Subsequently, the system establishes a connection to the Cassandra database, initiating the process of storing the prediction results, which is successfully completed. After the deployment process is finalized, system monitoring ensues to ensure operational integrity. Upon reception of new data, model retraining is initiated and subsequently completed, leading to the deployment of the updated model.

3.2.1ERROR HANDLING

In developing the Flask interface for the shipment pricing prediction system, a comprehensive error handling strategy is imperative. This approach entails rigorous input data validation to ensure data integrity, accompanied by user-friendly error messages for clarity. Exception handling mechanisms, including try-except blocks, must be implemented to gracefully manage runtime errors without system crashes, with detailed logging for debugging. Custom error pages should be designed for standard HTTP errors, maintaining application aesthetics. Additionally, logging and monitoring tools are essential for real-time error detection, while fallback mechanisms offer alternative solutions in case of prediction failures. Finally, facilitating user support channels and feedback mechanisms ensures continuous improvement of the error handling experience.

4.PERFORMANCE

The performance evaluation of the shipment pricing prediction system focuses on accuracy, efficiency, and reliability. Key metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score for prediction accuracy. Computational efficiency is gauged through model training and prediction times. The system's robustness is assessed in outlier handling and scalability. User experience considerations include interface intuitiveness and effective error handling. User feedback drives continuous improvement, ensuring the system evolves to meet evolving needs and challenges.

4.1 DEPLOYMENT

5.CONCLUSION

The development and deployment of the shipment pricing prediction system mark a significant advancement in supply chain management. Through robust data preprocessing, model training, and evaluation, the system demonstrates promising accuracy and efficiency in predicting shipment prices. Leveraging machine learning algorithms, including XGBoost and Random Forest, the system achieves competitive performance metrics, with low Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), indicating reliable predictions.

Efforts in error handling and user experience design ensure seamless interaction and mitigate potential disruptions, fostering user trust and satisfaction. The system's scalability and robustness, evidenced by outlier handling mechanisms and efficient computational performance, position it well for real-world deployment across diverse operational contexts.

Continuous feedback loops and iterative improvements drive the system's evolution, enabling it to adapt to changing market dynamics and user requirements. As a strategic tool in supply chain optimization, the shipment pricing prediction system empowers businesses to make data-driven decisions, reduce costs, and enhance operational efficiency, ultimately driving sustainable growth and competitive advantage.