

Stores Sales Prediction

Revision Number: 1.1

Last date of revision: 03/09/2021

Document Version Control

Date Issued	Version	Description	Author
28/08/2021	1	Initial HLD - V1.0	Ashok Kumar S M C Shravan
03/09/2021	1.1	Final HLD – V1.1	Ashok Kumar S M C Shravan

Contents

Document Version Control	2
Abstract.....	4
1 Introduction	5
1.1 Why this High-Level Design Document?.	5
1.2 Scope.	5
1.3 Definitions	5
2 General Description.....	6
2.1 Product Perspective	6
2.2 Problem statement.....	6
2.3 PROPOSED SOLUTION	6
2.4 FURTHER IMPROVEMENTS	6
2.5 Data Requirements	7
2.6 Tools used.	8
2.7 Hardware Requirements.	9
2.8 Constraints	10
2.9 Assumptions.....	10
3 Design Details.....	11
3.1 Process Flow.	11
3.1.1 Model Training and Evaluation.....	11
3.1.2 Deployment Process.....	12
3.2 Event log.....	12
3.3 Error Handling.....	12
4 Performance.....	13
4.1 Reusability.....	13
4.2 Application Compatibility	13
4.3 Resource Utilization	13
4.4 Deployment.	13
5 Conclusion	14
6 References.....	14

Abstract

The value for a particular product or item keeps on changing from time to time. Any business cannot improve its financial performance without estimating the demand of the customer and future sales of products or items accurately. In this project we are trying to predict the sales of stores and Big Marts using different machine learning techniques and trying to determine the best algorithm suited to our particular problem statement.

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 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.

1.3 Definitions

<i>Term</i>	<i>Description</i>
<i>Database</i>	Collection of all the information monitored by this system
<i>IDE</i>	Integrated Development Environment

2 General Description

2.1 Product Perspective

The Stores Sales Prediction is a machine learning-based model which will help us to predict the sales of products in the different stores.

2.2 Problem statement

To create a model for predicting the future sales of the product in a store or Big Mart and to implement the following use cases.

- To keep track of individual item sales data in order to forecast future client demand.
- To discover more anomalies by mining the data store from the data warehouse.
- To discover the common pattern by mining the data store from the data warehouse.

2.3 PROPOSED SOLUTION

The solution proposed here is a Stores Sales Prediction System which will help to predict the sales of the products in the stores based on the provided datasets and helps to estimate future sales by implementing the classic machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing and convenient machine learning algorithms.

2.4 FURTHER IMPROVEMENTS

Stores sales prediction can increase the accuracy of the output even more with better machine learning algorithms. Extra data can be added to the datasets which helps in providing better accurate sales. An effective sales management system can be added to keep the sales in check. The project can be further improved by keeping tabs on factors impacting the sales prediction

2.5 Data Requirements

Data requirements are completely dependent on the given problem statement.

- We need to use the dataset set which has been provided with both test and train data.
- The test and train datasets should have both input and output variables.
- The data sets should consist of the following:
 1. Item_Weight: Weight of product
 2. Item_Fat_Content: Whether the product is low fat or not
 3. Item_Visibility: The % of total display area of all products in a store allocated to the particular product
 4. Item_Type: The category to which the product belongs

5. Item_MRP: Maximum Retail Price (list price) of the product
 6. Outlet_Identifier: Unique store ID
 7. Outlet_Establishment_Year: The year in which store was established
 8. Outlet_Size: The size of the store in terms of ground area covered
 9. Outlet_Location_Type: The type of city in which the store is located
 10. Outlet_Type: Whether the outlet is just a grocery store or some sort of supermarket
- All these will have different values and they act as the input where the user enters the input data.
 - Item_Outlet_Sales is the sales of the product in the particular store. This is the outcome variable to be predicted.

2.6 Tools used

The tools used to build the model are Python programming language and frameworks such as NumPy, Pandas, Matplotlib and Seaborn, Scikit Learn.



- Jupyter Notebook is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Heroku is used for deployment of the model.
- Web application is created in Notepad++.
- Front end development is done using HTML/CSS
- Backend development is done using Flask
- GitHub is used as version control system.

2.7 Hardware Requirements

- PC/ Laptop



2.8 Constraints

The Stores Sales Prediction is limited to the data provided in the datasets. Only the data provided in the dataset can be used for predicting the sales of the product

2.9 Assumptions

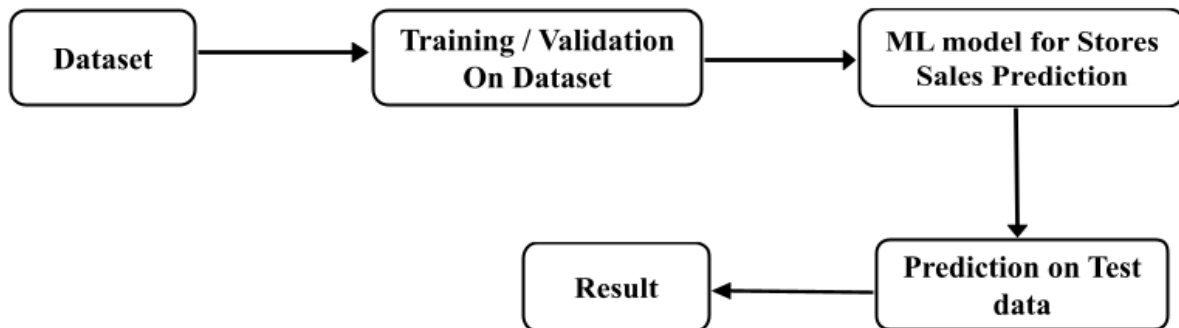
The main objective of the project is to implement the use cases as previously mentioned for an already existing dataset provided for doing the project. A machine learning based algorithm named Gradient Boosting Regression algorithm is used for predicting the sales. It can also be assumed that every part of the project is working as per the programmer's requirements and the end users' need.

