☐ High Level Design(HLD)

Purchasing capabilities of a customer

Document Version Control

Date Issued	Version	Description	Author
23-7-22	1	Update HLD Description	Anwesha Das
24-7-22	2	Update index	
	3	Update workflow	

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 ::

o Present all of the design aspects and define them in detail
o Describe the hardware and software interfaces
oDescribe the performance requirements
o Include design features and the architecture of the project
o List and describe the non-functional attributes like:
o Security
o Reliability
o Maintainability
o Portability
o Reusability
o Application compatibility

- Resource utilisation
- 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.

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 localised 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.

Dataset Description:: The shape of data is `(8950, 18)`.

The detailed about Dataset

Custid - identification of Credit Card holder (Categorical)

BALANCE - Balance amount left in customers account to make purchases

BALANCE_FREQUENCY - How frequently the Balance is updated, score between 0 and 1

PURCHASES - Amount of purchases made from account

ONEOFF_PURCHASES - Maximum purchase amount done in one-go

INSTALLMENTS_PURCHASES - Amount of purchase done in instalment

CASH_ADVANCE - Cash in advance given by the user

PURCHASES_FREQUENCY - How frequently the Purchases are being made, score between 0 and 1

ONE OFF PURCHASE FREQUENCY - How frequently Purchases are happening in one-go

PURCHASE INSTALMENTS FREQUENCY - How frequently purchases in instalments are being done

CASH ADVANCE FREQUENCY - How frequently the cash in advance being paid

CASH ADVANCE TRX - Number of Transactions made with "Cash in Advance"

PURCHASES_TRX - Number of purchase transactions made

CREDIT_LIMIT - Limit of Credit Card for user

PAYMENTS - Amount of Payment done by user

MINIMUM_PAYMENTS - Minimum amount of payments made by user

PRC FULL PAYMENT - Percent of full payment paid by user

TENURE - Tenure of credit card service for user

2.4 Tools Used-

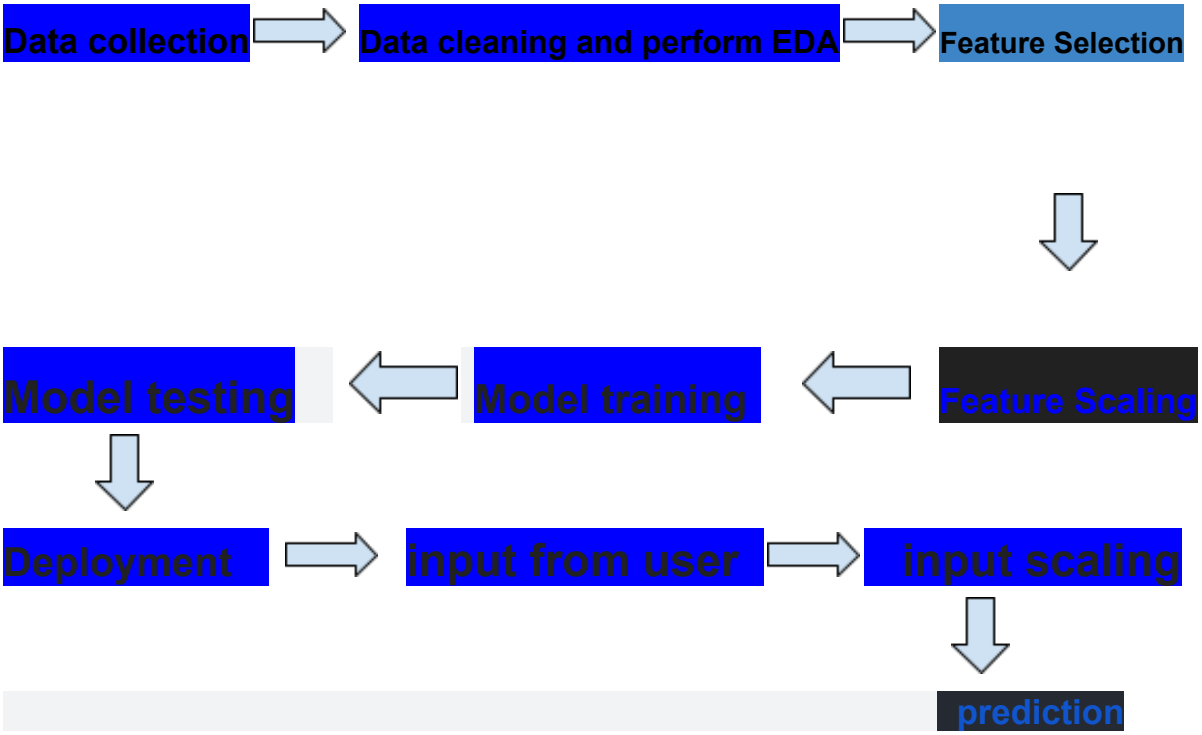
Python programming language and frameworks such as NumPy, Pandas, Scikit Learn, Flask, Matplotlib, Seaborn are used to build the whole model.

- VsCode is used as an IDE.
- For visualisation of the plots Matplotlib, Seaborn and Plotly are used.
- Heroku is used for deployment of the model.
- Front end development is done using HTML/CSS
- Python Flask is used for backend development.
- GitHub is used as a 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. Developers can choose logging methods. We 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 Application 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 the complete process of Machine Learning. Here is a summary of the complete process:
1. We cleaned our dataset properly by removing all null value.
2. We perform feature scaling.
3. We have performed PCA for visualisation of Clusters.
4. We train our model on k means Clustering.
5. We saved the model in pickle format for the further 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 problem

4.2 Application Compatibility

The different module of this project is using Python as an interface between them. Each module has its own job to perform and it is the job of the Python to ensure the proper transfer of information.

4.3 Resource Utilisation

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

5.0 conclusion

To group the customers based on their spending,earning and behaviour of data.