High Level Design (HLD)

Purchasing Capability of Customers

Document Version Control

Date Issued	Version	Description	Author
25-Sep-2021	1	Update HLD Description.	Prasad Aher.
28-Sep-2021	2	Update Index	Subhabrata Nath
2-Oct-2021	3	Update Work Flow	Vinayak Dumbre.

INDEX

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 hardware andsoftware 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**

2 General Description

2.1 Problem Statement:

Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs. Using the above data, companies can then outperform the competition by developing uniquely appealing products and services.

2.2 Proposed Solution :

The most common ways in which businesses segment their customer base are:

1. **Demographic information includes gender, age, familial and marital status,**
   income, education, and occupation.
2. **Geographical information which differs depending on the scope of the company. For localized businesses, this info might pertain to specific towns or counties. On the other hand, it might mean a customer's city, state, or even country of residence for larger companies.**
3. **Psychographics, such as social class, lifestyle, and personality traits.**
4. **Behavioural data, such as spending and consumption habits, product/service usage, and desired benefits.**

Advantages of Customer Segmentation

1. **Determine appropriate product pricing.**
2. **Develop customized marketing campaigns.**
3. **Design an optimal distribution strategy.**
4. **Choose specific product features for deployment.**
5. **Prioritize new product development efforts. You have to create a project to segment a customer for a banking client based on their transaction history. You have to find out the purchasing**

capabilities of customers for a different part of the customer.

Dataset Description :

This case requires developing a customer segmentation to define Purchasing Capability of Customer. The sample Dataset summarizes the usage behaviour of about 9000 active credit card holders over the period of 6 months. The file is at a customer level with 18 behavioural variables.

Following is the Data Dictionary for Credit Card dataset :-

CUSTID	**: Identification of Credit Card holder (Categorical)**
BALANCE	**: Balance amount left in their account to make purchases (**
BALANCEFREQUENCY	**: How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)**
PURCHASES	**: Amount of purchases made from account**
ONEOFFPURCHASES	**: Maximum purchase amount done in one-go**
INSTALLMENTSPURCHASES	**: Amount of purchase done in instalment**
CASHADVANCE	**: Cash in advance given by the user**
PURCHASESFREQUENCY	**: How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)**
ONEOFFPURCHASESFREQUENCY	**: How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)**
PURCHASESINSTALLMENTSFREQUENCY	**: How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)**
CASHADVANCEFREQUENCY	**: How frequently the cash in advance being paid**
CASHADVANCETRX	**: Number of Transactions made with "Cash in Advanced"**
PURCHASESTRX	**: Numbe of purchase transactions made**
CREDITLIMIT	**: Limit of Credit Card for user**
PAYMENTS	**: Amount of Payment done by user**

MINIMUM_PAYMENTS	**: Minimum amount of payments made by user**
PRCFULLPAYMENT	**: Percent of full payment paid by user**
TENURE	**: Tenure of credit card service for user**

2.4 Tools Used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask, Matplotlib, Seaborn, and are used to build the whole model.

