Low-Level Design (LLD)

Store Sales Prediction

Dipendra Singh

Dinansh Bhardwaj

Document Version Control

Date Issued	Version	Description	Author
18th Feb 2020	1.1	First Draft	Dipendra Singh
20th Feb 2020	1.2	Added Workflow chart	Dipendra Singh
20th Feb 2020	1.3	Added Exception Scenarios Overall, Constraints	Dinansh Bhardwaj
26th Feb 2021	1.5	Added user I/O flowchart	Dipendra Singh
31st Feb 2021	1.7	Added dataset overview and updated user I/O flowchart.	Dinansh Bhardwaj

Contents

Abstract

In this project, we applied machine learning techniques to a realworld problem of predicting stores sales. This kind of prediction enables store managers to create effective staff schedules that increase productivity and motivation. We used the popular open-source statistical programming language Python. We used feature selection, model selection to improve our prediction result. In view of the nature of our problem, Root Mean Square Error (RMSE) is used to measure the prediction accuracy.

1. Introduction

1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Deep SSP System. 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.

- Present all of the design aspects and define them in detail
- Describe the user interface is 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
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

This software system will be a Web application. This system will be designed to predict the Sales of the product in the particular store.

1.3 Constraints

The Store Sales Prediction must be user-friendly, as automated as possible and users should not be required to know any of the workings. We will use only a few columns depending on model behavior.

1.4 Risks

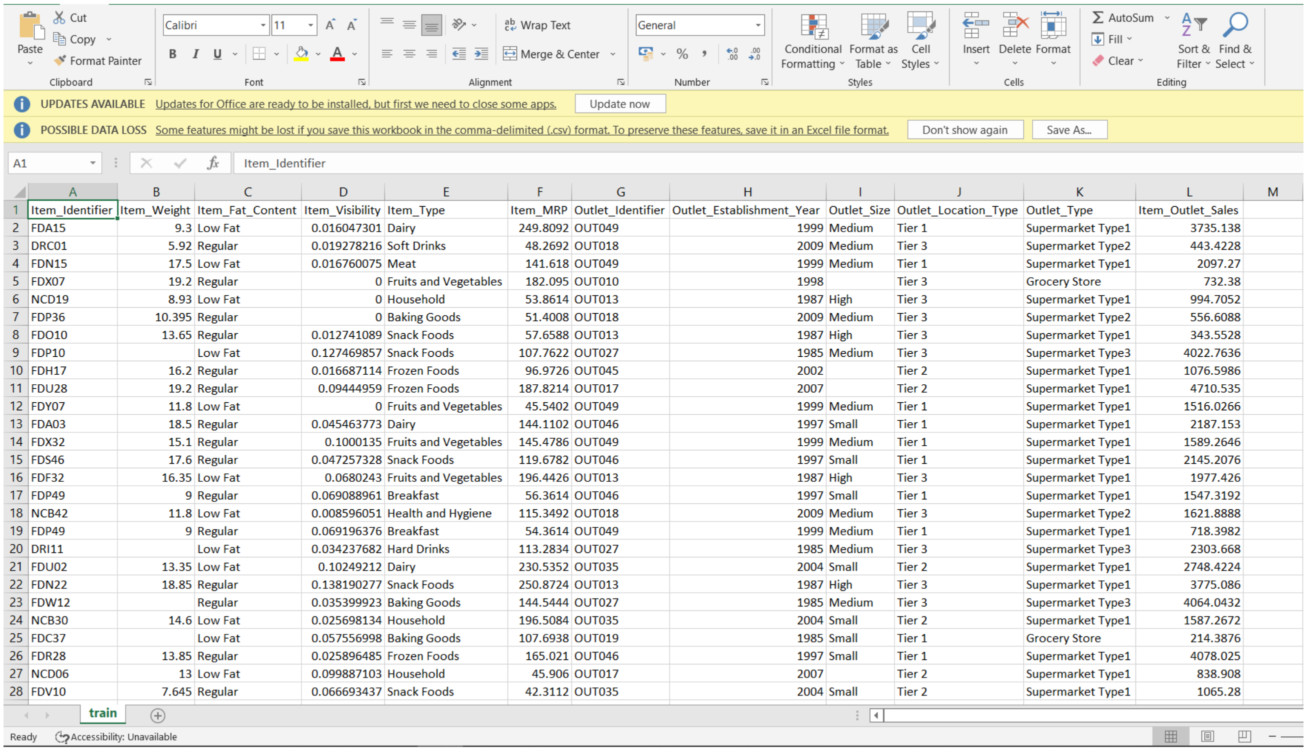
Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

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

2. Technical specifications

2.1 Dataset



UPDATES AVAILABLE Updates for Office are ready to be installed, but first we need to close some apps.