3 Design Details

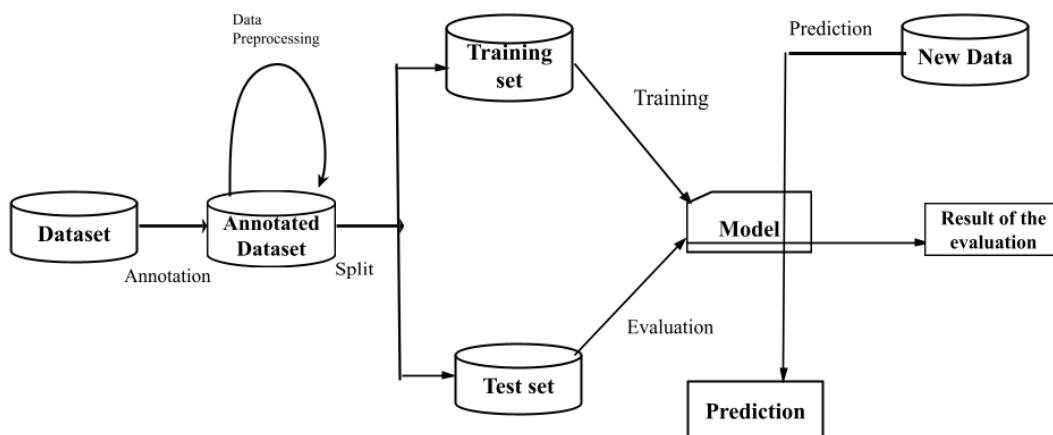
3.1 Process Flow

For predicting the sales of stores, we will use a machine learning model. The process flow diagram is as shown below.

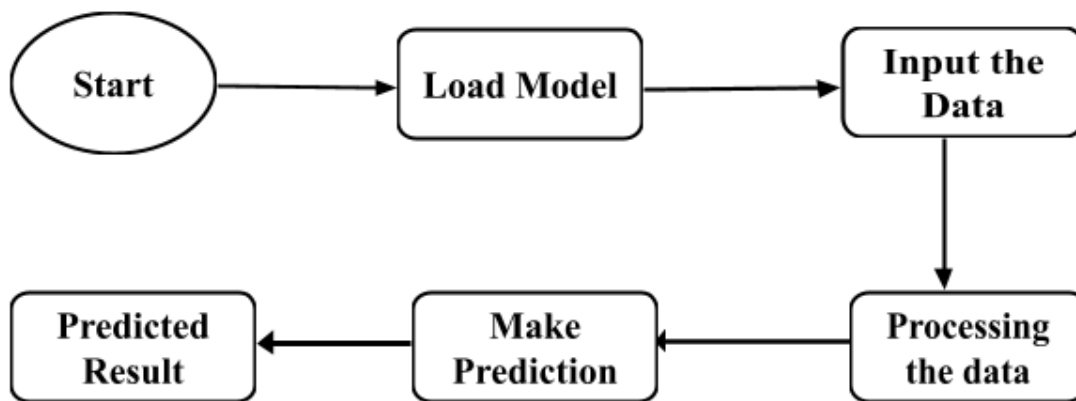
Proposed methodology



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



3.2 Event log

Every event is logged in the system so that the user can understand whichever 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. Developer can choose logging method. You can choose database logging/ File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

3.3 Error Handling

When it comes to error handling, there are most possibly no errors because all the errors are handled during the execution of the program

4 Performance

The Stores Sales Prediction System is used for predicting the sales of different stores. In the designed webapp, the input is given for a particular product it provides us with the necessary result with the highest accuracy and as fast as possible which helps to predict the sales of the particular product in stores or Big Marts.

4.1 Reusability

The written code for this project can be reused as many number of times possible without any errors or issues.

4.2 Application Compatibility

The main component in the project will be using Python for writing the code and web app which helps to retrieve and submit the data from the database over the internet.

4.3 Resource Utilization

When the input is given, it will take the necessary actions to get the accurate results from the provided information.

4.4 Deployment



5 Conclusion

The Designed web application for Stores Sales Prediction will help to predict the sales of the different stores of Big Mart with the help of the dataset provided.

6 References

1. <https://medium.com/analytics-vidhya/rossmann-store-sales-prediction-998161027abf#:~:text=Sales%20forecasting%20refers%20to%20the,a%20specific%20period%20of%20time.>
2. Google.com for understanding the topic
3. <https://www.kaggle.com/brijbhushannanda1979/bigmart-sales-data> for dataset