High Level Design

Store Sales Prediction

First Draft : 12/12/21



# Document Version Control

| **Date Issued** | **Version** | **Description** | **Author** |
| --- | --- | --- | --- |
| 12/12/2021 | 1 | HLD-V1.0 | Raghav Kachroo |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_heading=h.gjdgxs)

[Abstract 4](#_heading=h.30j0zll)

1. [Introduction 5](#_heading=h.1fob9te)
   1. [Why this High-Level Document? 5](#_heading=h.3znysh7)
   2. [Scope 5](#_heading=h.2et92p0)
   3. [Definitions 5](#_heading=h.tyjcwt)
2. [General Description 6](#_heading=h.3dy6vkm)
   1. [Product Perspective 6](#_heading=h.1t3h5sf)
   2. [Problem Statement 6](#_heading=h.4d34og8)
   3. [Proposed Solution 6](#_heading=h.1ci93xb)
   4. [Data Requirements… 6](#_heading=h.2s8eyo1)
   5. [Tools Used 7](#_heading=h.17dp8vu)
   6. [Constraints… 8](#_heading=h.3rdcrjn)
   7. [Assumptions… 8](#_heading=h.26in1rg)
3. [Design Details 9](#_heading=h.lnxbz9)
   1. [Process flow 9](#_heading=h.35nkun2)
      1. Model Training & Evaluation 9
      2. [Deployment Process 10](#_heading=h.1ksv4uv)
   2. [Event log 10](#_heading=h.44sinio)
   3. [Error Handling 10](#_heading=h.2jxsxqh)
4. [Performance… 11](#_heading=h.z337ya)
   1. [Reusability 11](#_heading=h.3j2qqm3)
   2. Application Capability 11
   3. [Resource Utilization 11](#_heading=h.1y810tw)
   4. [Deployment 11](#_heading=h.4i7ojhp)
5. [Conclusion 12](#_heading=h.2xcytpi)



# Abstract

Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In order to predict store sales machine learning can be used so that Big Marts can adjust their inventory likewise. This study demonstrates how different regression algorithms can forecast the presence of the sales. Different regression algorithms such as Linear regression, Random Forest Regressor are implemented.

# Introduction

## Why this High-Level 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 model 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 the design aspects and define them in detail
    - Describe the user interface being implemented
    - Describe the hardware and software interfaces
    - Describe the performance and requirements
    - Include design features and the architecture of the project
    - List and describe the non-functional attributes like:
      * Security
      * Reliability
      * Maintainability
      * Portability
      * Reusability
      * Application compatibility
      * Resource utilization
      * Serviceability

## Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture, application flow (Navigations), and technology architecture. The HLD uses non-technical to miIdIy-technical term which should be understandable to the administrator of the system.

## Definitions

| **TERM** | **DESCRIPTION** |
| --- | --- |
| **Database** | Collection of all the information monitored by this system |
| **IDE** | Integrated Development Environment |
| **AWS** | Amazon Web Services |

# General Description

## Product Perspective

The Store Sales Prediction is a machine learning based multi regression model which will help us to predict the sales of the products.

## Problem Statement

Shopping malls and Big Marts keep track of individual item sales data in

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

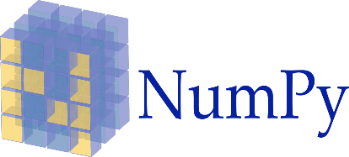
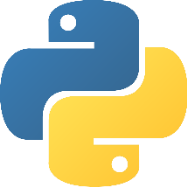
particular item details. By mining the data store from the data warehouse, more

anomalies and common patterns can be discovered.

## Data Requirements

| **Name** | **Data Type** | **Description** |
| --- | --- | --- |
| **Item\_Identifier** | String | Unique product ID |
| **Item\_Weight** | Float | Weight of product |
| **Item\_Fat\_Content** | String | Whether the product is low fat or not |
| **Item\_Visibility** | Float | The % of a total display area of all products in a store allocated to the particular product |
| **Item\_Type** | String | The category to which the product belongs |
| **Item\_MRP** | Float | Maximum Retail Price (list price) of the product |
| **Outlet\_Identifier** | String | Unique store ID |
| **Outlet\_Establishment\_Year** | Integer | The year in which the store was established |
| **Outlet\_Size** | String | The size of the store in terms of ground area covered |
| **Outlet\_Location\_Type** | String | The type of city in which the store is located |
| **Outlet\_Type** | String | Whether the outlet is just a grocery store or some sort of supermarket |
| **Item\_Outlet\_Sales** | Float | Sales of the product in the particular store. This is the outcome variable to be predicted. |

## Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Streamlit, Heroku, Git.







