**Architecture Design**

**Customer Personality Analysis**

Revision Number – 1.0

# Last Date of Revision: 12 – 05 -2024

Harshal Zarikar

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 12– 05 - 2024 | 1.0 | First Draft | Harshal Zarikar |

**Contents**

1. Abstract

2. Introduction

2.1 Why architecture Design Document?

2.2 Scope

2.3 Definitions

3. Technical specification

3.1 Dataset overview

3.2 Segmentation of customers into groups

3.3 Logging

4. Technology Stack

5. Proposed Solution

6. Workflow

7. Key Performance indicators (KPI)

**1.Abstract**

The Architecture Design Document for Customer Personality Analysis outlines the technical specification and proposed solution for clustering customers into groups based on various factors. It covers the dataset overview, clustering customers and predicting which cluster a customer belongs to, logging, technology stack, workflow, and key performance indicators (KPIs) to ensure successful implementation.

**2.Introduction**

**2.1 Why Architecture Design Document?**

An Architecture Design Document is essential to ensure the successful development and deployment of a project. It helps to define the technical specification, technology stack, and proposed solution for a project.

**2.2 Scope**

The scope of this Architecture Design Document is to provide technical specifications and a proposed solution for analysing customers’ personality based on various factors.

**2.3 Definitions**

Customer Personality Analysis: Customer personality analysis is the process of identifying and understanding the unique characteristics and traits that make up an individual customer's personality. This information can be used by companies to tailor their marketing and sales efforts to better target and serve each customer's specific needs and preferences.

Dataset: A collection of data used to train and test the machine learning model.

**3.Technical Specification**

**3.1 Dataset Overview**

The dataset used for this project contains information on the customer's date of birth, income, expenditure, education level, marital status, number of kids and other relevant factors.it is segmented into clusters for understanding particular cluster. Then, it is split into training and testing datasets to train and test the machine learning model for predicting the cluster to which a customer might belong based on his/her data.

**3.2 Analysing customer personality**

The data is clustered into 3 groups using unsupervised machine learning model. Then the cluster and data are used in training a classification model. The model is then evaluated on the testing dataset to ensure that it performs well on unseen data.

**3.3 Logging**

Logs are used to track the process of cluster formation, evaluation. It is also used to track the performance of the machine learning model during training and testing.

**4.Technology Stack**

The technology stack used for this project includes:

**Python:** The programming language used for data preprocessing and model development.

**Scikit-learn**: A machine learning library used for model development and evaluation.

**Streamlit**: Streamlit is an open-source Python library that allows users to create interactive web apps to predict insurance premiums and making it easy to share and visualize data.

**5.Proposed Solution**

The proposed solution for analysing the customer’s personality involves the following steps:

**Data preprocessing**: The dataset is cleaned and pre-processed to remove missing values and outliers.

**Model development**: A machine learning algorithm is used on the pre-processed dataset to segment the dataset into clusters. The clusters formed are used in the classification algorithm to predict the cluster a particular customer may belong to.

**Model evaluation**: The model is evaluated on the testing dataset to ensure that it performs well on unseen data.

**Web application**: A web application is created using Streamlit to allow users to input their information and receive a cluster number that customer belongs to.

**6.Workflow**

The workflow for this project is as follows:

**Data collection:** Collect the dataset containing information on customers and their income and spending information on a product.

**Data preprocessing**: Clean and preprocess the dataset to remove missing values and outliers.

**Model development**: Train a machine learning algorithm on the pre-processed dataset to cluster the customers data.

**Model evaluation**: Evaluate the model's performance on the testing dataset to ensure that it performs well on unseen data.

**Web application development:** Develop a web application using Streamlit to allow users to input their information receive a cluster number that customer belongs to.

**Deployment**: Deploy the web application to a server for public use.

**7.Key Performance Indicators (KPIs)**

The following KPIs will be used to measure the success of the project:

**Model accuracy**: The accuracy of the machine learning model in segmenting the customers’ data into clusters.

**User engagement**: The number of users who interact with the web application.

**Response time**: The time it takes for the web application to respond to user requests.

**Server uptime**: The percentage of time the server hosting the web application is operational.