# HIGH LEVEL DESIGN

CUSTOMER SEGEMENTATION

## DOCUMENT VERSION CONTROL

| DATE ISSUED | VERSION | DESCRIPTION | AUTHOR |
|---|---|---|---|
| 3/7/2024 | 1 | Initial HLD v1 | Abishek kumar |
| | | | |

# Table of Contents

## Abstract

In today's highly competitive business environment, understanding customer bahavior is curical for companies aiming to enhance their marketing strategies and optimize customer experiences. This project focuses on developing a mahine learning model for customer experiences.Project aim to identify unique customer segments that can be targeted with personalized marketing efforts.

The project employs machine learning techinque, K-means clustring, to analyze customer data nad uncover hidden pattens. Extensive preprocessing steps, such as data cleaning, normalization, and feature engineering, are performed to enhance tha quality and reliability of the input data. The ultimate guide of this project is to imorve customer satiisfaction and increased revenue.

# 1. Introduction

## 1.1 Why This high level documentation

This purose 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 referece manual for how the modules interact at a high level.

The HLD will

- Present all of the design  ascpect and define them in detail.
- Describe the user interface being implementd
- Describe the hardware and software interface.
- Describe the performance requirements.
- Include design features and the architecture of the project.
- List the describe and non functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatiability
    - Serviceability

## 1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture (layers), application flow (navigation), and technology architecture. The HLD uses non technical to mildy-technical terms which should be understandable to the administrators of the system.

## 1.3 Definitions

Customer segmentation is the process of dividing a company's customer base into distinct groups of individuals that share similar characteristics relevant to marketing, such as buying habits, interests, and demographics. This allows businesses to tailor their marketing strategies to meet the specific needs of each segment.

# 2 General Description

## 2.1 Product perspective

Customer segmentation is machine learning based cluster the customer based on their income and expences. By integrating this model, businesses can gain actionable insights into their customer base. Allowing for more personalized and effective marketing stretegies.

## 2.2 Probelem Statement

In today's competitice market, businesses struggle to effectively unserstand cater to the deiverse neesd of their customer base. Traditional marketing strategies often fall short in addressing invividual customer perferences, leading to suboptimal engagement and lost revenue opportunities, reduce customer satisfaction. This project aims to address these challenges by developing a machine learning based project by identifying these segements, and companies can tailor their marketing strategies to meet the specific needs of each group.

## 2.3 Purposed Solution

To address the challenges of effectivelu understanding and catering to diverse customer needs, this project proposes the development of a machine learning based customer segementaion model .

- Data collection and preparation
- Clustering Alogo implementation.
- Model evaluation and validation
- Integrating and deployment.

By implementing this solution, businesses can gain actionable insights into customer behaviour, allowing for personalized marketing efforts, imporved customer engagement.

## 2.4 Technical Requirements

To sucessfully develop and implement the customer segmentation model, the follwing

Technical like

- Data infrastructure.
- Processing and preprocessing,
- Machine learning environement,
- Computational resources.
- Model Evaluation and Validation.
- Integration and Deployment.
- User interface and Reporting.

Data infrastructure access to comprehensive customer data, including transactional record, demographic information, and interaction history. Tools for data cleaning and preprocessing, such as python libararies (pandas , numpy) used and capabilities for feature extraction and engineering to create meaningful input features for the model.

For machine lerning environment support jupyter notebooks for expoloratory data analysis and model development.machine learning libraries such as scikit-learn for clustering algo, and additional libraries like SciPy and Matplotlib for data analysis and visualization. Hardware computational resources, including high-performance CPU or GPU for training and testing the machine learning models. Utilization of colud computing platforms AWS for scalable computing power and storage as needed. And Tools to compute evaluation metrics such silhouette score for evaluation and libararies for viualizing clusters and segmentation results, such as Matplotlib, Seaborn, or Plotly.

For Integration Deployment  Frameworks (Flask FastAPI) for developing APIs to integrate the segmentation model with existing CRM systems. Plateforms for deploying model in production environment, such as Docker for containerization

By meeting theses tchnical requirement, the project can effectively develop, validate and deploy a robust customer segmentation model that integrates seamlessly with existing business processes and provides valuable insights for personalied marketing strategies.

## 2.5 Data Requirements.

The Data require for the model is completely depend upon problem statement.

