Low Level Design (LLD)

Stores Sales Prediction

Revision Number: 1.6
Last date of revision: 15 Sep 2021

SOMAY

Document Version Control

Date Issued	Version	Description	Author
15 SEP 2021	1.1	First Draft	SOMAY
15 SEP 2021	1.2	Added Workflow chart	SOMAY
15 SEP 2021	1.3	Added Exception Scenarios Overall, Constraints	SOMAY
15 SEP 2021	1.4	Added user I/O flowchart	SOMAY
15 SEP 2021	1.5	Added dataset overview and updated user I/O flowchart.	SOMAY
15 SEP 2021	1.6	Restructure and reformat LLD	SOMAY

Contents

Abstract

The purpose of this research is to construct a **sales prediction** model for retail **stores** using the deep learning approach, which has gained significant attention in the rapidly developing field of machine learning in recent years. Using such a model for analysis, an approach to **store** management could be formulated.

1 Introduction

1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Stores Sales Prediction 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.

The main objective of the project is to predict the stores sales month wise .

Stores Sales is the part of Economy

- Contain a item weight, item fat content, item mrp, item sales outlet year
- Allow access to evidence-based tools that providers can use to make decisions about a sales of the strore.
- Automate and streamline provider workflow

A Stores Sales Prediction contains item details, such as:

- Item sales
- Sales profit
- Item quantity
- Item Location
- Item Value

This project shall be delivered in two phases:

Phase 1: All the functionalities with PyPi packages.

Phase2: Integration of UI to all the functionalities.

1.2 Scope

This software system will be a Web application This system will be designed to predict the Stores Sales and whether it is a compensated, primary or secondary item for better customer management, improved interventions, and more efficient Store sales resource allocation. More specifically, Early predicting of any preventable sales is important for better stores management. This system is designed to predict the Sales Prediction from store information such as item value, item location, item fat content, procedures.

1.3 Constraints

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

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

em_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type
FDA15	9.300	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1
DRC01	5.920	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3
FDN15	17.500	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1
FDX07	19.200	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	Tier 3
NCD19	8.930	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	Tier 3
FDP36	10.395	Regular	0.000000	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3
FDO10	13.650	Regular	0.012741	Snack Foods	57.6588	OUT013	1987	High	Tier 3
FDP10	NaN	Low Fat	0.127470	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3
FDH17	16.200	Regular	0.016687	Frozen Foods	96.9726	OUT045	2002	NaN	Tier 2
FDU28	19.200	Regular	0.094450	Frozen Foods	187.8214	OUT017	2007	NaN	Tier 2

2.1.1 Sales dataset overview

The Sales dataset consists of a table with 5681 records and 11 features. Features are distributed as 7 continuous features and 4 categorical features.

2.1.2 Input schema

Feature name	Datatype	Size	Null/Required
Item_weight	int	3	Required

2.2 Predicting Disease

- The system presents the set of inputs required from the user.
- The user gives required information.
- The system then predicts that the user is having sales is high or not.
  Also, it tells whether a user is having compensated, primary or secondary dataset resources.

2.3 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.

2.4 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 MongoDB.

