

Low Level Design

Customer Personality Analysis

REVISION NUMBER – 1.2

Last date of Revision: 01/01/2023

Authored by: Aditya Jain

Aditya Sharma

Document Version Control

Date	Version	Description	Author
30/12/2022	1.0	Abstract Introduction Architecture	Aditya Sharma/Aditya Jain
31/12/2022	1.1	Architectural Design	Aditya Sharma/Aditya Jain
01/01/2023	1.2	Deployment Unit test Cases	Aditya Sharma/Aditya Jain

Contents

Document Version Control	2
Abstract.....	4
1. Introduction	5
1.1 What is LLD Document ?	5
1.2 Scope	5
2. Architecture	6
3. Architecture Design	6
3.1 Data Collection	6
3.2 Data Description	7
3.3 Importing Data into Database.....	8
3.4 Exporting Data from Database.....	8
3.5 Data Preprocessing	8
3.6 Model Creation.....	8
3.7 Data from User	9
3.8 Data Validation	9
3.9 Rendering the Results.....	9
4. Deployment.....	10
4.1 Unit Test Cases	10

Abstract

As the legal cannabis industry emerges from its nascent stages, there is increasing motivation for retailers to look for data or strategies that can help them segment or describe their customers in a succinct, but informative manner. While many cannabis operators view the state-mandated traceability as a necessary burden, it provides a goldmine for internal customer analysis. Traditionally, segmentation analysis focuses on demographic or RFM (recency- frequency-monetary) segmentation. Yet, neither of these methods has the capacity to provide insight into a customer's purchasing behavior. With the help of 4Front Ventures, a battle-tested multinational cannabis operator, this report focuses on segmenting customers using cannabis-specific data (such as flower and concentrate consumption) and machine learning methods (K-Means and Agglomerative Hierarchical Clustering) to generate newfound ways to explore a dispensary's consumer base. The findings are that there are roughly five or six clusters of customers with each cluster having unique purchasing traits that define them. Although the results are meaningful, this report could benefit with exploring more clustering algorithms, comparing results across dispensaries within the same state, or investigating segmentations in other state markets

1. Introduction

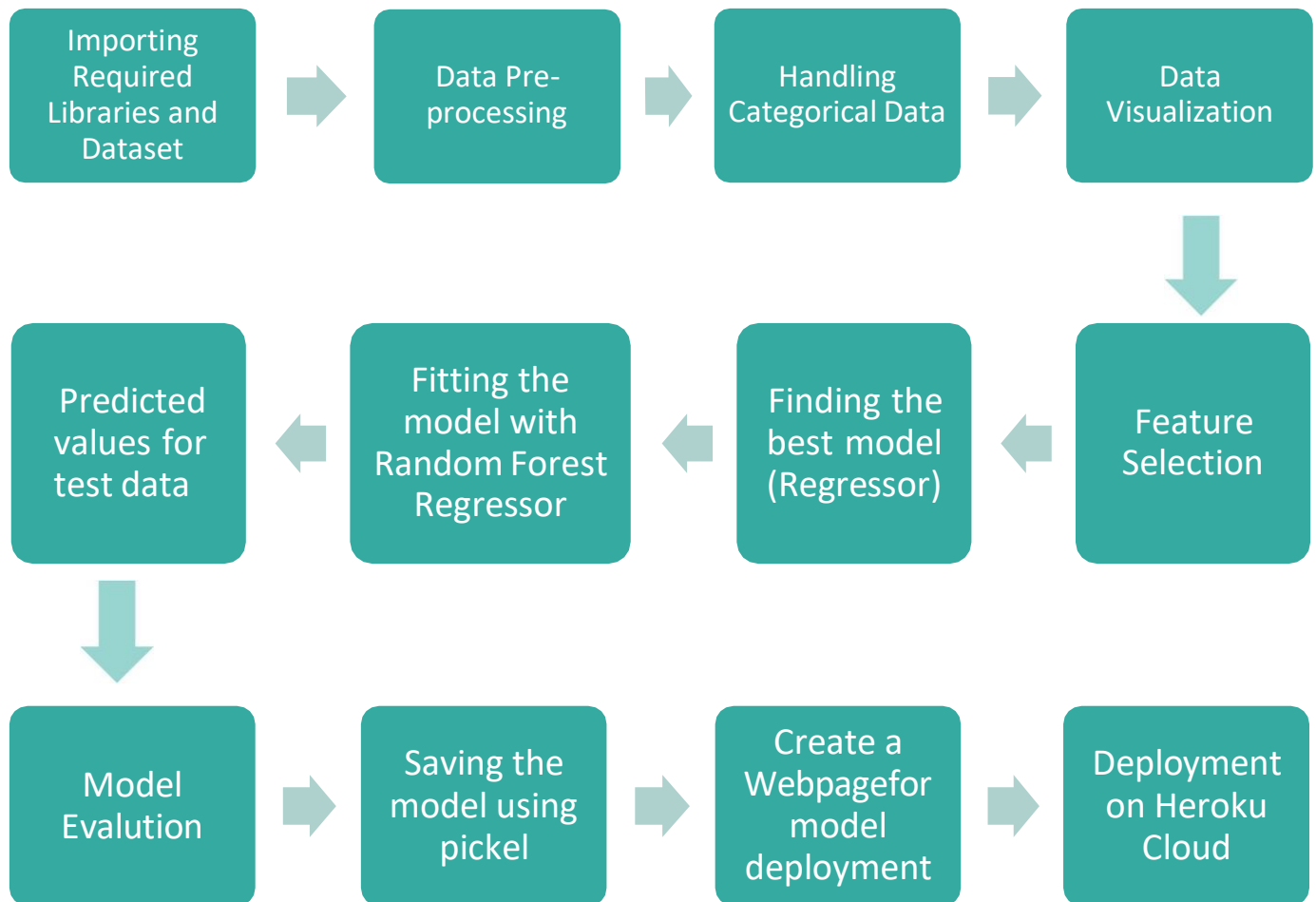
1.1 What is LLD Document ?