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

A1 Item_Identifier

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales	
2	FDA15	9.3	Low Fat	0.016047301	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138	
3	DRC01	5.92	Regular	0.019278216	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228	
4	FDN15	17.5	Low Fat	0.016760075	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27	
5	FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38	
6	NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052	
7	FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3	Supermarket Type2	556.6088	
8	FDO10	13.65	Regular	0.012741089	Snack Foods	57.6588	OUT013	1987	High	Tier 3	Supermarket Type1	343.5528	
9	FDP10		Low Fat	0.127469857	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3	Supermarket Type3	4022.7636	
10	FDH17	16.2	Regular	0.016687114	Frozen Foods	96.9726	OUT045	2002		Tier 2	Supermarket Type1	1076.5986	
11	FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017	2007		Tier 2	Supermarket Type1	4710.535	
12	FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049	1999	Medium	Tier 1	Supermarket Type1	1516.0266	
13	FDA03	18.5	Regular	0.045463773	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermarket Type1	2187.153	
14	FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049	1999	Medium	Tier 1	Supermarket Type1	1589.2646	
15	FDS46	17.6	Regular	0.047257328	Snack Foods	119.6782	OUT046	1997	Small	Tier 1	Supermarket Type1	2145.2076	
16	FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013	1987	High	Tier 3	Supermarket Type1	1977.426	
17	FDP49	9	Regular	0.069088961	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermarket Type1	1547.3192	
18	NCB42	11.8	Low Fat	0.008596051	Health and Hygiene	115.3492	OUT018	2009	Medium	Tier 3	Supermarket Type2	1621.8888	
19	FDP49	9	Regular	0.069196376	Breakfast	54.3614	OUT049	1999	Medium	Tier 1	Supermarket Type1	718.3982	
20	DRI11		Low Fat	0.034237682	Hard Drinks	113.2834	OUT027	1985	Medium	Tier 3	Supermarket Type3	2303.668	
21	FDU02	13.35	Low Fat	0.10249212	Dairy	230.5352	OUT035	2004	Small	Tier 2	Supermarket Type1	2748.4224	
22	FDN22	18.85	Regular	0.138190277	Snack Foods	250.8724	OUT013	1987	High	Tier 3	Supermarket Type1	3775.086	
23	FDW12		Regular	0.035399923	Baking Goods	144.5444	OUT027	1985	Medium	Tier 3	Supermarket Type3	4064.0432	
24	NCB30	14.6	Low Fat	0.025698134	Household	196.5084	OUT035	2004	Small	Tier 2	Supermarket Type1	1587.2672	
25	FDC37		Low Fat	0.057556998	Baking Goods	107.6938	OUT019	1985	Small	Tier 1	Grocery Store	214.3876	
26	FDR28	13.85	Regular	0.025896485	Frozen Foods	165.021	OUT046	1997	Small	Tier 1	Supermarket Type1	4078.025	
27	NCD06	13	Low Fat	0.099887103	Household	45.906	OUT017	2007		Tier 2	Supermarket Type1	838.908	
28	FDV10	7.645	Regular	0.066693437	Snack Foods	42.3112	OUT035	2004	Small	Tier 2	Supermarket Type1	1065.28	

2.1.1 Store Sales dataset overview

We have trained (8523) and tested (5681) data set, train data set has both input and output variable(s). We need to predict the sales for the test data set.

- **Item_Identifier**: Unique product ID
- **Item_Weight**: Weight of product
- **Item_Fat_Content**: Whether the product is low fat or not
- **Item_Visibility**: The % of the total display area of all products in a store allocated to the particular product
- **Item_Type**: The category to which the product belongs
- **Item_MRP**: Maximum Retail Price (list price) of the product
- **Outlet_Identifier**: Unique store ID
- **Outlet_Establishment_Year**: The year in which store was established
- **Outlet_Size**: The size of the store in terms of ground area covered
- **Outlet_Location_Type**: The type of city in which the store is located
- **Outlet_Type**: Whether the outlet is just a grocery store or some sort of supermarket
- **Item_Outlet_Sales**: Sales of the product in the particular store. This is the outcome variable to be predicted.

2.1.2 Input schema

Feature name	**Datatype**	**Null/Required**
Outlet Identifier	int	Required
Item Fat Content	Object	Required
Item MRP	Float / Int	Required
Item Type	Object	Required
Item Visibility	float	Required
Outlet Establishment year	Int	Required
Outlet Size	Object	Required
Outlet Type	Object	Required

2.2 Predicting Sales

- The system displays the choices of the Outlet Identifier.
- The User chooses the target Outlier Identifier by clicking one of the available Outlier.
- The User gives the Item MRP.
- The system presents the set of inputs required from the user.
- The user gives the required information.
- The system should be able to predict Outlet Sales based on user information.

