

**SALES PRICE PREDICTION**

High Level Design

Domain: Machine Learning

Creator: Aravind S

Date: 23.07.2023

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Date issued | Version | Description | Author |
| July 23, 2023 | 1.1 | First Draft | Aravind S |

Contents

[Abstract 3](#_Toc141014914)

[Introduction 4](#_Toc141014915)

[What is High-Level Design Document? 4](#_Toc141014916)

[Scope 4](#_Toc141014917)

[General Description 5](#_Toc141014918)

[Definitions 5](#_Toc141014919)

[Product Description 5](#_Toc141014920)

[Problem Statement 5](#_Toc141014921)

[Proposed solution 5](#_Toc141014922)

[Further improvements 5](#_Toc141014923)

[Data requirements 5](#_Toc141014924)

[Tools used 6](#_Toc141014925)

[Hardware Requirements 8](#_Toc141014926)

[Constraints 8](#_Toc141014927)

[Assumption 8](#_Toc141014928)

[Design Details 9](#_Toc141014929)

[Process Flow 9](#_Toc141014930)

[Deployment Process 9](#_Toc141014931)

[Error Handling 10](#_Toc141014932)

[Performance 10](#_Toc141014933)

[Reusability 10](#_Toc141014934)

[Resource Utilization 11](#_Toc141014935)

[Deployment 11](#_Toc141014936)

[Conclusion 11](#_Toc141014937)

# Abstract

The present study focuses on developing a sales price prediction model using machine learning algorithms to help businesses optimize their pricing strategies and enhance revenue generation. In the ever-evolving landscape of commerce, accurate sales price prediction is essential for maintaining competitiveness and profitability.

The dataset utilized for this research comprises historical sales data from a diverse range of products across various industries. Features such as Item Fat Content, Item Type, Outlet Identifier, Outlet Size, Outlet Location Type, Outlet Type for predicting the Sale value of the product.

Several regression models were implemented and evaluated to identify the most suitable approach for sales price prediction. The models employed include linear regression, decision tree regression, random forest regression, and gradient boosting regression. Feature engineering and selection techniques were also applied to enhance model performance and mitigate the impact of irrelevant features.

To evaluate the models' predictive capabilities, the dataset was divided into training and testing sets, using various evaluation metrics such as mean absolute error (MAE), mean squared error (MSE), and R-squared (R2) to assess the models' accuracy and generalization capabilities.

The experimental results reveal that the Cat boosting regression model outperforms the other models, exhibiting the lowest prediction error and the highest R-squared value. This finding highlights the significance of ensemble learning techniques in capturing complex relationships between sales prices and their influencing factors.

The practical implications of this study are twofold: First, it provides businesses with a powerful tool for optimizing pricing strategies, enabling them to set competitive prices while maximizing profitability. Second, the study contributes to the growing body of knowledge in the field of machine learning applications for sales forecasting and pricing optimization.

In conclusion, this research demonstrates the efficacy of machine learning algorithms in predicting sales prices accurately. By leveraging historical data and advanced regression techniques, businesses can make data-driven decisions to stay ahead in a highly competitive market. Further research could explore the integration of additional data sources and advanced deep learning models to enhance predictive accuracy and robustness.

# Introduction

## What is High-Level Design Document?

The goal of this HLD or a high-level design 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 design aspects and define them in detail

Describe all user interfaces being implemented

Describe the hardware and software interfaces

Describe the performance requirements

Include design features and architecture of the project

List and describe the non-functional attributes such as security, reliability, maintainability, portability, reusability, application compatibility. resource utilization, serviceability

## Scope

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

# General Description

## Definitions

|  |  |
| --- | --- |
| Term | Description |
| ML | Machine Learning |
| API | Application programming Interface |
| IDE | Integrated Development Environment |
| UI | User Interface |

## Product Description

A Web UI which predicts the sales price of the product based on the given information

## Problem Statement

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

## Proposed solution

A Web UI that predicts the sales price of the product based on the given information, which it has been trained on, is a useful tool for determining the optimal product price. It offers businesses a powerful solution for fixing their product prices accurately.

## Further improvements

This project can be further improved by connecting the data to a NO SQL or to a cloud database and retrieving the data from the source.

## Data requirements

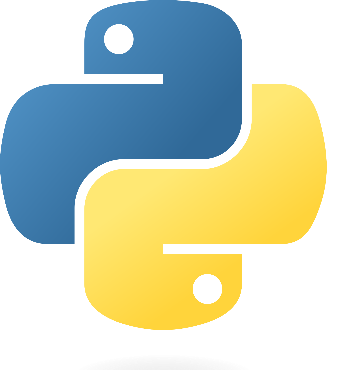
The data requirements for the project is the same as the features given in the dataset which are:

