HIGH LEVEL DESIGN (HLD)

BIG MART SALES PREDICTION

Revision Number : 3.0

Last Date of Revision:02/12/2021

Document Version Control

Date Issued	Version	Description	Author
01/09/2021	1.0	1. Document Created	Shubham Bhatt
15/09/2021	2.0	1. There are many blank pages 2. Explanation for Design Details in required 3. HLD Content- Add more information about its uses and application. 4. Naming for Images is required	Shubham Bhatt
28/09/2021	3.0	1. Changes according to your deployment	Shubham Bhatt

Content

Abstract

Bigmart is a big supermarket chain, with stores all around the country , the aim is to predict sales of the store (that is, Item Outlet Sales) as a function of parameters. We perform analysis using 12 different features like supermarket type, item weight etc. The features differ with respect to the store type, items selected , item weight etc.. The dataset comprises 8523 samples and 12 features, aiming to predict the sales of bigmart store. It can also be used as a multi-class classification problem if the response is rounded to the nearest integer.

1. INTRODUCTION

1.1 Why This High-Level Design Document

The purpose of this High Level Design (HLD) Document is to add the 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.

1.2 Scope

The HLD documentation presents the structure of the system, such as the 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.

1.3 Definitions

Term	Definitions
IOS	Item outlet Sales
Database	Collection of all the information monitored by this system
IDE	Integrated Development Environment
AWS	Amazon Web Service

1.4. Overview

The HLD will:

• present all of the design aspects and define them in detail

• describe the user interface being implemented

• describe the hardware and software interfaces

• describe the performance requirements

• include design features and the architecture of the project

1.5 Uses

- This document is designed to help in operational requirement and can be used as a reference manual for how the modules interact.

- HLD briefly describes about the platforms/products/services/processes, flow of traffic that it depends on and includes any important changes that need to be made to them.

- HLD is the input for creating the LLD (Low Level Design) since the key communication items are displayed in HLD which are then converted to detailed communication in LLD, showing connectivity and physical level

1.6 Application

1] Website Development

2] Application Development

3] Data Science Project

2. GENERAL DESCRIPTION

2.1 Product Perspective

The Bigmart Sales solution system is machine learning model which helps to find out the co-relation between the input variable and output variable of sales of the outlet. Each row of dataset represents the one outlet parameter. So our task is to predict the output on the basis of given input variable.

2.2 Problem Statement

The effect of twelve input variables (Item Identifier, Item weight , Item Visibility , Item Fat Content, Outlet Identifier , Item MRP, Outlet size etc.) on output variable, namely Item outlet Sales, of the outlet in big mart is calculated using a statistical machine learning framework. We have to use a number of classical and non-parametric statistical analytic tools to carefully analyze the strength of each input variable's correlation with each of the output variables in order to discover the most strongly associated input variables. We need to estimate the sales; we can compare a traditional linear regression approach to a sophisticated state-of-the-art nonlinear nonparametric method, random forests.

2.3 Proposed Solution

The target feature Item Outlet Sale can be predicted using regression algorithms. So from the above problem statement we can find out that sales from the given input attribute. The sales will help you to find out the different parameter that is required for the estimation in terms of the prediction. In the given model we take data from the client whatever we required for the predicting the sales.

2.4 Technical Requirement

In this Project the requirements to get outlet sales through various platform. For that, in this project we are going to use different technologies. Here are some requirements for this project.

- Model should be exposed through API or User Interface, so that anyone can test model.
- Model should be deployed on cloud (Azure, AWS, and GCP) for the public used.

- Cassandra database should be integrated in this project for any kind of user input.

2.5 Data Requirements

Data Requirement completely depends on our problem.

- For training and testing the model, we are using Bigmart sales prediction dataset.
- From user we are taking following input :
  - Item Weight :- Float Value By User
  - Item Fat content :- Integer Value By User
  - Item Visibility :- Integer Value By User
  - Item Type :- Integer Value By User
  - Item MRP :- Float Value By User
  - Outlet size :- Integer Value By User
  - Outlet Location Type :- 1,2,3,4
  - Years Established :- Integer Value By User
  - Supermarket type :- 0,1

2.6 Tools Used

Python programming language and frameworks such as NumPy, Pandas and Scikit-learn are used to build the whole model.

- Jupyter Notebook is used as IDE
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- AWS used for deployment of the model.
- Cassandra is used to insert, delete, retrieve and update the database.
- Front end development is done using HTML/CSS.
- Python flask is used for backend development
- • GitHub is used as version control system.

