Architecture Design

SHIPMENT PRICING PREDICTION

Project By: Apurv Bhusari

Document Version control

Date issued	Version	Description	Author
22/02/2023	1	Initial Architecture	Apurv Bhusari

Contents

ABSTRACT

The goal of this project is to develop a model that can accurately predict supply chain shipment pricing based on a variety of factors. The market for supply chain analytics is expected to experience significant growth over the next few years, as organizations increasingly recognize the benefits of being able to forecast future events with a high degree of certainty. By accurately predicting supply chain pricing, supply chain leaders can address challenges, reduce costs, and improve service levels.

1. Introduction

1.1 What is Architecture design document?

Any software needs the architectural design to represents the design of software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures.

The goal of Architecture Design (AD) is to give the internal design of the actual program code for the `Shipment Pricing Prediction`. AD describes the class diagrams with the methods and relation between classes and program specification

1.2 Scope

Architecture Design (AD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software, architecture, source code, and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work. And the complete workflow.

1.3 Constraints

We only predict the expected estimated Shipment cost for customers based on some past consignment shipment records the chances are many of our predicted values may vary to some extent.

2. Technical Specification

2.1 Dataset

The dataset is comprised of 10,324 records with 33 attributes. The data is in structured format.The data is a stored locally and is exported as a CSV file to be used for Data Pre-processing and Model Training.

The dataset looks as follows:

ID	Project Code	PQ #	PO / SO #	ASN/DN #	Country	Managed By	Fulfill Via	Vendor INCO Term	Shipment Mode	PQ First Sent to Client Date	PO Sent to Vendor Date	Scheduled Delivery Date	Delivered to Client Date	Delivery Recorded Date	Product Group	Sub Classification
1	100-CI-T01	Pre-PQ Process	SCMS-4	ASN-8	Côte d'Ivoire	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	2-Jun-06	2-Jun-06	2-Jun-06	HRDT	HIV test
3	108-VN-T01	Pre-PQ Process	SCMS-13	ASN-85	Vietnam	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	14-Nov-06	14-Nov-06	14-Nov-06	ARV	Pediatric
4	100-CI-T01	Pre-PQ Process	SCMS-20	ASN-14	Côte d'Ivoire	PMO - US	Direct Drop	FCA	Air	Pre-PQ Process	Date Not Captured	27-Aug-06	27-Aug-06	27-Aug-06	HRDT	HIV test
15	108-VN-T01	Pre-PQ Process	SCMS-78	ASN-50	Vietnam	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	1-Sep-06	1-Sep-06	1-Sep-06	ARV	Adult
16	108-VN-T01	Pre-PQ Process	SCMS-81	ASN-55	Vietnam	PMO - US	Direct Drop	EXW	Air	Pre-PQ Process	Date Not Captured	11-Aug-06	11-Aug-06	11-Aug-06	ARV	Adult

	Vendor	Item Description	Molecule/Test Type	Brand	Dosage	Dosage Form	Unit of Measure (Per Pack)	Line Item Quantity	Line Item Value	Pack Price	Unit Price	Manufacturing Site	First Line Designation	Weight (Kilograms)
	RANBAXY Fine Chemicals LTD.	HIV, Reveal G3 Rapid HIV-1 Antibody Test, 30 T...	HIV, Reveal G3 Rapid HIV-1 Antibody Test	Reveal	NaN	Test kit	30	19	551.0	29.00	0.97	Ranbaxy Fine Chemicals LTD	Yes	13
	Aurobindo Pharma Limited	Nevirapine 10mg/ml, oral suspension, Bottle, 2...	Nevirapine	Generic	10mg/ml	Oral suspension	240	1000	6200.0	6.20	0.03	Aurobindo Unit III, India	Yes	358
	Abbott GmbH & Co. KG	HIV 1/2, Determine Complete HIV Kit, 100 Tests	HIV 1/2, Determine Complete HIV Kit	Determine	NaN	Test kit	100	500	40000.0	80.00	0.80	ABBVIE GmbH & Co.KG Wiesbaden	Yes	171
	SUN PHARMACEUTICAL INDUSTRIES LTD (RANBAXY LAB...	Lamivudine 150mg, tablets, 60 Tabs	Lamivudine	Generic	150mg	Tablet	60	31920	127360.8	3.99	0.07	Ranbaxy, Paonta Shahib, India	Yes	1855
	Aurobindo Pharma Limited	Stavudine 30mg, capsules, 60 Caps	Stavudine	Generic	30mg	Capsule	60	38000	121600.0	3.20	0.05	Aurobindo Unit III, India	Yes	7590

