



Consignment Pricing Prediction (Machine Learning Project)

Sooraj M.A

Akhila Saj

Objective:

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

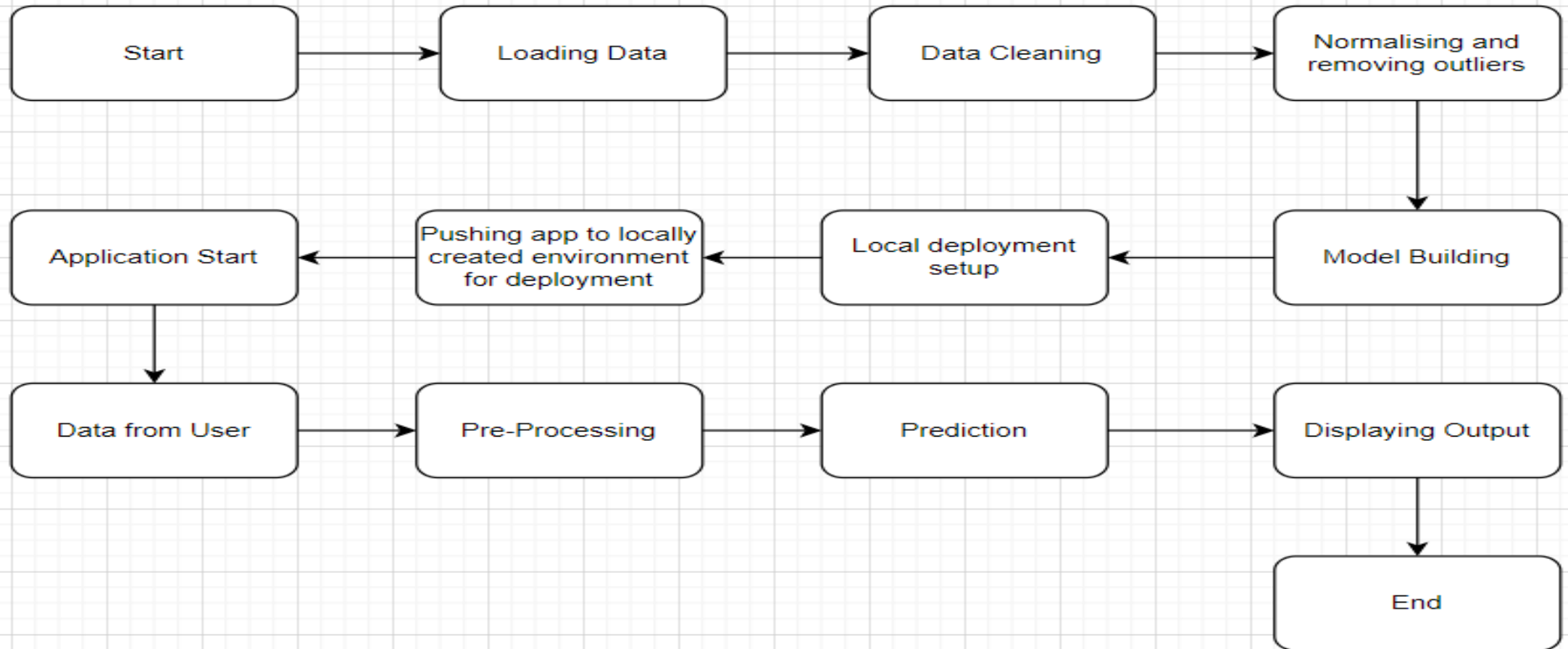
Benefits:

- Consignment Price Prediction
- Identify the supply chain difficulties
- Enhance the service levels

Data Sharing Agreement

- Sample file name (SCMS_Delivery_History_Dataset)
- Length of date stamp (8 digits)
- Length of time stamp (digits)
- Number of columns
- Column names
- Column data type

Architecture



Data Validation

- File name Validation: File name validation as per the DSA.
- Name and Number of columns: It will check for number of columns and name of columns as per schema file.
- Data types of columns: The datatype of the columns is given in the schema file.
- Null values: If any column contain null values, then that respective value should be imputed.

Model Training

1. Exploratory Data Analysis :

- Performing EDA to get insights from the data like distributions, trends etc.

2. Data Pre-processing:

- Check for null values in the columns. If present impute the null values.
- Encode the categorical values using Categorical encoder.

- Removing the outliers.
- Checking co-relation and removing features with less co-relation with the target variable.
- Perform min-max scalar to scale down the values.

Model Selection:

- The data is trained with different models. We have tried Linear Regressor, Xgboost regressor, Random Forest Regressor and CAT-BOOST regressor. Among the four models Random Forest regressor has the highest r^2 -score and adjusted r^2 -score. So this model is considered.

Prediction

- The testing files are shared and we perform the same validation operations, data transformation and data insertion on them.
- Pre-processing is done on this files before prediction.

Q&A:

1. Explain about the project.

Consignment price prediction is a machine learning web app designed to predict the consignment prediction. The model will predict the cost of a consignment based on input features provided by the user.

2. What is the source of data?

The dataset is taken from Kaggle problem statement.

3. What was the type of data?

The data was the combination of numerical and categorical values.

4. What's the complete flow you followed in this project?

Refer slide 5 for better understanding.

5. What techniques were used for data pre-processing?

- Converting categorical values using Label encoder.
- Removing unwanted attributes.
- Removing the features which is having less co-relation with the target variable.
- Imputing the missing values
- Converting the continuous values into normal distribution.
- Removing the outliers
- Scaling using Min-Max scalar

6 . What models were used

- Linear Regressor
- XgBoost Regressor
- Random Forest Regressor
- CATBOOST Regressor

7. How Training was done?

The pre-processed data is divided into train and test dataset. This train dataset is used for training the model and the test dataset is used for the validation of the model.

8. How Prediction was done?

The user need to provide the required details, then the model will take input and provide prediction output which will be shown to the user through the UI.

9. What is the size of the data?

3.698 Mb

10. Which tools were used for the implementation of this model?

- IDE : PyCharm, Jupyter notebook
- Deployment : Local host
- Front end : HTML