The main goal of the LLD document is to give the internal logic design of actual code implementation and supply the outline of the machine learning model and its implementation. Additionally, it provides the description how our project will designed end - to - end.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture



3. Architecture Design

This project is designed to make an interface for the User to predict the customer behaviour.

3.1 Data Collection

The data for this project is collected from the Kaggle Dataset, the URL for the dataset is https://raw.githubusercontent.com/amankharwal/Website-data/master/marketing_campaign.csv

3.2 Data Description

The dataset contain 29 column and 2240 row which give us identity and other detail of customer which help us to predict the customer behavior and the dataset column income contain 24 null values.

First give a brief introduction to all attributes in the dataset we use:

- People
 - ID: Customer's unique identifier
 - Year_Birth: Customer's birth year
 - Education: Customer's education level
 - Marital_Status: Customer's marital status
 - Income: Customer's yearly household income
 - Kidhome: Number of children in customer's household
 - Teenhome: Number of teenagers in customer's household
 - Dt_Customer: Date of customer's enrollment with the company
 - Recency: Number of days since customer's last purchase
 - Complain: 1 if customer complained in the last 2 years, 0 otherwise
- Products
 - MntWines: Amount spent on wine in last 2 years
 - MntFruits: Amount spent on fruits in last 2 years
 - MntMeatProducts: Amount spent on meat in last 2 years
 - MntFishProducts: Amount spent on fish in last 2 years
 - MntSweetProducts: Amount spent on sweets in last 2 years
 - MntGoldProds: Amount spent on gold in last 2 years
- Promotion
 - NumDealsPurchases: Number of purchases made with a discount
 - AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
 - AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise

-
- AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
 - AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
 - AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
 - **Response:** 1 if customer accepted the offer in the last campaign, 0 otherwise
- Place
 - NumWebPurchases: Number of purchases made through the company's web site
 - NumCatalogPurchases: Number of purchases made using a catalogue
 - NumStorePurchases: Number of purchases made directly in stores
 - NumWebVisitsMonth: Number of visits to company's web site in the last month

3.3 Data Preprocessing

- Checked for info of the Dataset, to verify the correct datatype of the Columns.
- Checked for Null values, because the null values can affect the accuracy of the model.
- Converted all the illegal values into legal values.
- Performed Labeled encoding and One hot Encoding on the desired columns.
- Checking the distribution of the columns to interpret its importance.
- Now, the info is prepared to train a Machine Learning Model.

3.4 Model Creation

The Pre - processed info is now envisioned and drawn insights helps us to select the feature that improves the accuracy of the model. The info is randomly used for modelling with different machine learning algorithms to create a model to predict the customer behaviour. After performing on different algorithms, we use Random Forest Regression to create a model and then also perform

Hyperparameter Tuning to improve the accuracy of the model.

3.5 Data from User

The data from the user is retrieved from the created streamlit web page.

3.6 Data Validation

The data provided by the user is then being processed by streamlit file and validated. The validated data is then sent to the prepared model for the prediction.

4. Deployment

The tested model is then deployed to Heroku. So, users can access the project from any internet devices.

4.1 Unit test case

Test Case Description	Pre-Requisites	Expected Results
Verify whether the Webpage is accessible to the User or not.	Webpage URL should be defined.	Webpage should be accessible to the User.
Verify whether the Webpage is completely loads for the User or not	1. Webpage URL is accessible. 2. Webpage is deployed.	The Webpage should be completely loads for the User when it is accessed.
Verify whether the User is able to enter data in input fields or not.	1. Webpage URL is accessible. 2. Webpage is deployed. 3. Webpage input fields are editable.	The User is able to enter data in input fields.
Verify whether the User is able to submit details or not.	1. Webpage URL is accessible. 2. Webpage is deployed. 3. Webpage input fields are editable.	The User is able to submit details to process.
Verify whether the User gets recommended results on submitting the details or not.	1. Webpage URL is accessible. 2. Webpage is deployed. 3. Webpage input fields are editable	The User gets recommended results on submitting the details.