

High Level Design (HLD)

Store Sales Prediction

Written By:

Mohamed Afaque Mulla

Version: 1.0

Date: 16-02-2023

Abstract

The goal of the project is to develop a system that can forecast future consumer demand for specific products in retail settings. by looking into a product's previous sales data. The sales of products at large shopping centers are tracked in order to predict future demand. Manufacturing and product warehouses are useful for storing a lot of things. The major objective of this study is to examine historical data, identify significant relationships among various characteristics, and develop a system that can make predictions about how much a certain product will be in demand. This technology will assist in controlling the storage capacity of warehouses.

1. Introduction

The desire to better serve local customer needs while preserving the benefits of central distribution presents a significant problem for large businesses. Methods focusing on store level optimization become more valuable as consumer demand for mass customization rises. The capacity to precisely forecast upcoming sales in the context of retail data mining directly translates into a number of high impact and doable actions.

Product assortment optimization, sales anomaly detection, customer segment targeting, new product distribution, and new store stocking are common applications that depend on precise item sales modelling.

With the help of additional data, it is possible to divide the data set into several groups and create unique models for each group. Building a rolled-up model that is a linear combination of all the sub models is an easy way to deal with this abundance of models. The issue then becomes figuring out the sub-model mixing weights, with each sub-model corresponding to a distinct partition. The mixing ratios can be used as a stand-in for the significance of client traits and the settings of the business environment in precisely modelling sales at the shop level.

The techniques selected for this investigation were subject to significant limitations. We made the explicit choice that the benefits of procedures that are simpler to explain to the client at any point in the study outweighed small methodological improvements.

We built our strategy with the knowledge that the data we received would eventually be continuously updated. It was decided to use quick online approaches because of the frequent updating requirements and the magnitude of the related data sets. The more processing-intensive aspects of our approach are supported by these techniques.

1.1 Why these High-Level-Documents?

The goal of these High-Level Design (HLD) Documents is to supplement the current project description with the information required to depict an appropriate for coding. The purpose of this text is to aid in the early detection of discrepancies. Also, it serves as a reference guide for the high-level interactions between the modules.

As for the HLD:

- Outline each part of the design in detail and present it.
- Describe the user interface being implemented.
- Describe the needed Python libraries for the coding.

- Describe the performance requirements.
- Include design features and the architecture of the project.
- List and describe the non-functional attributes like:
 - ❖ Security
 - ❖ Reliability
 - ❖ Maintainability
 - ❖ Portability
 - ❖ Reusability
 - ❖ Resource Utilization
 - ❖ Serviceability
 - ❖ Application Compatibility

1.2 Scope

The HLD documentation details the system's architecture, including the database architecture, application architecture (layers), application flow (navigation), and technological architecture. The HLD employs words that should be understandable to system administrators that are neither overly technical nor inherently non-technical.

1.3 Definition

ML – Machine Learning

DB – Database, a cloud platform where the data will be stored.

API – Application Programming Interface.

2. General Description

2.1 Product Perspective

A Machine Learning model called The Store Sales Prediction uses historical data to forecast future product demand. It will provide the figure used to calculate product sales.

2.2 Problem Statement

To create a system that can take information about a product and estimate how much demand there will be for it in the future. The application we create must be able to deliver outcomes.

2.3 Proposed Solution

The significant relationships between various qualities will be discovered using EDA, and a machine-learning algorithm will be used to forecast future sales demand. The client will input the necessary feature and receive results via the online application. The system will collect features, which will then be sent to the backend for validation and pre-processing before being fed to a machine learning model with hyperparameter tuning to forecast the result.

2.4 Data Requirements

The data needed for the project's construction is already there on the dashboard. Many product descriptions and historical sales figures are included in the Store Sales Forecast data. We will utilise the provided dataset to build the ML model. There are 8523 rows in the data, each including a variety of product-related details such as product id, category, store id, and location.