2.3 Logging

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

- The System identifies at what step logging is required
- The System should be able to log each system flow.
- Developers can choose logging methods. We can choose database logging/ File logging as well.
- The 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.

2.4 Database

We used a given dataset for this project which is a Cassandra database.

2.5 Deployment

❖ HEROKU

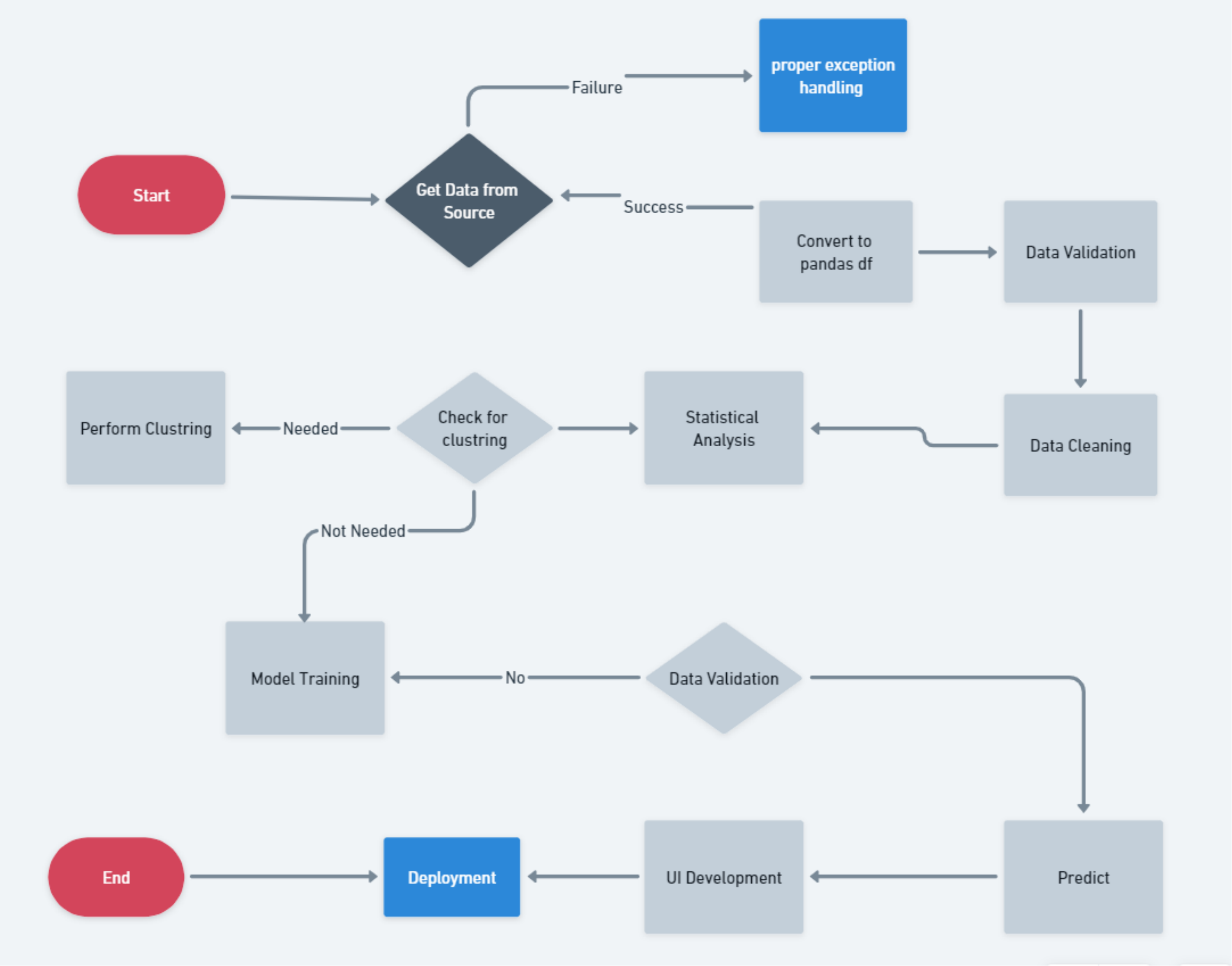
3. Technology stack

Front End	HTML/CSS
Backend	Python Flask
Database	Cassandra AstraDB
Deployment	Heroku