Figure No. 1 Entire Technology That Is Used in Project

2.7 Constraints

The Big Mart Sales prediction system must be user friendly, errors free and users should not be required to know any of the back-end working.

3. DESIGN DETAILS

3.1 Process Flow

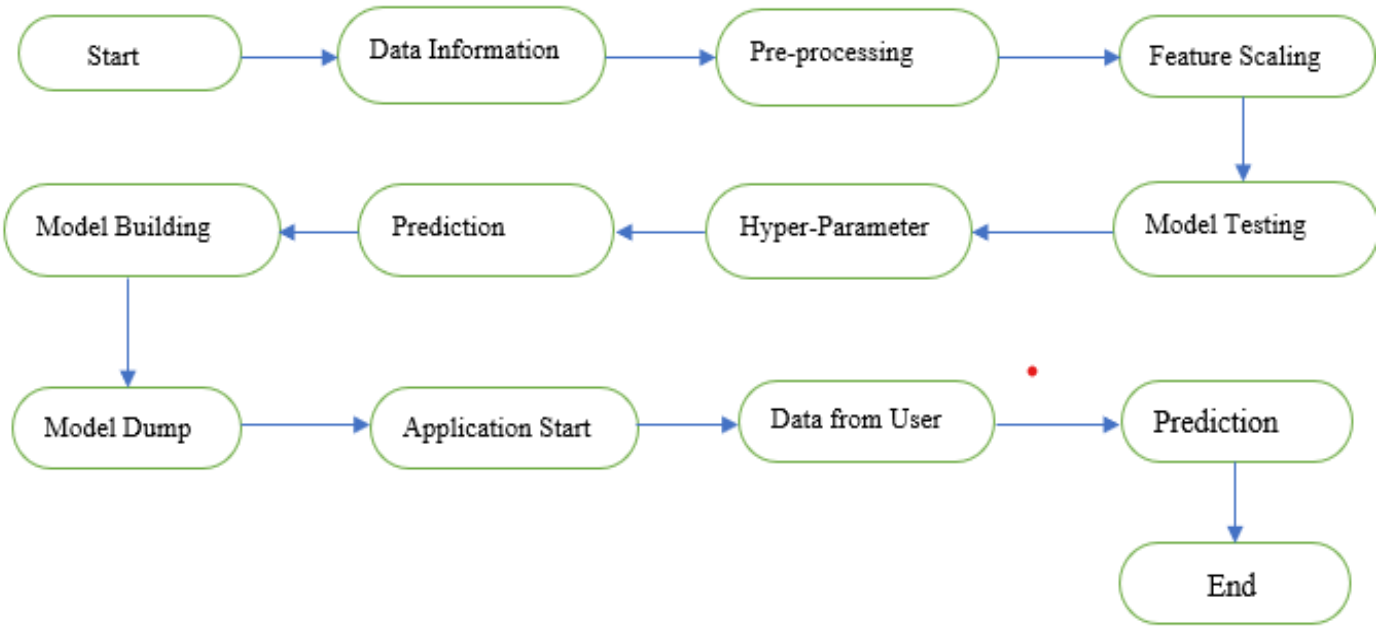

Figure No. 2 The Process Flow

3.2 Deployment Process

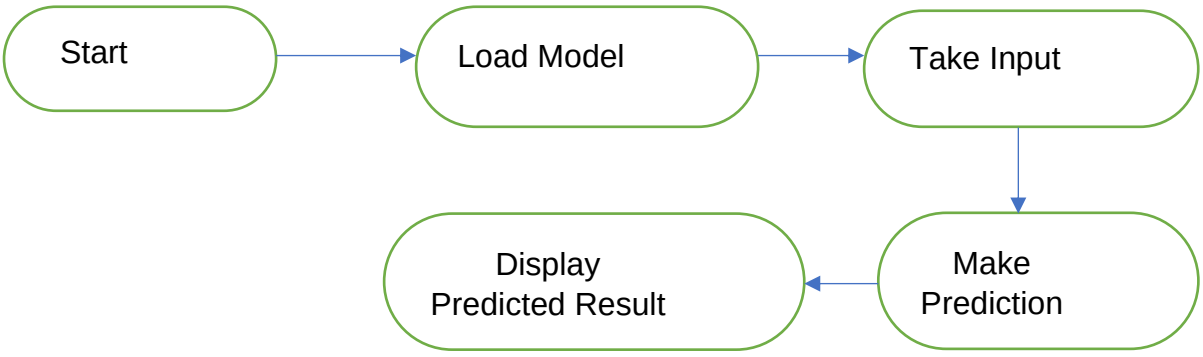

Figure No. 3 The Deployment Process

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong?

