Low Level Design (LLD)

STORE SALES PREDICTION

First Draft : 11th October 2021

Jayesh Sehgal

Ayush Kashyap

# 

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 11th Oct 2021 | 1.1 | First Draft | Ayush Kashyap, Jayesh Sehgal |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_gjdgxs)

[**Abstract** 4](#_30j0zll)

[1](#_1fob9te) Introduction 5

[1.1](#_3znysh7) Why this Low-Level Design Document? 5

[1.2](#_2et92p0) Scope 6

[1.3](#_tyjcwt) Constraints 6

[1.4](#_3dy6vkm) 1.3 Risks 6

[1.5](#_1t3h5sf) 1.4 Out of Scope 7

[2](#_4d34og8) Technical specifications 7

[2.1 Predicting Disease 7](#_2s8eyo1)

[2.2 Logging 7](#_17dp8vu)

[2.3 Database 7](#_3rdcrjn)

[3](#_26in1rg) Technology stack 8

[4](#_lnxbz9) Proposed Solution 8

[5](#_35nkun2) Model training/validation workflow 9

[6](#_1ksv4uv) User I/O workflow 10

[7](#_44sinio) Exceptional scenarios 11

**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 the how different regression algorithms can forecasts the presence of the sales. Different regression algorithms such as Linear regression, Random Forest Regressor are implemented

# Introduction

## Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Store Sales Prediction. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The main objective of the project is to predict the sales of the product.

Store Sales Prediction is essential for companies :

* Helps in forecasting the sale of different products
* Helps in product warehousing
* It provides an estimate on how the market will react to the product.

A Store Sales Prediction System contains information on product’s:

* Unique ID
* Item Information
* Outlet information

## Scope

This software system will be a Web application This system will be designed to predict the sales of the product and with that it will provide an estimate on how well the product do in the market. More specifically, It will give the company an idea of what products can increase their overall profit in long run and it will gain profit by selling that high sale product and also by reducing the warehouse capacity of low sale product.

## Constraints

The Store Sales Prediction System should be easy to use and should give an accurate prediction regarding the sales.

## Risks

Document specific risks that have been identified or that should be considered.

## Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

# Technical specifications

## 2.1 Dataset

## 2.1.1 Sales dataset overview

The sales data set consists of 8523 training data sets with 12 feature variables

## 2.1 Predicting Sales

* The system presents the set of inputs required from the user.
* The user gives required information.
* The system then predicts approximate sale

## 2.2 Logging

We should be able to log every activity done by the user.

* The System identifies at what step logging 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 be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

## 2.3 Database

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User gives required information.
2. The system stores each and every data given by the user or receives on request to the database. Database which we have chosen is Cassandra.

2.4 Deployment

* AWS



# Technology stack

|  |  |
| --- | --- |
| **Front End** | HTML/CSS |
| **Backend** | Python Streamlit |
| **Database** | Cassandra |
| **Deployment** | AWS |

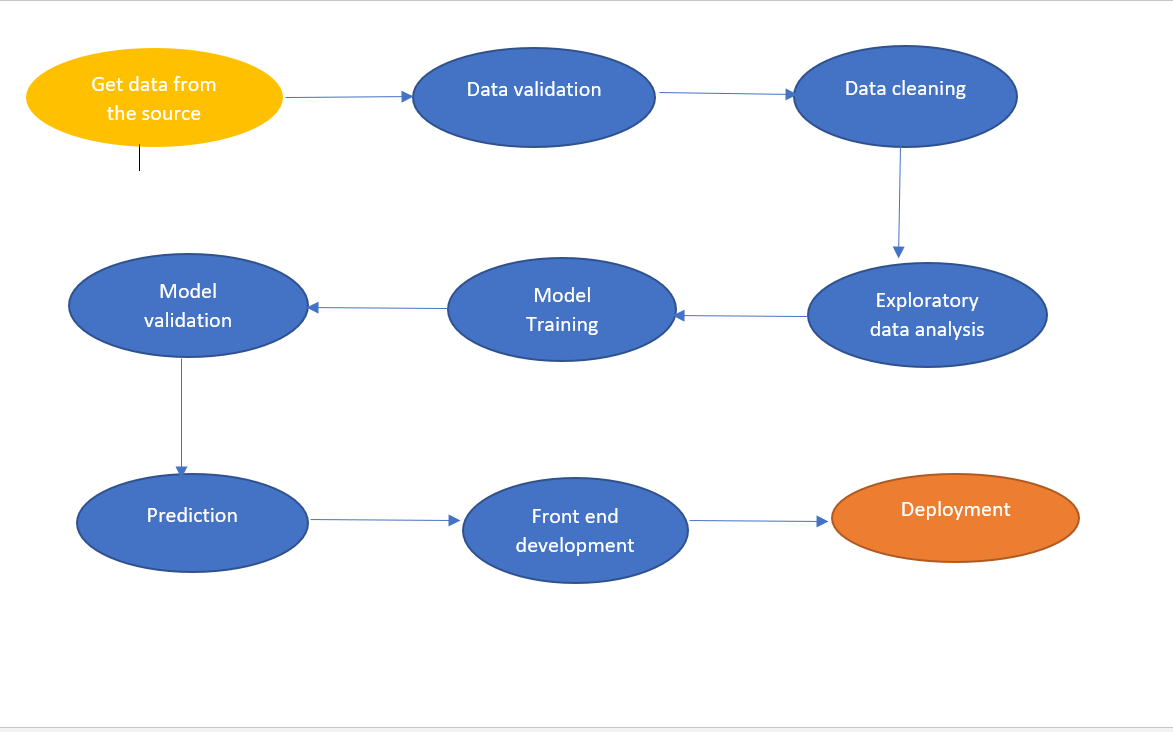
# Proposed Solution

After data processing we apply the algorithms and build a models to check accuracy and the model with least error will be selected to be dumped in a pickle while so that we can deploy the model through AWS console which will make the model user friendly overall.

The user can just input the values on the Front end page and the model can draw out results.

For data processing, some of the concepts of missing values, outliers and correlation is kept in mind while altering the dataset and for model training regression algorithms are used to get the best accuracy.

# System Workflow



# 

# Exceptional scenarios

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Exception | Mitigation | Module |
| 11th Oct 2021 | 1.1 | First Draft |  |
|  |  |  |  |

# Test cases

|  |  |  |  |
| --- | --- | --- | --- |
| Test case | Steps to perform test case | Module | Pass/Fail |
|  |  |  |  |

# 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 model with highest CV score. Therefore, the Linear Regression model fits better and exhibits accuracy