# Objective

- ✓ Development of a predictive a model that can accurately predict supply chain shipment pricing based on a variety of factors.

- ✓ By accurately predicting supply chain pricing, supply chain leaders can address challenges, reduce costs, and improve service levels.

Benefits

- **Improved cost management**: By accurately predicting supply chain pricing, supply chain leaders can better manage costs and identify opportunities for cost savings. This can help to improve the overall financial performance of the organization.

- **Enhanced decision-making**: With accurate pricing predictions, supply chain leaders can make more informed decisions about which suppliers to use, which transportation modes to employ, and which inventory management strategies to adopt.

- **Increased efficiency**: By better predicting demand and identifying bottlenecks in the supply chain, organizations can improve their overall efficiency and reduce waste.

- **Improved customer satisfaction**: Accurate pricing predictions can help organizations to offer more competitive prices to their customers, leading to improved customer satisfaction.

Architecture
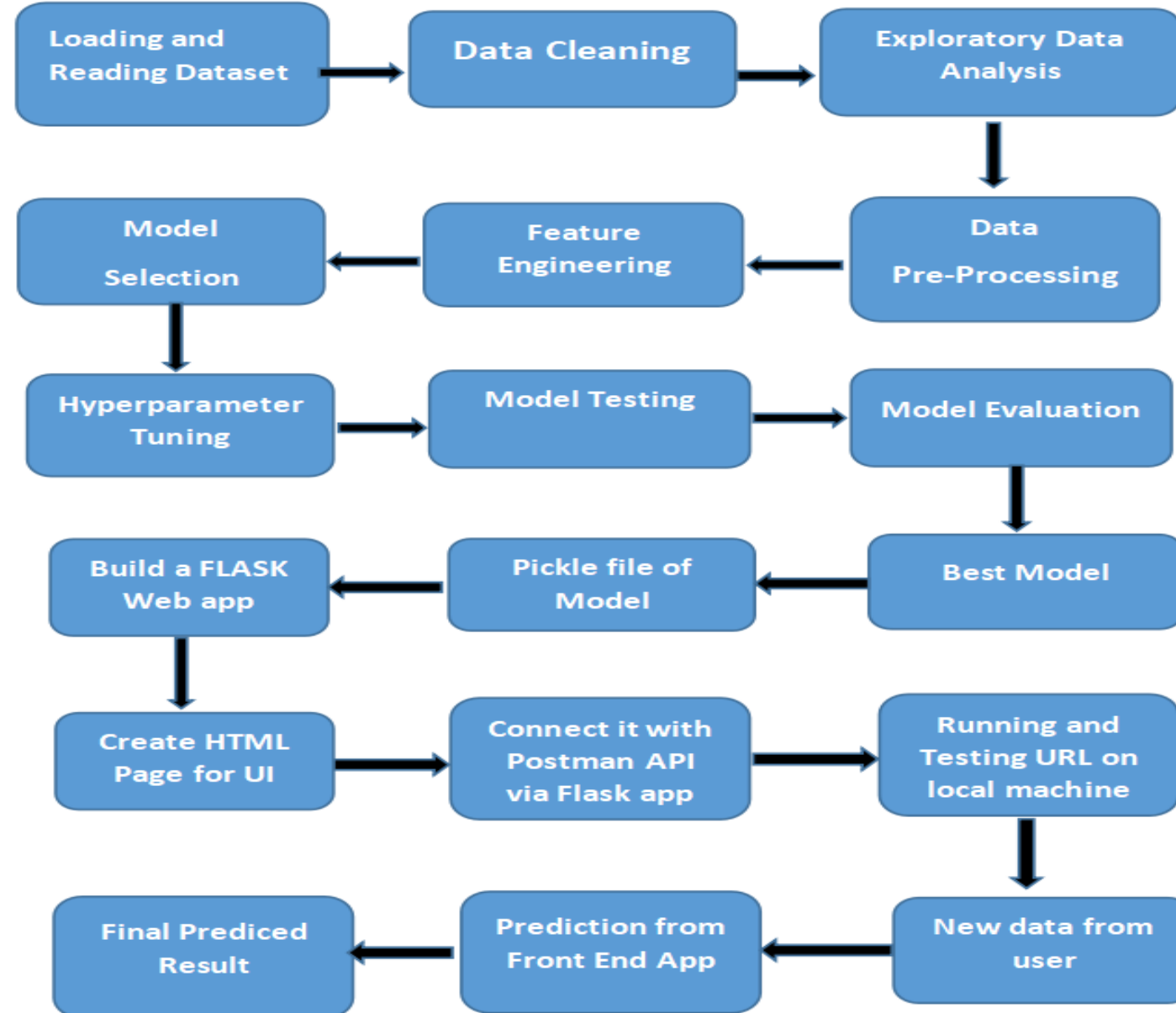
Process Flow
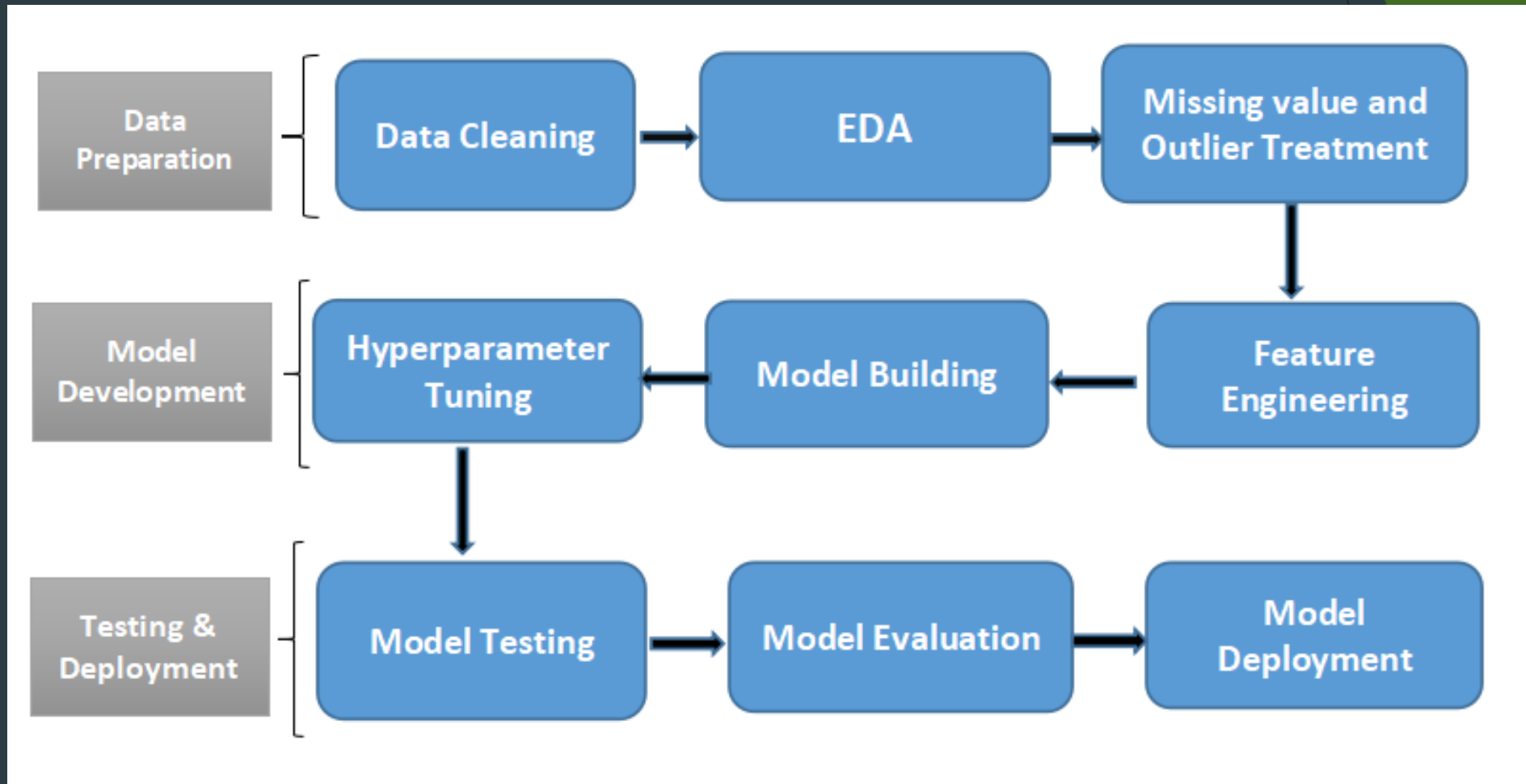
Data Validation and Data Transformation