An error will be defined as anything that falls outside the normal and intended usage.

4 . Performance

Solution of Big mart Sales prediction is used to predict the Sales, so it should be as accurate as possible so that it should give as much as possible accurate prediction. That's why before building this model we followed complete process of Machine Learning. Here are summary of complete process:

1. First we cleaned our dataset properly by removing all null value and duplicate value present in dataset and if not present leave the dataset as it is.
2. Second we have to handle the outlier form the dataset so that it can't effect on the accuracy. So in our dataset there is no outlier present.
3. Then after that we scale down the whole dataset in between 0 & 1 with help of Min Max scalar.
4. Then we split the whole data set into train-test split with test size will be 30% of the whole dataset.
5. After performing above step we ready for model training. In this step, I trained my dataset on different Regression algorithm such as Linear, Lasso, Ridge, Elastic Net, and Random Forest etc.
6. After that I applied hyper-parameter tuning on all models which I have described above.
7. After that I saved my model in pickle file format for model deployment.
8. After that my model was ready to deploy.

4.1 Reusability

We have done programming of this project in such way that it should be reusable. So that anyone can add and contribute without facing problem.

4.2 Application Compatibility

The different component for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is job of the Python to ensure proper transfer of information.

4.3 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

We have deployed the project on Local cloud service only.

4.5 User Interface

We have created an UI for user by using HTML and CSS.

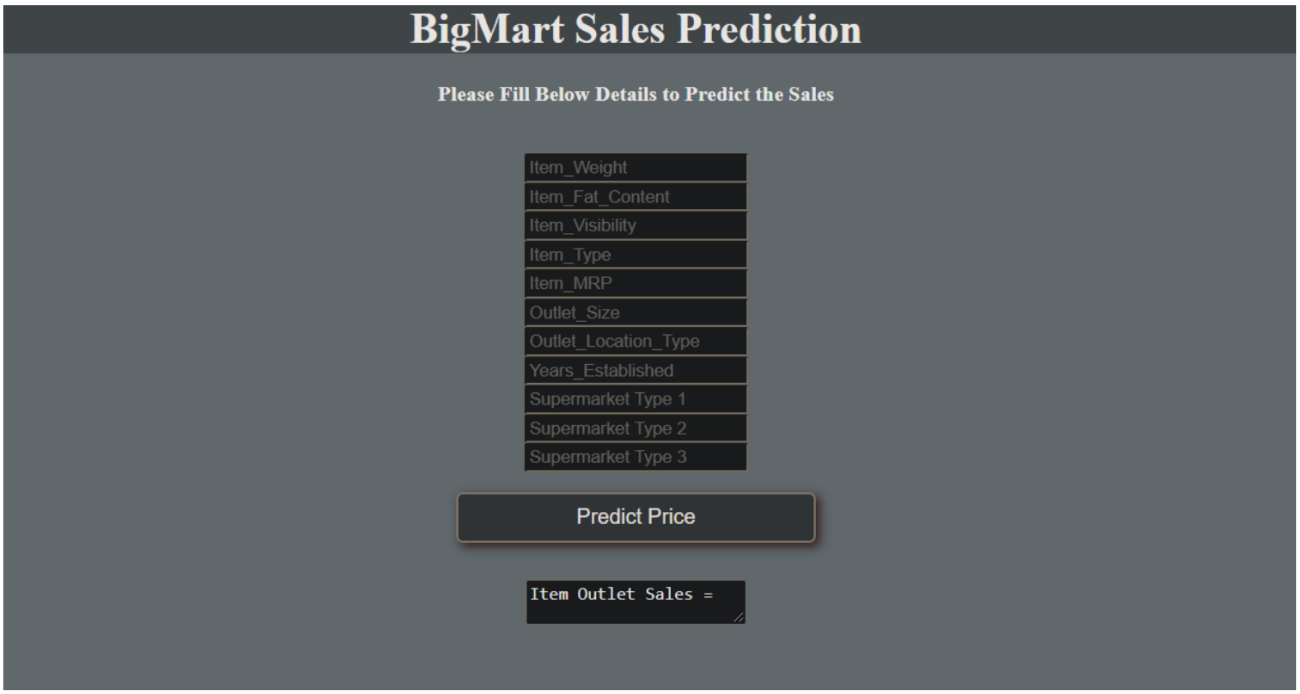

Figure No. 4 User Interface

5. Conclusion

The Big Mart Sales Prediction application will find out the Sales of Outlets. From this prediction the Outlets can arrange the items according to their needs.