* + - PyCharm is used as an IDE.
    - For visualization of the plots, Matplotlib, Seaborn are used.
    - Heroku is used for deployment of the model.
    - MongoDB is used to retrieve, insert, delete, and update the database
    - Frontend development is done using HTML/CSS
    - Python Streamlit is used for backend development.
    - GitHub is used as a version control system.

## Constraints

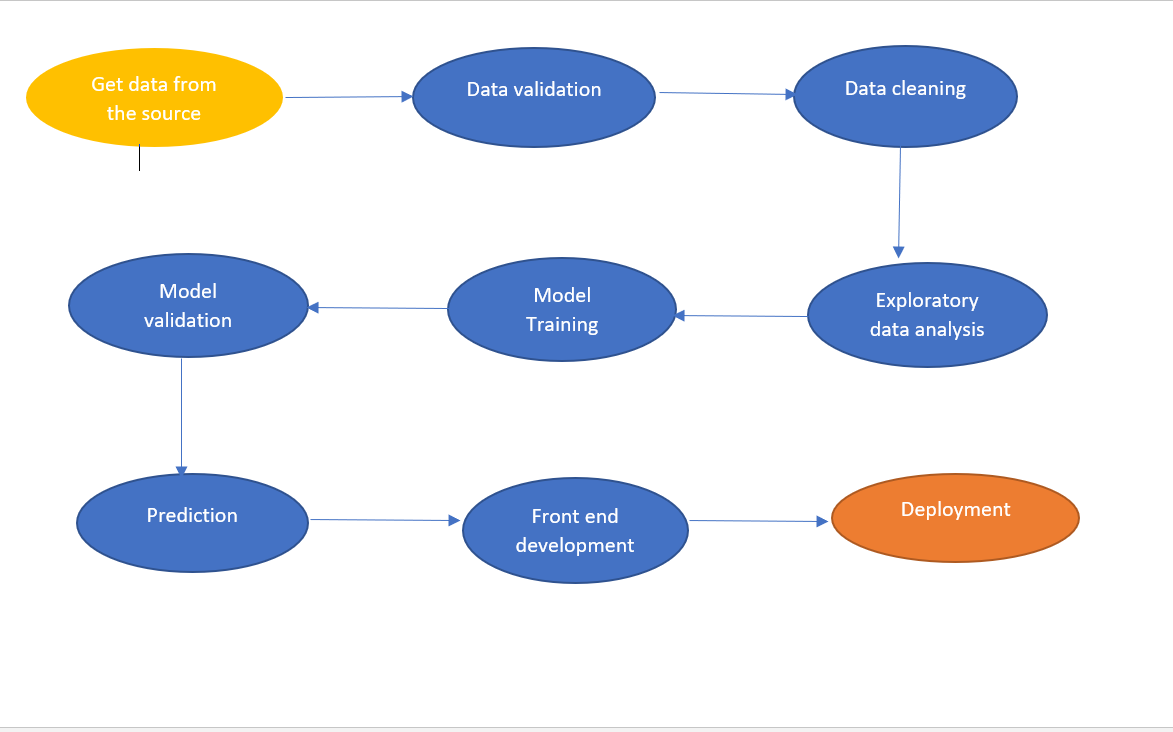
The Store Sales Prediction application must be user friendly, as automated as possible and users should not be required to know any of the workings.

# Design Details

## Process Flow

For predicting the sales, we will use a regression model. Below is the process flow diagram as shown below.

## Proposed Methodology



* + 1. Model Training and Evaluation



## Deployment Process





## Event Log

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

### Initial Step-By-Step Description:

* The system identifies at what step logging is required.
* The system should be able to log each and every system flow.
* Developers can choose logging methods. You can choose database logging / File logging as well.
* System should not hang even after using loggings. Logging just because we can easily debug issues, so logging is mandatory to do.

## Error Handling

Should error 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

# Performance

Adjusted R-squared and R-squared values for Linear Regression Model are higher than other models.

Also its RMSE value is low as compared to other models with the highest CV score. Therefore, the Linear Regression model fits better and exhibits accuracy

## Reusability

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

## Application Compatibility

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

## Resource Utilization

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

## Deployment



# Conclusion

In this project, the basics of machine learning and the associated data processing and modeling algorithms have been described, followed by their application for the task of sales prediction in Big Mart shopping centers at different locations. On implementation, the prediction results show the correlation among different attributes considered and how a particular location of medium size recorded the highest sales, suggesting that other shopping locations should follow similar patterns for improved sale