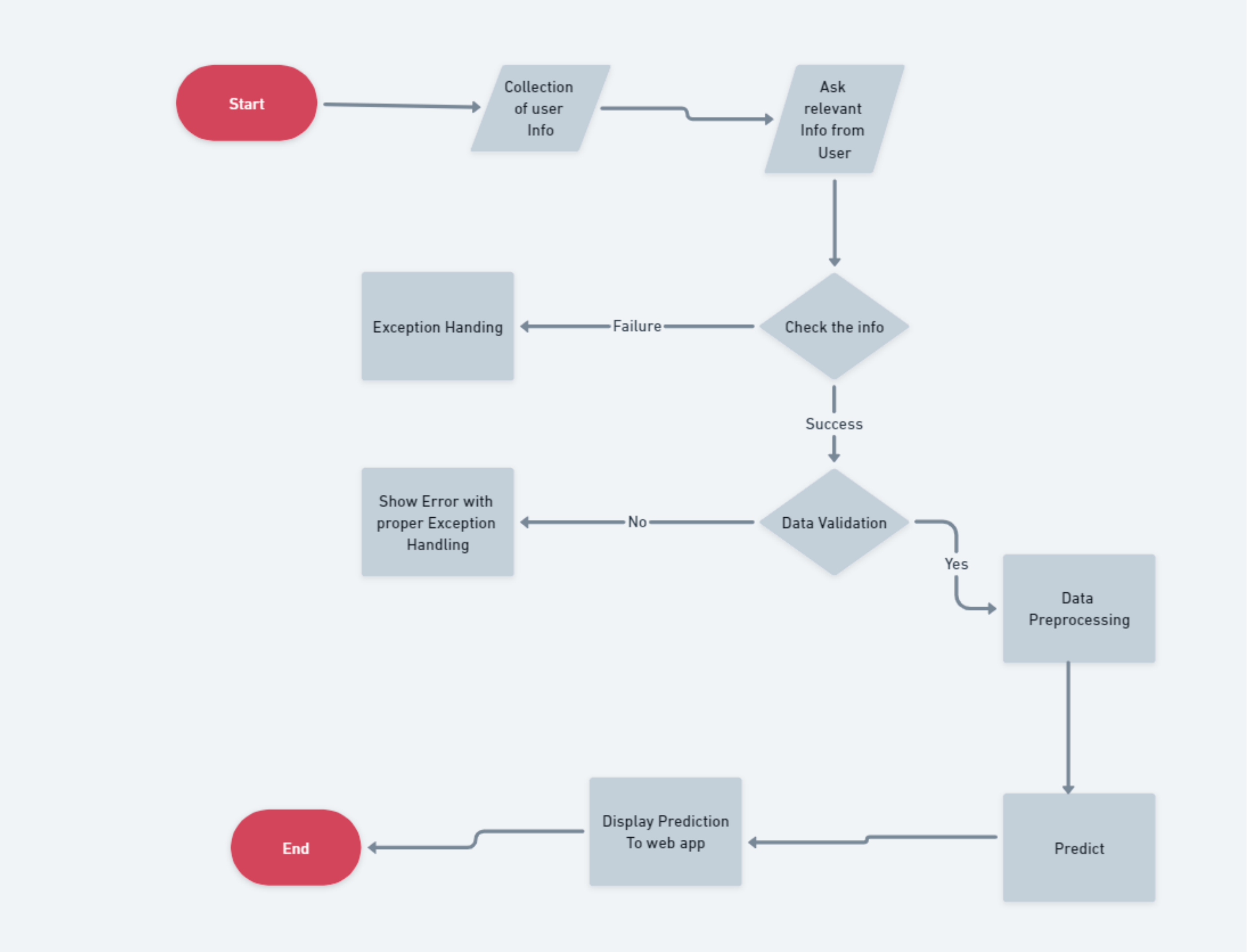
4. Proposed Solution

The proposed solution for this project is a Machine learning algorithm that can be implemented to predict Sales of the Store. Considering various features like Outlet Identifier, Outlet Type, Item MRP, Outlet Establishment year, Outlet Size.. as input from the web app, the implemented regression model will predict the sales.
Here, we have used Gradient Boosting Regressor.
However, drawing a baseline model is important since it tells us how well other models have performed compared to the base model.

5. Model training/validation workflow



6. User I/O workflow



7. Exceptional scenarios

Step	Exception	Mitigation	Module
18th Feb 2022	1.1	First Draft	Dipendra Singh
20th Feb 2022	1.2	Added Workflow chart	Dipendra Singh

8. Test cases

Test case	Steps to perform test case	Module	Pass/Fail

9. Performance

We can observe that the accuracy of the predicted output was seen at 69% using Gradient Boosting Regressor. Other Regression models such as KNeighors regressor and decision tree have given good accuracy above 68% and 68% respectively.