➢ Missing Values – All the missing values were replaced with the value being repeated the most number of times.

➢ Numerical Columns - All the numerical features were standardized, preventing any data leakage by using data pipelines.

➢ Categorical Columns - Either label encoding or one hot encoding was done to treat the categorical features

Model Training

Export Data :

➢ The accumulated data is exported to python and read using pandas.

Data Preprocessing:

➢ Performing EDA to get insight of data like identifying distribution ,outliers treatment,trend among data etc.

➢ Check for null values in the columns. If present impute the null values.

➢ Encode the categorical values with numeric values.

➢ Perform Standard Scalar to scale down the values.

➢ Create New features in accordance with business domain helpful for model building process.

Model Selection

➢ Different Regression models were compared and hyperparamater tuning was done via gridsearchcv on the best performing one that is Light-gbm Regressor.

Prediction

➢ The model is made in such a way to maximise the results and also other performance metrics so that the predictions are as accurate as possible

➢ The training R-square for our model is 0.998273 and test R-square is 0.991598.

➢ Features such are 'Days to Process','Line Item Insurance','Shipment Mode', 'Freight Cost' are of importance.

Q & A

Q1) What's the source of data?

The main source is kaggle and for the prediction the data for training is provided by the client in the form answers to certain questions asked which the user has to input.

Q 2) What was the type of data?

The data was the combination of numerical and Categorical values.

Q 3) What's the complete flow you followed in this Project?

Refer slide number 4 and 5 for better understanding

Q 4) What techniques were you using for data pre-processing?

- ▶ Removing unwanted attributes
- ▶ Visualizing relation of independent variables with each other and output variables
- ▶ Checking and changing Distribution of continuous values
- ▶ Outliers Treatment
- ▶ Cleaning data and imputing if null values are present.
- ▶ Converting categorical data into numeric values.
- ▶ Scaling the data

Q 5) How training was done or what models were used?

▶ First, we started with data cleaning, EDA and feature engineering

▶ Then, outliers and ambiguities were removed from the data and categorical features data transformation was applied for categorical columns like one hot encoding, label encoding, etc.

▶ Before training the model the dataset is divided into the training set and testing set.

▶ The scaling was performed on the training and testing set.

▶ The categorical columns were converted into numeric values.

▶ We train the different models using a training set by using multiple regression algorithms like Linear Regression, CART,Random Forest,ANN,XGBoost,Light-GBM and Grid Search CV for best parameters.

▶ Then we test each model with a test set, we will find the accuracy of train and test predictions using evaluation metrics like RMSE (Root mean squared error) and r2_score (R-squared).

▶ The model which gives close train and test accuracy with the least RMSE train test difference is chosen.

Q 6) How Prediction was done?

Some questions were asked to the client like Shipment Mode,Manufacturing Site,Unit Price etc and

the responses are taken as inputs which are then feeded to the model as a single test case and the

predictions are then returned on the clients screen after a interval of two seconds in which the model

processes the input data to get the output.

Q 7) What are the different stages of deployment?

To deploy a model, we used the following steps:

1.Save the trained model as a pickle file using Python's pickle library.

2.Create a Flask app in Python, which will act as the server for your model.

3.Define the routes for the Flask app, which will determine the behavior of the server when it receives different HTTP requests.

4.In the routes, we loaded the pickle file and use it to make predictions based on the input received in the request.

5.We created HTML templates to display the results of the predictions on a website.

6.Test the Flask app using Postman or a similar API testing tool to ensure it is working correctly.