Low Level Design (LLD)

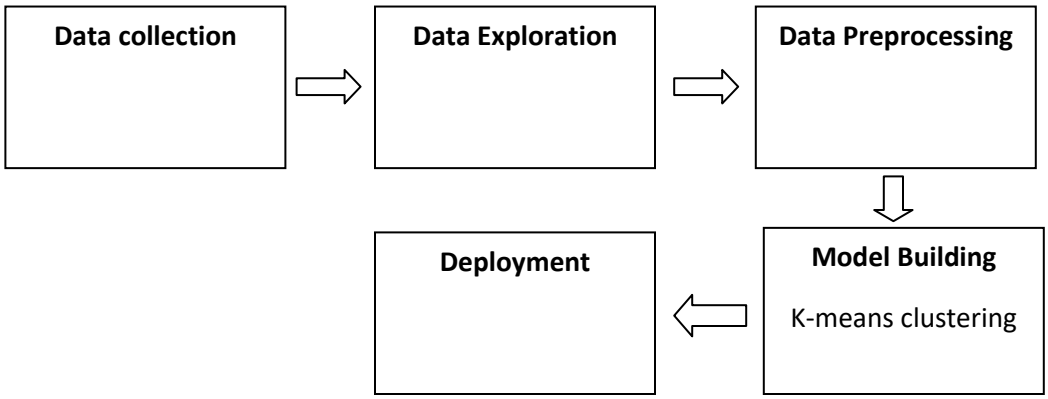
Mall Customer Segmentation

Table of contents

1. Introduction

The Low-Level-Design (LLD) gives an understanding of the internal logic design of the code for the flight fare prediction system. The LLD highlights different aspects of the architecture, the data involved and the various steps undertaken to reach the final goal of the project.

2. Architecture



3. Architecture Description

3.1 Data Description

The dataset used for the project is the Mall customer segmentation data which is in a .csv format consisting of 200 rows of data relating to different customers and 5 variables like Customer ID, age, gender, spending score and annual income

3.2 Import Data

Data is stored and imported to Python in CSV format which is then used for data pre-processing and model training and testing

3.3 Data Pre-processing

The dataset was fairly clean and had no null values or duplicated rows. The Annual income had few outliers and they were replaced with the upper limit of the box plot. For ease of clustering only 2 variables: Annual income and spending score were used and customers were grouped according to these features. The values were standardized to keep them within a range as K-means clustering requires so.

3.4 Model Building

After cleaning and standardizing the values, elbow method and silhouette score method was used to find the ideal number of clusters. The ideal number of clusters were then applied in the K-means clustering model to find groupings of clusters.

3.5 Deployment

The model will be deployed as an API using FastAPI where the user can input the relevant values to find which group the customer belongs to.