2.5 Deployment

3 Technology stack

Front End	HTML/CSS
Backend	Python Flask
Database	MongoDB

Deployment	Heroku/AWS

4 Proposed Solution

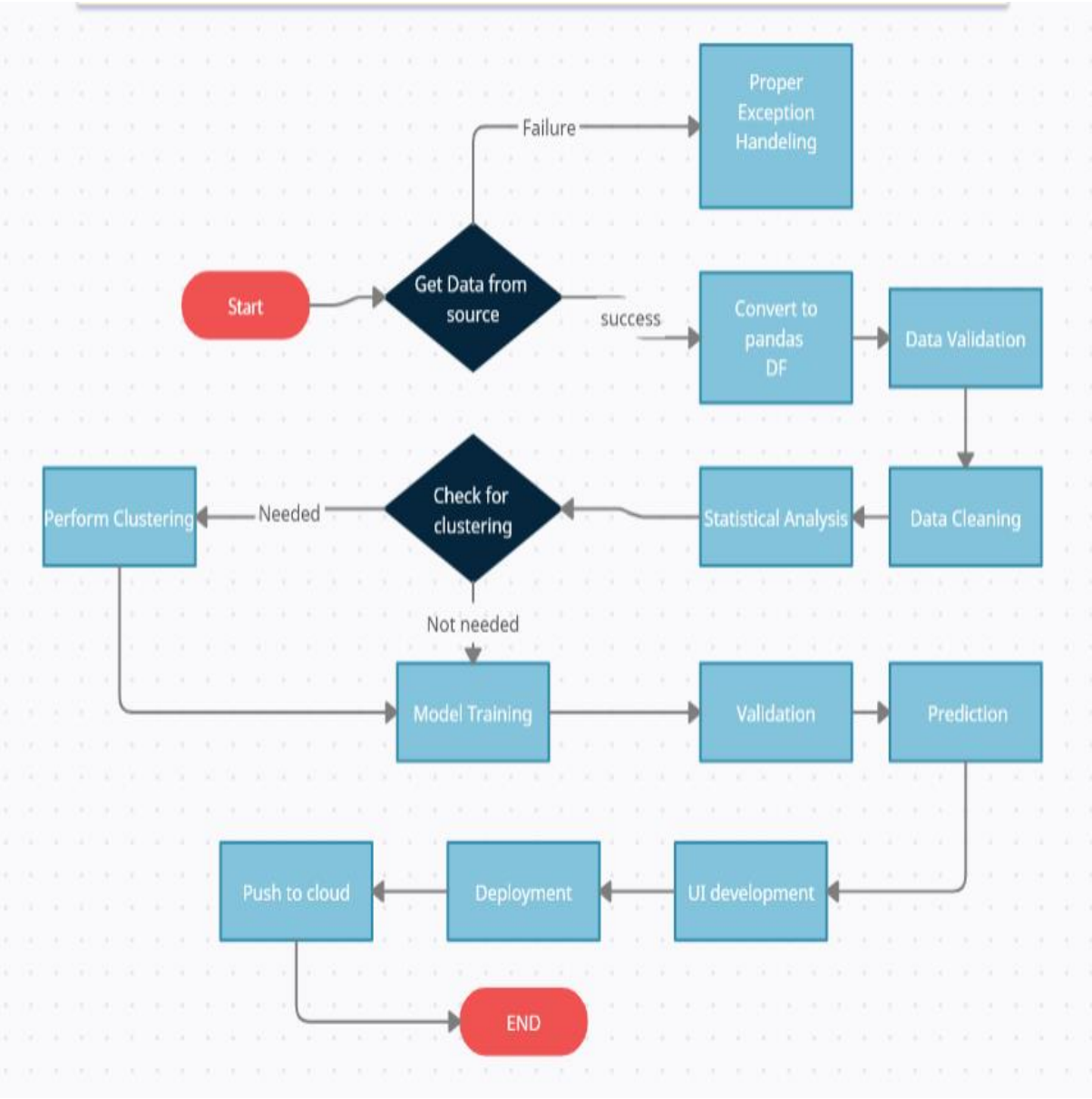
The proposed solution for this project is Machine learning algorithms can be implemented to predictthe stores sales. Considering various features like item type, outlet size, item mrp, item year as inputs from the web app, the implemented classification model will predict the output as store sales

Here, we have used Random Forest Classifier to predict whether the stores sales is high or not.

