Gaurav Ghosh

INEURON

ARCHITECTURE

Customer Segmentation and Clustering

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issue | Version | Description | Author |
| 12-12-2024 | 1.1 | First Draft |  |
| 12-12-2024 | 1.2 | Added Workflow chart |  |
| 12-12-2024 | 1.3 | Added Exception Scenarios Overall,  Constraints |  |
| 12-12-2024 | 1.4 | Added KPIs |  |
| 12-12-2024 | 1.4 | Added user I/O flowchart |  |
| 19-12-2024 | 1.5 | Added dataset overview and updated user  I/O flowchart. |  |
| 23-12-2024 | 1.6 | Restructure and reformat LLD |  |
|  | 1.7 |  |  |
|  | 1.8 |  |  |

Contents

[Document Version Control 1](#_Toc188652054)

[Abstract 3](#_Toc188652055)

[1 Introduction 4](#_Toc188652056)

[1.1 Why is this Low-Level Design Document? 4](#_Toc188652057)

[1.2 Scope 4](#_Toc188652058)

[1.3 Constraints 4](#_Toc188652059)

[1.4 Risks 4](#_Toc188652060)

[1.5 Out of Scope 4](#_Toc188652061)

[2 Technical Specification 5](#_Toc188652062)

[2.1 Dataset 5](#_Toc188652063)

[2.1.1 Customer dataset overview 5](#_Toc188652064)

[2.2 Predicting Cluster 5](#_Toc188652065)

[2.3 Logging 5](#_Toc188652066)

[2.4 Deployment 5](#_Toc188652067)

[3 Technology Stack 6](#_Toc188652068)

[4 Proposed Solution 7](#_Toc188652069)

[5 Model training / Validation workflow 8](#_Toc188652070)

[6 User I/O workflow 9](#_Toc188652071)

[9 Key Performance Indicator (KPI) 10](#_Toc188652072)

# Abstract

In today's increasingly competitive market, selling products has become a challenging task due to shifting customer needs and desires, as well as the growing number of retailers and sellers. Customer Segmentation and Clustering models offer valuable insights into customer preferences and behaviour. By analysing clustered customer data, retailers can develop targeted strategies to strengthen their business and meet customer expectations effectively.

# 1 Introduction

## 1.1 Why is this Low-Level Design Document?

The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Customer Segmentation and Clustering. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

## 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.

## 1.3 Constraints

We will be selecting particular attribute column for predictions.

## 1.4 Risks

* Limited data can hinder meaningful pattern detection.
* Overrepresented groups may bias clustering outcomes.
* The model might struggle with large datasets or real-time updates.

## 1.5 Out of Scope

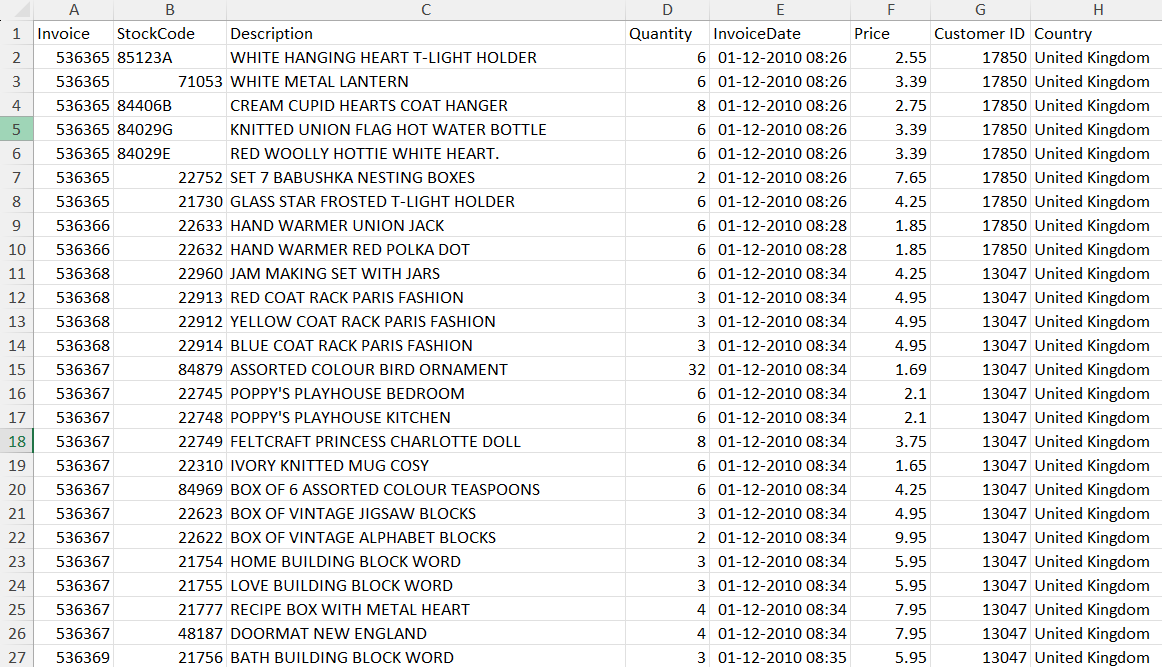
Delineate specific activities, capabilities, and items that are out of scope for the project.

