High Level Design(HLD)

Store Sales Prediction

Revision Number : 1.0
Last Date of Revision : 04/06/2022

Contents

Document Version Control

Date Issued	Version	Description	Author
04/06/2022	1	Initial HLD-V1.0	Amit Kumar Pandit

Abstract

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered.

1 Introduction

1.1 Why this High Level Design Document

The purpose of this High-Level-Design(HLD) Document is to add necessary details 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 a detail
- Describe the user interface being implemented
- Includes design features and architecture of the project
- List and describe the non functional attribute like:
  - Reliability
  - Maintainability
  - Portability
  - Reusability
  - Application compatibility
  - Serviceability

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 administrator of the system.

1.3 Definitions

Term	Definition
Database	Collection of all the information uploaded by the system or used by the system
IDE	Integrated Development Environment
AWS	Amazon Web Services

2   General Description

2.1   Product Perspective
The model helps to predict the sales of a product in a particular outlet

2.2   Problem Statement
Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered.

We have to build a solution that should able to predict the sales of the different stores of Big Mart according to the provided dataset.

2.3   Dataset Description
We have train (8523) and test (5681) data set, train data set has both input and output variable(s). We need to predict the sales for test data set.

2.4   Attribute Information
Item_Identifier: Unique product IDItem_Weight:
Weight of product
Item_Fat_Content: Whether the product is low fat or not
Item_Visibility: The % of total display area of all products in a store allocated to the particular product
Item_Type: The category to which the product belongs
Item_MRP: Maximum Retail Price (list price) of the product
Outlet_Identifier: Unique store ID
Outlet_Establishment_Year: The year in which store was established
Outlet_Size: The size of the store in terms of ground area covered
Outlet_Location_Type: The type of city in which the store is located
Outlet_Type: Whether the outlet is just a grocery store or some sort of supermarket
Item_Outlet_Sales: Sales of the product in the particulat store. This is the outcome variable to be predicted.

2.4 Proposed Solution

Here a system has been created which predicts the sales of a product in a particular outlet.

2.5 Tool used

- Python is used as programming language
- Flask is used as backend development
- Pycharm is used as IDE
- Front end development is done by using HTML/CSS/JS/Bootstrap
- MongoDB is used as a database to store data and logging

- Scikit-learn is used to create models
- For visualization of plots Matplotlib, Seaborn are used
- Github is used as version control
- AWS is used to deploy the application

2.5 Constraints

The Store Sales Prediction system must be user-friendly, as automated as possible, and the user should not be required to know any other work.

2.6 Assumption

The main objective of this project is to implement the use cases as previously mentioned(2.2 Problem Statement) as fast as possible, so that organizations can test the cases. It is also assumed that all aspects of the project have the ability to work together in the way the organization is expecting.

3   Design Details

3.1 Process Flow
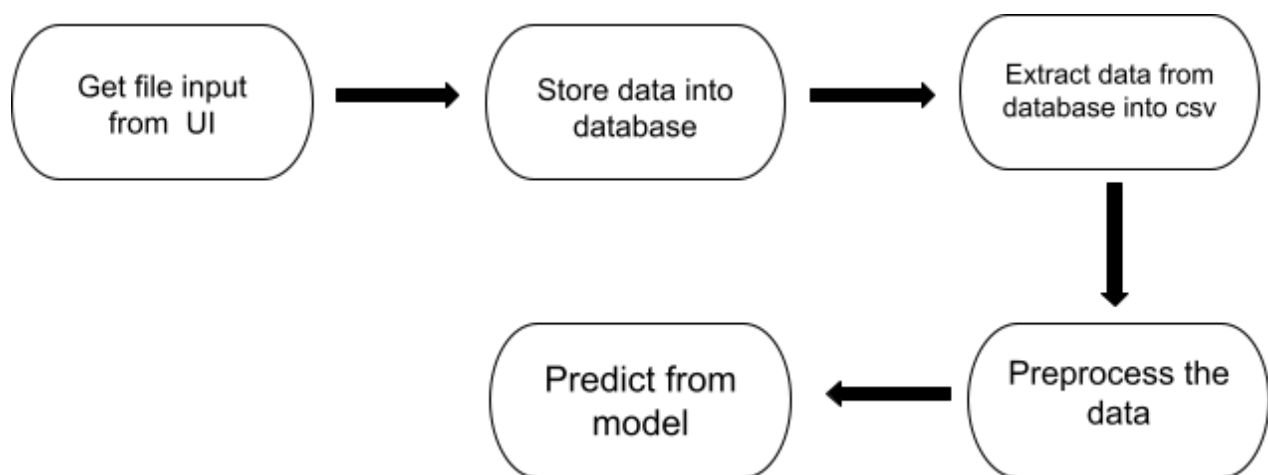
3.2 Event Log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description

1. The system identifies at what step logging required
2. The system should be able to log each and every system flow
3. Developers can choose a logging method. You can choose database logging or file logging as well
4. System should not hang even after so many loggins. Logging just because we can debug issues, so logging is mandatory.

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

This system will help to predict the sales of a item in a outlet based on previous data

4.1 Reusability

The code written should have the ability to be reused with no problems.

4.2 Application Compatibility

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

4.3 Resource Utilisation

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

4.4 Deployment

5   KPIS (Key Performance Indicators)

1. Organization can predict the sales without any experiment

6   Conclusion

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered.

7   References

1. https://docs.python.org/3/
2. https://xgboost.readthedocs.io/en/stable/
3. https://scikit-learn.org/stable/
4. https://getbootstrap.com/