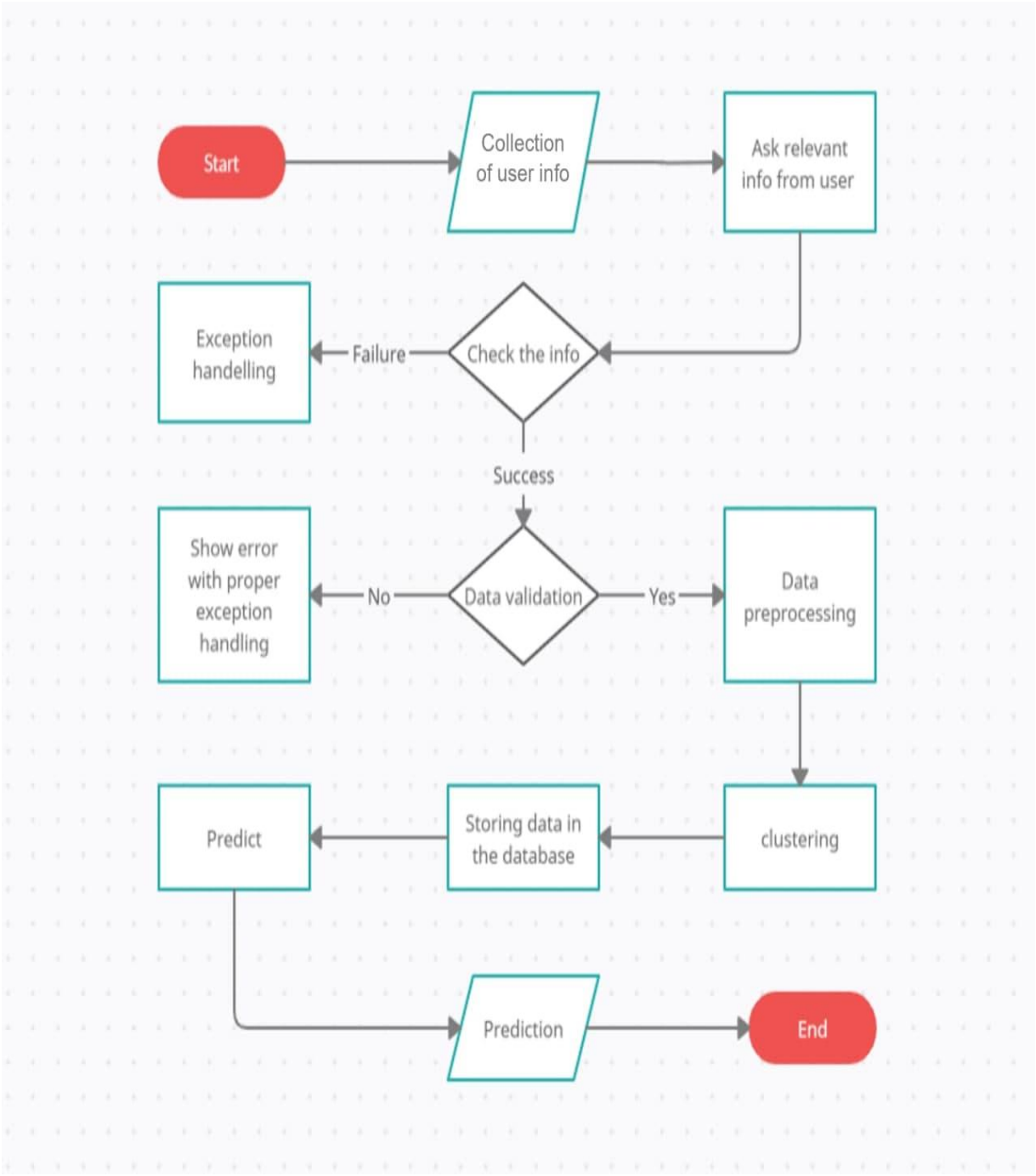
However, drawing a baseline model is important since it tells us how well other models have performed compared to base model. Here, the base model for Store sales dataset isLinear regression

1. Baseline Model: Linear Regression
2. Actual Model: Random Forest

5 Model training/validation workflow



6 User I/O workflow



7 Exceptional scenarios

Step	Exception	Mitigation	Module
15 Sep 2021	1.1	First Draft	Somay
15 Sep 2021	1.2	Added Workflow chart	Somay

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 80% using Random Forestclassifier. Other classification models such as logistic regression and decision tree have given good accuracy above 80