# 2 Technical Specification

## 2.1 Dataset

### 2.1.1 Customer dataset overview

The dataset contains over 10000 entries, including fields such as **Invoice, StockCode, Discription, Quantity, InvoiceDate, Price, Customer ID and Country**. This dataset provides valuable information for training the model to achieve its optimal performance.



## 2.2 Predicting Cluster

* The system will ask for the CSV file to be predicted.
* The user will enter the CSV file to be predicted.
* The system will provide Elbow graph, Silhouette Score graph and 2D and 3D Scatter plot graph for better understanding.

## 2.3 Logging

* The system uses a CSV file for logging.
* The system will not hung up with multiple logging.

## 2.4 Deployment

* Github

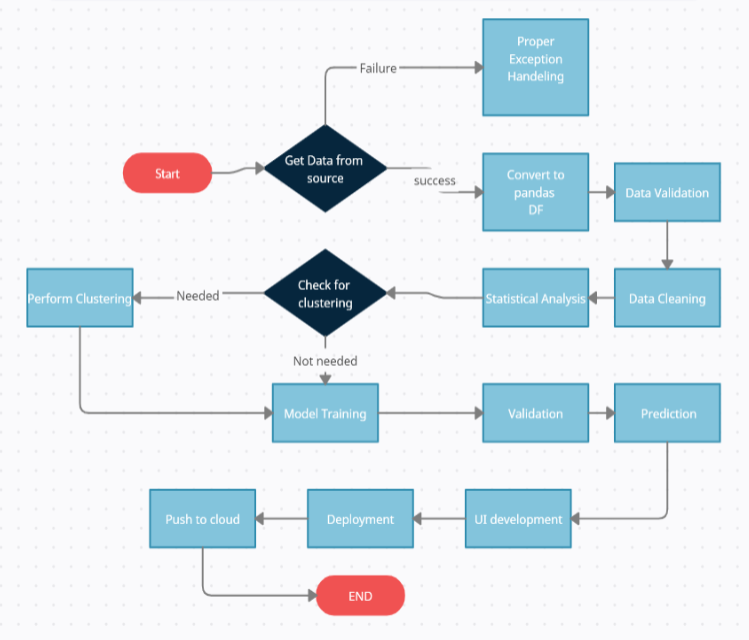
# 3 Technology Stack

|  |  |
| --- | --- |
| Front End | HTML, CSS |
| Back End | Python |
| Database | - |
| Deployment | Github |

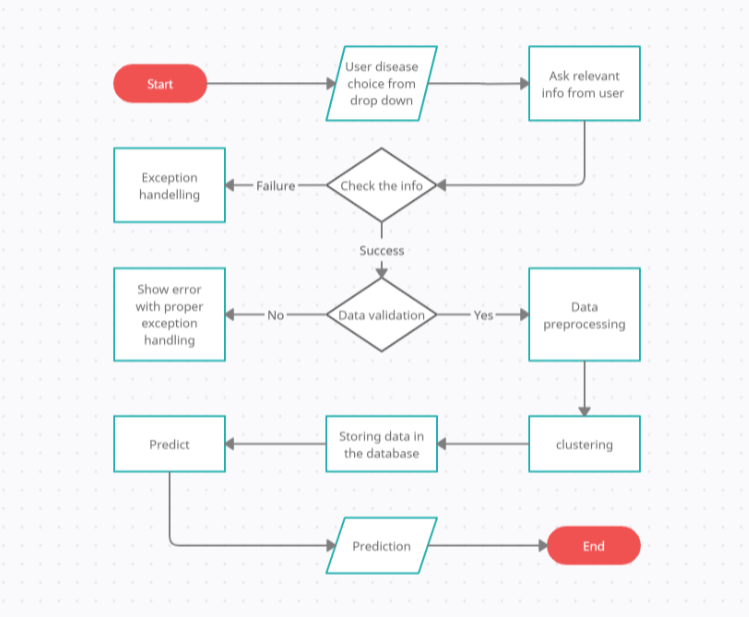
# 4 Proposed Solution

Based on the research paper, we utilize historical data to make predictions, with K-Means clustering identified as the most suitable approach. This machine learning algorithm clusters data based on past purchase behaviour, allowing for effective segmentation. To ensure reliability, a baseline model is first developed. This baseline is crucial as it validates the accuracy of predictions and establishes trust in the model's outcomes, enabling its use for future predictions.

# 5 Model training / Validation workflow



# 6 User I/O workflow



# 7 Exceptional Scenarios

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Exception | Mitigation | Module |
|  | 1.1 | First Draft | Gaurav Ghosh |
|  | 1.2 | Added Workflow chart |  |

# 8 Test Cases

|  |  |  |  |
| --- | --- | --- | --- |
| Test case | Steps to perform  test case | Module | Pass/Fail |
|  |  |  |  |
|  |  |  |  |

# 9 Key Performance Indicator (KPI)

* How well the data is grouped into meaningful clusters.
* How distinct the clusters are from each other.
* Speed of creating clusters from the data.
* Increase in revenue from targeted marketing strategies.
* Percentage of customers in each group making purchases or taking desired actions.
* Effectiveness of clustering in guiding business strategies.