2.2 Logging

We should be able to log every activity done by the user

- ✓ The system identifies at which step logging require.
- ✓ The system should be able to log each and every system flow.
- ✓ The system should not be hung even after using so much logging. Logging is just because we can easily debug issuing so logging is mandatory to do.

2.3 Deployment

To deploy a model, you can use the following steps:

- ✓ Save the trained model as a pickle file using Python's pickle library.
- ✓ Create a Flask app in Python, which will act as the server for your model.
- ✓ Define the routes for the Flask app, which will determine the behavior of the server when it receives different HTTP requests.
- ✓ In the routes, you can load the pickle file and use it to make predictions based on the input received in the request.
- ✓ You can also create HTML templates to display the results of the predictions on a website.
- ✓ Test the Flask app using Postman or a similar API testing tool to ensure it is working correctly.

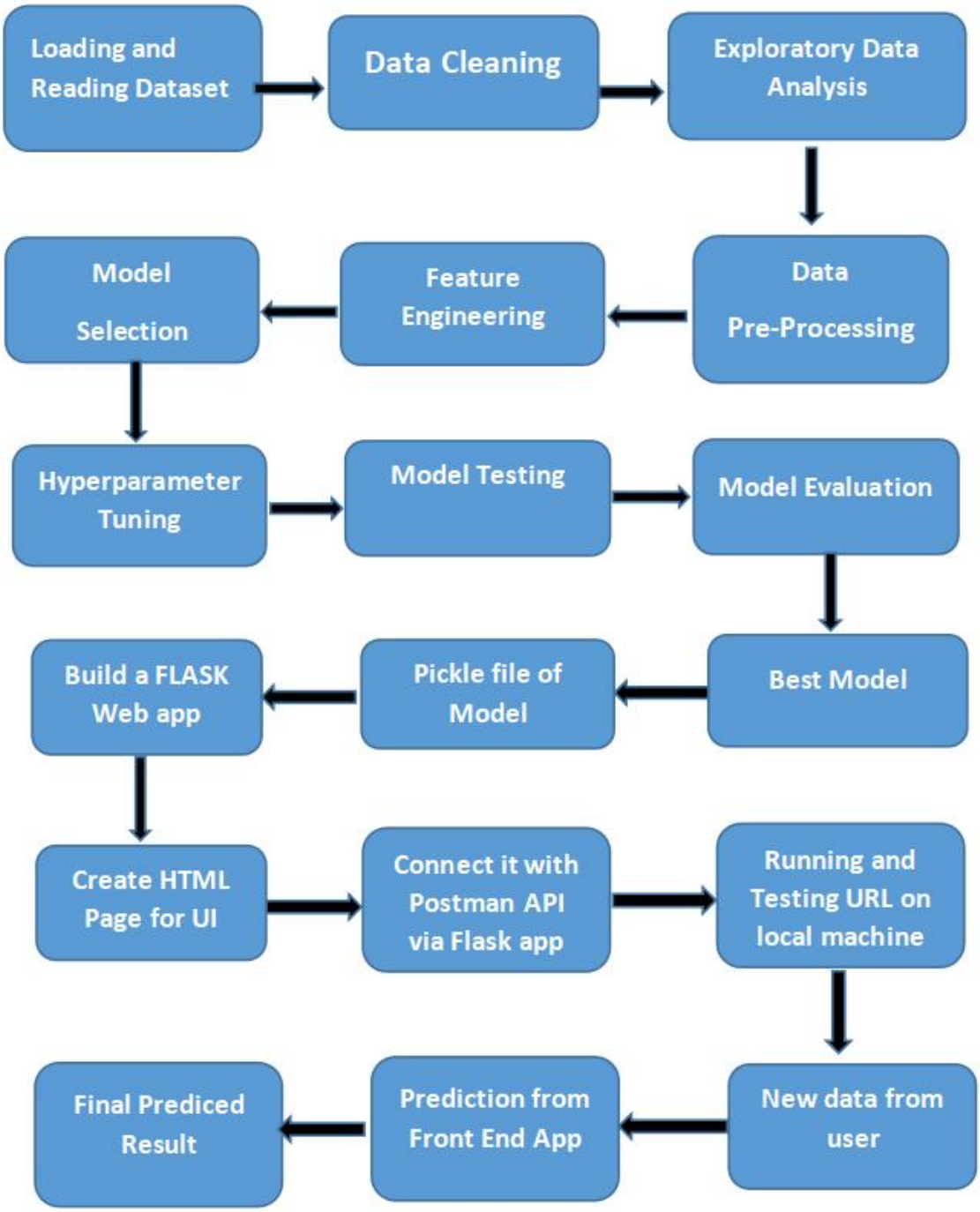
3. Technology Stack

Front End	HTML/CSS
Backend	Python/ Flask
Deployment	Local Machine Host

4. Proposed Solution

- ✓ To accurately estimate the Shipment cost, we will utilize a combination of EDA and machine learning techniques.
- ✓ The client will input relevant information through our web application, which will be passed to the backend for validation and preprocessing.
- ✓ From there, the data will be fed into a machine learning model that has been optimized through hyper parameter tuning.
- ✓ The model's output will be the final prediction for the Shipment cost, which will be displayed to the client through the web application.

5. Architecture



5.1 Data Gathering

Data source: https://www.kaggle.com/datasets/divyeshardeshana/supply-chain-shipment-pricing-data

Data Dictionary and Source: https://data.usaid.gov/HIV-AIDS/Supply-Chain-Shipment-Pricing-Data/a3rc-nmf6

5.2 Data Cleaning

