High Level Design (HLD)
Black Friday Sales Prediction

Revision Number: 1.0

Last date of revision: 01/08/2023

Jothimalar Paulpandi

Document Version Control

Date Issued	Version	Description	Author
01/08/2023	1.0	First Version of Complete HLD	Jothimalar Paulpandi

Contents

Abstract

Retail is the sale of goods and services from individuals or businesses to the end-user. The retail industry provides consumers with goods and services for their everyday needs. In retail one of crucial part is to understand the consumer behaviour and make various arrangements for the sales of the company. Accurately predicting customer's behaviour can be a daunting task. Providing personalized offers for the costumers based on their purchasing pattern can help in increasing the sales of the company.

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.

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
- List and describe the non-functional attributes like:
    - o Security
    - o Reliability
    - o Maintainability
    - o Portability
    - o Reusability
    - o Application compatibility
    - o Resource utilization
    - o Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the 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 Product Perspective & Problem Statement

A retail company "ABC Private Limited" wants to understand the customer purchase behaviour (specifically, purchase amount) against various products of different categories. They have shared purchase summary of various customers for selected high-volume products from last month.

The data set also contains customer demographics (age, gender, marital status, city_type, stay_in_current_city), product details (product_id and product category) and Total purchase_amount from last month.

Now, they want to build a model to predict the purchase amount of customer against various products which will help them to create personalized offer for customers against different products.
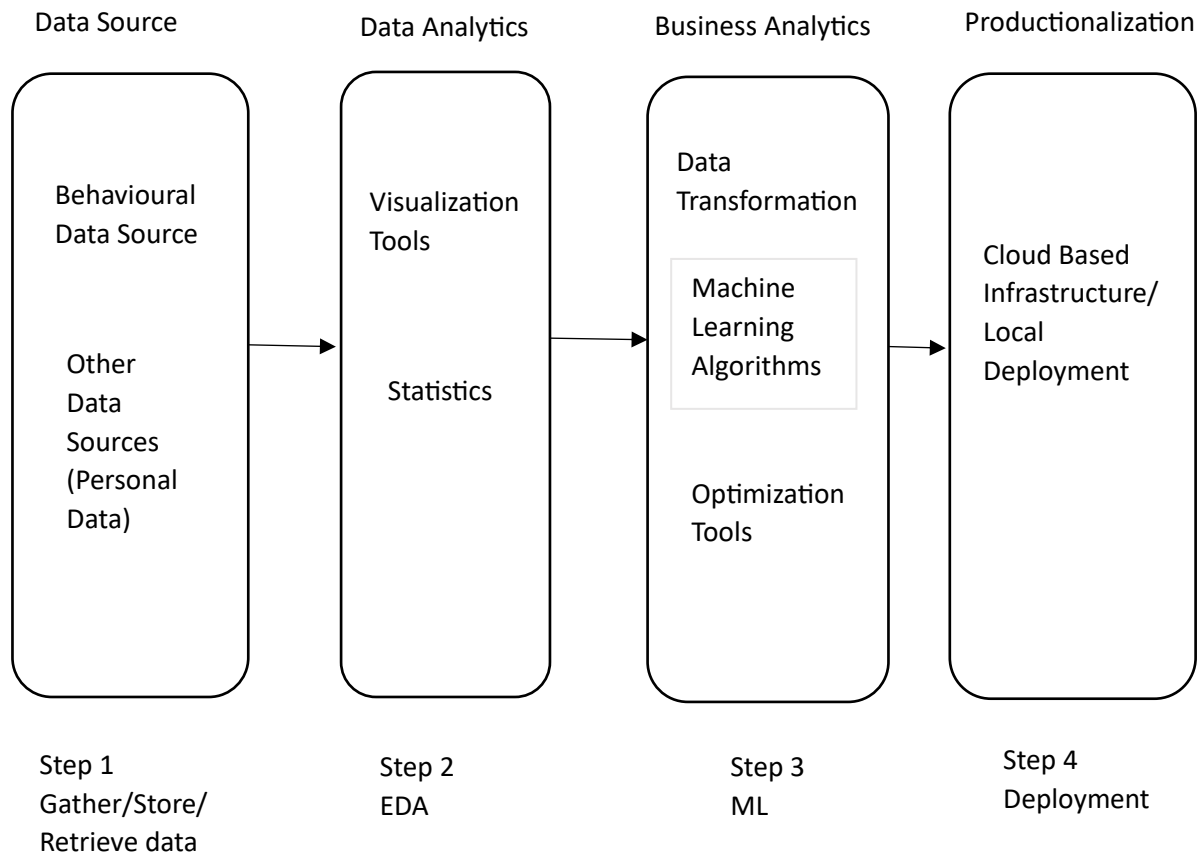
2.2 Tools used

Machine learning tools and libraries such as Python, Pandas, NumPy, Flask, Matplotlib, Seaborn, Scikit-learn and HTML are used to build the whole framework.

3 Design Details

3.1 Functional Architecture

Data Source	Data Analytics	Business Analytics	Productionalization
Behavioural Data Source Other Data Sources (Personal Data)	Visualization Tools Statistics	Data Transformation Machine Learning Algorithms Optimization Tools	Cloud Based Infrastructure/ Local Deployment
Step 1 Gather/Store/ Retrieve data	Step 2 EDA	Step 3 ML	Step 4 Deployment

3.2 Optimization

- Minimize the number of fields by removing irrelevant features.
- Minimize the number of records by removing duplicate records.
- Remove outliers.
- Handle missing data.
- Use Count/Frequency based or Label encoding rather than One Hot encoding when the number of unique categories is large.
- Perform Hyperparameter Tuning for the selected model.
- Perform Standardization / Normalization to bring all the features under same scale.
- Use Regularization techniques to maintain the Bias-Variance trade-off.

4 KPIs

4.1 KPIs (Key Performance Indicator)

We use RMSE (Root Mean Squared Error) as a KPI metric to evaluate the performance of the model. RMSE is a common metric for measuring the accuracy of machine learning models, especially for regression tasks. It is calculated by taking the square root of the mean squared error (MSE), which is the average of the squared differences between predicted and actual values. RMSE is sensitive to outliers and penalizes large errors more than small errors. A good RMSE value depends on the scale and context of the problem, but generally it should be as low as possible.

5 Deployment

Created the simple frontend page using HTML and deployed as a Rest API using Flask.

Gender: M ▾
Age: 0-17 ▾
Occupation: 0 ▾
City_Category A ▾
Stay_In_Current_City_Years: 0 ▾
Marital_Status: 0 ▾
Product_Category_1: nan ▾
Product_Category_2: nan ▾
Product_Category_3: nan ▾
Submit

4040.36