2.5 Tools Used

Python is one of the programming languages utilised here, and we will also use various other python-based libraries, such as Scikit-Learn for machine learning, Pandas for data manipulation, Numpy for numerical computing. We'll use GitHub to store all code files that should be accessible to the public.



2.6 Constraints

The system should be simple to use, and when utilising the web app, the user should receive all necessary messages. If the user makes a mistake on the web-app page, he or she should receive a suitable error message as well. The user should not be confused when using the system since all mistakes and results should be presented in the simplest way possible, and all buttons that will be added to the webpage should have accurate labels.

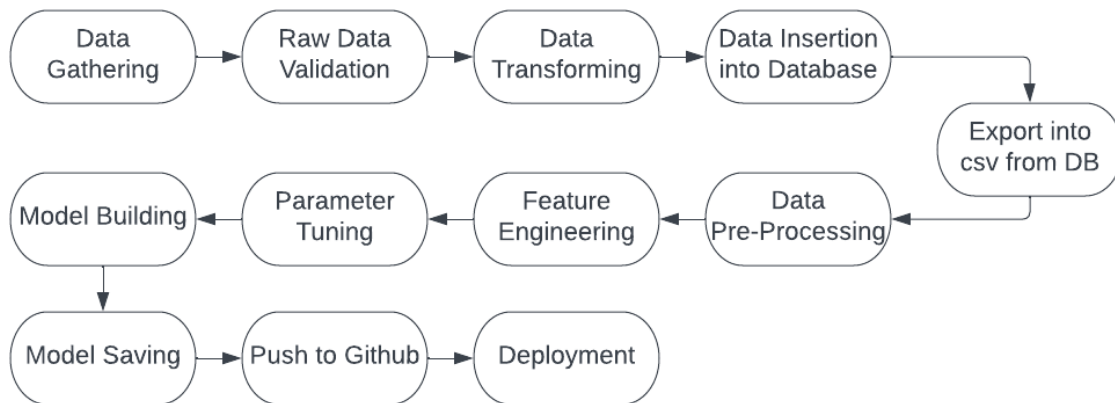
2.7 Assumptions

The major goal is to have a system in place that will generate an estimate of future demand for a product in retailers.

3. Design Details

3.1 Process Flow

For this project, we'll follow the workflow flow below. Modular coding, or using oops notions to build the complete project from beginning to end, will be the foundation of the method.



3.3 Handling the Error

If a processing issue occurred, the user should be presented with an error message that is entirely understandable to non-technical people. A clear error notice should be displayed so the user can correct his error and retry the process. Errors will happen, and they should all be handled correctly. And we need to manage and log any error that occurs with our application.

4. Performance

Machine-learning techniques are necessary for the Sales Price Prediction. We will test a variety of ML algorithms and select the one that best predicts the target. The data we give the algorithms will determine how well our system performs. And how well the model performs will rely on its completion. and the deployment server, as well as the web application. All of these parts should work together to make our programme work effectively.

4.1 Reusability

The project should adhere to all coding standards, and the code and modules developed during the building process should be written in a modular style. Each place should be able to use our technology without any problems. Additionally, it must handle any incorrect input values supplied by the user and issue a clear error notice so that the user can fix their error and supply accurate information in order to receive the desired result. Also, the system needs to be reusable under all circumstances with various inputs and values that are all consistent with how it was trained.

6. Conclusion

Business owners and manufacturers will soon be able to estimate future product demand thanks to the Sales Store Prediction. It will benefit the product supply chain as well as their ability to expand their business. We keep track of product sales history, prior performance, and store data. In order to discover the internal pattern and forecast the target price or the sales demand of the product in the future, we will analyse the historical data and create a Model Learning model.

7. Results

We perform the EDA firstly to understand the data. And run the data on many algorithms and the results are as follows:

Lasso Regression Accuracy: 0.54

Gradient Boosting Regressor Accuracy: 0.65

Random Forest Regressor Accuracy: 0.60

8. Reference

1. Retail Sales Prediction and Item Recommendations Using Customer Demographics at Store Level. - *Michael Giering*