Data cleaning is the process of detecting and correcting errors, inconsistencies, or missing values in a dataset. It is an essential step in data pre-processing and is often the most time-consuming part of the data analysis process.

5.3 Exploratory Data Analysis

Visualized the relationship between the dependent and independent features. Also checked relationship between independent features to get more insights about the data.

5.4 Feature Engineering

After pre-processing standard scalar is performed to scale down all the numeric features.

Even one hot encoding is also performed to convert the categorical features into numerical features

5.5 Model Selection

In This step we apply different machine learning algorithms to our processed data and select the model with best results for further Hyper-parameter tuning to make the best possible model that can be made for accurate and correct prediction.

Model building is an iterative process, and the specific steps and techniques used will depend on the nature of the data and the goals of the analysis

5.6 Model Evaluation

We evaluate our regression model based on performance metrics such as R-square,Adjusted R-square,Root mean square values.We can also compare the model's predictions to the actual values using a scatter plot.

5.7 Deployment Process

To deploy a model, you can use the following steps:

- ✓ Save the trained model as a pickle file using Python's pickle library.
- ✓ Create a Flask app in Python, which will act as the server for your model.
- ✓ Define the routes for the Flask app, which will determine the behavior of the server when it receives different HTTP requests.
- ✓ In the routes, you can load the pickle file and use it to make predictions based on the input received in the request.
- ✓ You can also create HTML templates to display the results of the predictions on a website.

5.8 GitHub

The whole project directory will be pushed into the GitHub repository

6. User Input / Output Workflow.