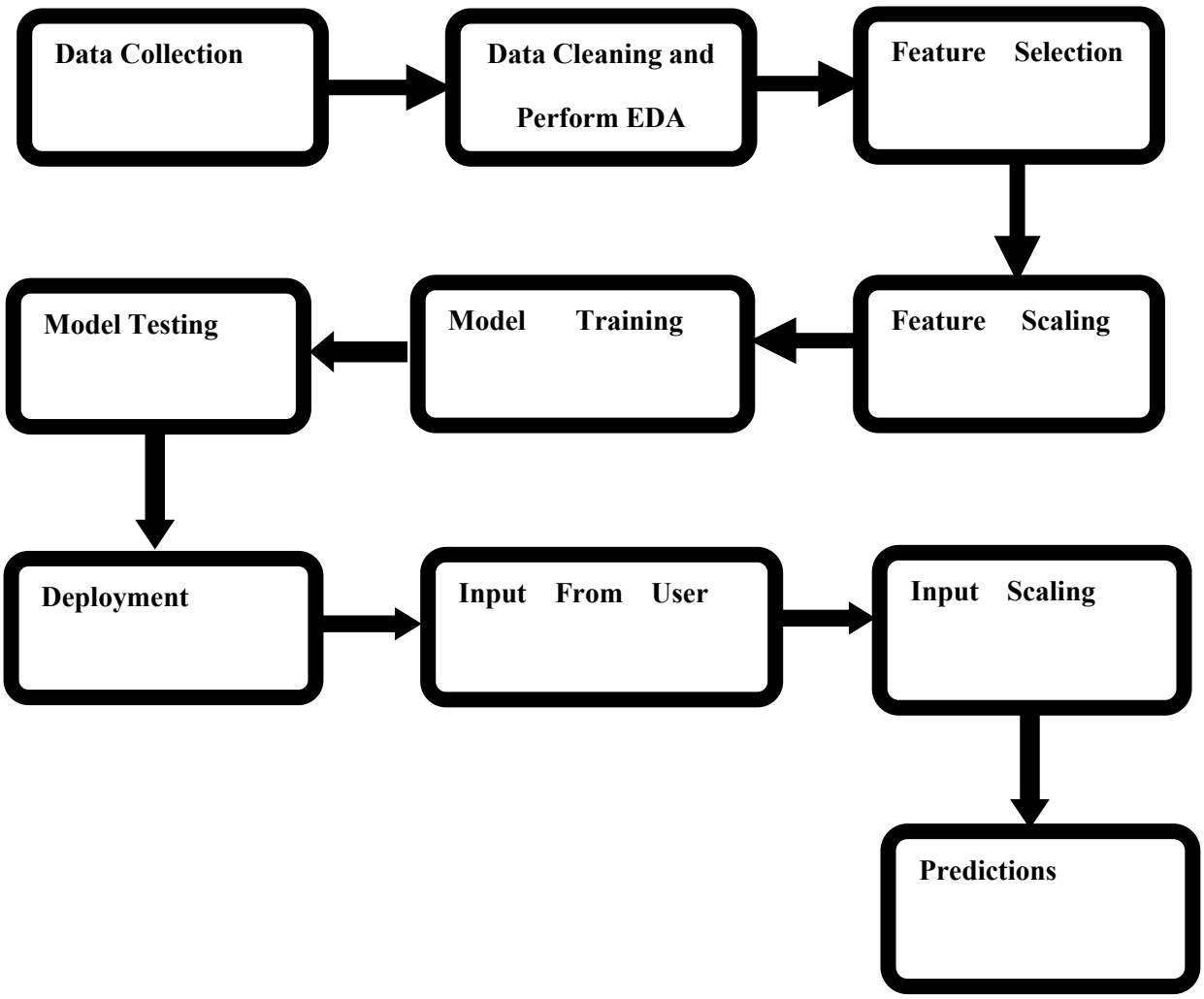
- **Pycharm / VsCode is used as IDE.**
- **For visualization of the plots, Matplotlib, Seaborn and Plotly are used.**
- **AWS / Azure / GCP/Heroku is used for deployment of the model.**
- **Cassandra is used to retrieve, insert, delete, and update the database.**
- **Front end development is done using HTML/CSS**
- **Python Flask is used for backend development.**
- **GitHub is used as version control system.**

2.5 Constraints:

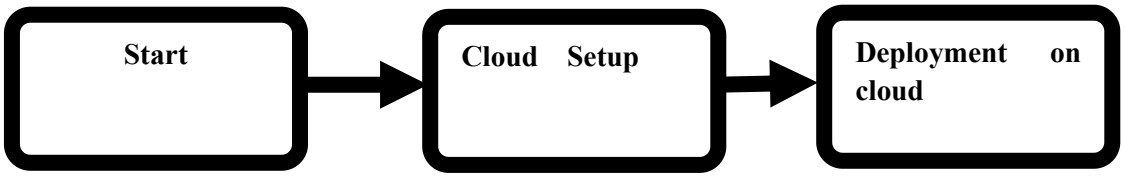
The Customer Segmentation solution must be user friendly, data validation and model selection must be as accurate as possible, and users should not be required to know any of the workings.

3 Design Details

3.1 Process Flow



3.2 Deployment



3.2 Event Log:

The system should log every event so that the user will know what 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

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

4. Performance

Solution of Customer Capabilities Apllication is used to predict the customer belongs to which group based on similarity between the spending ,earning and relative kinds of behaviour.so it should give as much as accurate possible results.

That's why before building this model we followed complete process of Machine

Learning. Here are summary of complete process:

1. We cleaned our dataset properly by removing all null value.

2. We perform feature scaling.

3. we have performed PCA for visualization of Clusters.

4. we train our model on kmeans Clustering.

5. we saved the model in pickel format for the furthur predictions.

6.After that our model was ready to deploy. I deployed this model on

heroku.

4.1 Re-usability

We have done programming of this project in such a way that it should be

reusable. So that anyone can add and contribute without facing any problems

4.2 Application Compatibility

The different module of this project is using Python as an interface between them.

Each modules have it is own job to perform and it is the job of the Python to ensure the proper transfer of information.

4.3 Resource Utilization

In this project, when any task is performed, it will likely that the task will use all the processing power available in that particular system until it's job finished.

5.0 Conclusion

To group the customers based on their spending,earning and behavioral data