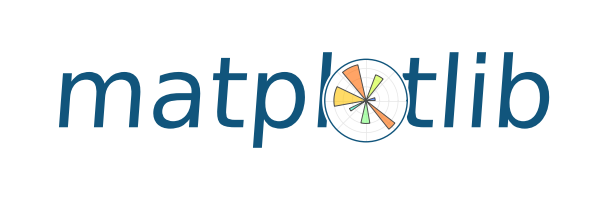
1. Item\_Identifier: Unique product ID
2. Item\_Weight: Weight of product
3. Item\_Fat\_Content: Whether the product is low fat or not
4. Item\_Visibility: The % of total display area of all products in a store allocated to the particular product
5. Item\_Type: The category to which the product belongs
6. Item\_MRP: Maximum Retail Price (list price) of the product
7. Outlet\_Identifier: Unique store ID
8. Outlet\_Establishment\_Year: The year in which store was established
9. Outlet\_Size: The size of the store in terms of ground area covered
10. Outlet\_Location\_Type: The type of city in which the store is located
11. Outlet\_Type: Whether the outlet is just a grocery store or some sort of supermarket
12. Item\_Outlet\_Sales: Sales of the product in the particulat store. This is the outcome variable to be predicted.

## Tools used

Sales price prediction involves the application of various machine learning and statistical modeling techniques. To implement these techniques effectively, you can use a combination of programming languages and libraries specifically designed for data analysis and machine learning. Some commonly used tools for sales price prediction include:

1. Python: Python is a popular programming language for data science and machine learning. It offers a wide range of libraries and frameworks, such as NumPy, pandas, scikit-learn, and TensorFlow, that facilitate data manipulation, model building, and evaluation.
2. Jupyter Notebook: Jupyter Notebook is an interactive development environment that allows you to combine code, visualizations, and explanatory text in a single document. It's commonly used for exploratory data analysis and model prototyping.
3. scikit-learn: This is a powerful machine learning library for Python that provides various algorithms for regression, classification, clustering, and more. It's widely used for building predictive models, including sales price prediction.
4. Matplotlib: Matplotlib is a widely-used data visualization library in Python. It provides a flexible and comprehensive set of functions to create various types of static, interactive, and animated plots.
5. Seaborn: Seaborn is a Python data visualization library built on top of Matplotlib. It provides a high-level interface for creating aesthetically pleasing statistical graphics.
6. Streamlit: Streamlit is a powerful Python library used for building interactive web applications for data science and machine learning projects.







# Hardware Requirements

* A computer system capable of running python. The chosen IDE (such as Visual Studio Code). The system should meet the minimum system requirements for the selected software.
* Minimum 1.10 GHz processor or equivalent.
* Between 1-2 GB of free storage
* Minimum 512 MB of RAM
* 3 GB of hard-disk space

## Constraints

The System should be user-friendly, the user should get all proper messages while using the web app. He/she also should get a proper error message if he/she has done something wrong On the web-app page. All the errors and results should be delivered in the easiest possible way and all the buttons are going to insert on the webpage should be labeled properly, so the user did not get confused to use the system.

# Assumption

The main objective is to implement a system that will produce approximate future demand for a product in stores.

# Design Details

## Process Flow

## Deployment Process

## Error Handling

In case of any processing errors, the system will display user-friendly error messages that are easily understandable to anyone. The error messages will be meaningful, helping users identify their mistakes and make necessary improvements for rerunning the process successfully. All potential errors will be handled with precision, ensuring smooth error management. The application will log every error for efficient tracking and management.

# Performance

Sales Price Prediction heavily relies on machine learning algorithms. The project involves training several ML algorithms to determine the best-fit model for accurate target prediction. The system's performance is closely linked to the quality and relevance of the data used to train these algorithms. Ultimately, the performance hinges on the effectiveness of the selected final model, the web application's functionality, and the deployment server's efficiency. By ensuring the seamless integration of all components, the program will run smoothly, providing reliable and precise sales price predictions.

## Reusability

During the project development phase, utmost care is taken to adhere to coding guidelines, ensuring that the code and modules are built in a modular fashion. This approach enhances code reusability, maintainability, and scalability. The system is designed to be flexible, allowing it to function optimally from any location. To provide a seamless user experience, the system is equipped to handle improper input values gracefully. When encountering such inputs, it generates meaningful error messages that guide the user in rectifying any mistakes, enabling them to input valid values and obtain accurate results. Moreover, the system's reusability is emphasized as it can efficiently handle various types of inputs, leveraging its comprehensive training. This versatility makes the system applicable to diverse scenarios, facilitating smooth integration into different projects and applications.

## Resource Utilization

Our application is designed to make efficient use of available resources, ensuring minimal internet consumption while interacting with APIs on the web page. Careful resource management ensures that the application does not impose a heavy computational burden, resulting in optimal performance without slowing down. When deployed on a cloud platform, the application leverages cloud resources effectively, enabling seamless operation and utilization of available cloud resources to ensure smooth functionality.

# Deployment

The deployment process is done with the help of streamlit which is a powerful Python library used for building interactive web applications for data science and machine learning projects. It allows data scientists and developers to create intuitive web interfaces for their models and visualizations with minimal effort.

# Conclusion

Sales Store Prediction aims to assist business owners and manufacturing companies in predicting future product demand. This predictive capability will facilitate business growth and optimize supply chain management. Leveraging historical records of products, product sales, and store information, the project involves extensive data analysis. Utilizing machine learning, an advanced model will be developed to identify internal patterns within the data and accurately predict the target value, which represents the future sales demand for the products.