- The customer segmentation dataset gather form kaggle.
- Dataset contain features like Gender, every marreied, income level, age, gender, education, spending score, Family size. Having 10000 datapoint.
- Well balanced dataset having four different target categories.

The dataset used for customer segmentaion project contains various features that provide detailed insights into each customer demograpic and behavioral charcheristics.

Age feature represent the age of customer having numerical types and age reanges from 18 to 65 in dataset, covering broad spectum of age groups. Age is curical demographic factor that influences purchasing behavior and perferences. Different age groups may exhibit distinct shopping patterns, helping to identify segements such as young adults, middle-aged progessional, and seniors.

Gender indicates the gender of the customer type like male, female, other. Commonly includes categories like ‘male’, ‘female’ and others. Gender can significantly impact shopping habits and product perferences. Segementing customer based on gender allows for more targeted marketing strategies.

Income level represent the annual income of the customer. It is categorical data having value (‘high’, ‘low’, ‘average’). Income level is a critical factor in understanding a customer’s purchasing power and spending capacity. It helps in creating segments based on economic status, enabling the design of appropriate product offerings and pricing strategies.

This feature describes the customer’s occupation or professional field. Categorical like student, engineer, techer, manager. Profession can affect a customer’s spending habits and product preferences. Understanding the professional background helps in creating segments that can be targeted with relevent marketing messages.

The dataset provides a comprehensive view of each customer's demographic and behavioral attributes, which are essential for effective customer segmentation. By analyzing these features, the machine learning model can identify distinct customer segments, enabling businesses to tailor their marketing strategies to the specific needs and preferences of each group. This targeted approach can lead to improved customer satisfaction, higher engagement, and increased revenue.


scikit
learn
git


Pandas


matplotlib
docker

## 3. Tools Used for Project

- Visual Studio Code is used as IDE.
- For Visualization of the plot, Matplotlib, seaborn are used.
- Pandas is used for feature engineering and data cleaning.
- AWS is used for deployment of the model
- Mongodb is used to retrive, insert, delete and update the database.
- Front end development is done using HTML/CSS.
- Github is used to as version control system.

## 4. Constrains Of Project

The customer segmentation project faces several constraints that need careful consideration to ensure successful implementation and accurate results. One of the primary constraints is data quality and availability. Incomplete or inconsistent data records can significantly impact the reliability of the segmentation model. Additionally, strict data privacy regulations, such as GDPR and CCPA, may limit access to certain customer data, necessitating robust compliance measures to protect customer privacy.

Computational resources also pose a significant constraint, as handling large datasets and complex algorithms requires substantial processing power and storage capacity. Selecting the appropriate clustering algorithm, such as K-means, hierarchical clustering, or DBSCAN, is critical to effectively addressing the specific characteristics of the data. However, scalability remains a challenge, as the algorithms must efficiently handle increasing data size and complexity.

Model performance is another constraint, with a particular emphasis on avoiding overfitting, which can lead to poor generalization on new data. Choosing suitable evaluation metrics is essential to accurately assess the quality of the segmentation. Integration challenges arise when ensuring that the segmentation model seamlessly integrates with existing CRM systems and data warehouses. Additionally, fostering user adoption among marketing and sales teams is crucial for the effective utilization of segmentation insights.

Interpretability of the segmentation results is vital, as the outputs must be understandable and actionable for business users, not just data scientists. Providing clear and intuitive visualizations helps stakeholders grasp the segmentation results. Furthermore, the project must adhere to

the timeline and budget constraints, including costs associated with data acquisition, computational resources, and tool licenses.

## 5. Conclusion

The segmentation results provide invaluable insights that help businesses tailor their offerings, improve customer engagement, and optimize resource allocation. This targeted approach not only enhances customer satisfaction but also drives revenue growth by focusing marketing efforts on the most relevant customer segments.

Despite the challenges related to data quality, computational resources, model performance, and integration, the project demonstrates that with careful planning and execution, these constraints can be effectively managed. The importance of ensuring data privacy, maintaining model scalability, and fostering user adoption are crucial takeaways for similar future projects.

In conclusion, the customer segmentation model serves as a vital tool for businesses seeking to understand and cater to their diverse customer needs. By leveraging machine learning, companies can transform raw data into actionable insights, ultimately achieving higher efficiency, improved customer loyalty, and sustained business growth. This project lays a strong foundation for ongoing improvements and adaptations, ensuring that the segmentation model remains relevant and valuable in the ever-evolving market landscape.