High Level Design (HLD)

Mall Customer Segmentation

Abstract

Shopping centers are always looking for ways to increase customers coming to their stores and making purchases to make profits in a huge margin. The first step to increasing profit would be to analyze the existing customers and understand their pattern of behavior and spending habits so as to target the right customers and to make their visit to the shopping centers frequent.

Machine Learning can help in this matter by grouping customers into different segments based on their spending habit. This project is done to help businesses understand their customers and apply marketing strategies and offers to the targeted customers. K-means clustering is done as a way of grouping customers into various segments

Contents

1. Introduction

The High-Level-Design (HLD) helps to add the necessary detail to the project description so that the model is suitable for coding. The document can avoid contradictions prior to coding and can be referred to as a manual

The HLD underlines the different design aspects, the user interface being used, performance requirements, design features and the architecture of the project

2. Description

2.1 Problem Perspective

The clustering problem is an unsupervised machine learning based model that will help us a group a particular customer based on their relevant details

2.2 Problem Statement

To create a solution to find groups of customers based on their spending behavior and annual income to target specific group of people and apply marketing strategies to shop more.

2.3 Proposed solution

The proposed solution is to build an unsupervised machine learning model like K-means clustering and find which cluster groups with similar features and likeliness.

2.4 Data Requirements

Data for the task depends solely on the problem statement. The dataset contains details pertaining to the customer like their shopping behavior, income, age and gender

The dataset used for the task is in CSV format and consists of numerical and categorical data types

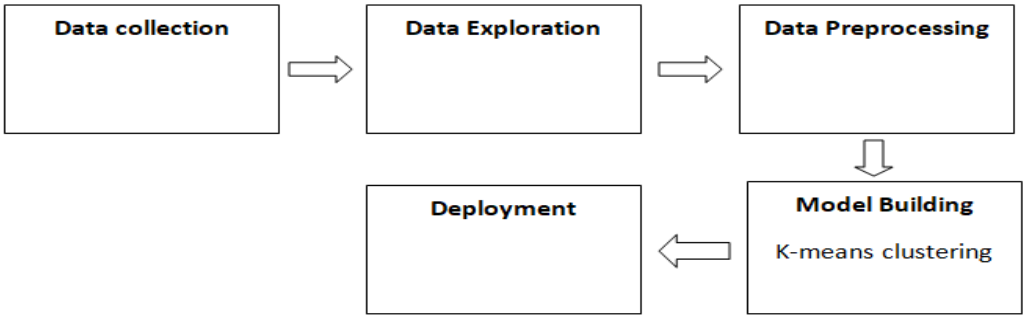
2.5 Tools used

Python is the main programming language used for the project and uses different libraries like pandas for reading the file and other tasks, numpy for mathematical operations, seaborn and

matplotlib for visualization, sklearn for building machine learning models. Jupyter notebook is used as the IDE

3. Design details

3.1 Methodology



3.2 Event log

The system logs all the steps so that the user will be aware of the processes working internally. The type of logging chosen for each depends upon the type of operation that is carried out like info, critical, debug or warning logging. Logging helps to debug issues, so it is a mandatory step

4. Performance

The unsupervised machine learning model is used to group customers with similar spending behaviour. It can help to make informed decisions and take necessary actions. Model retraining can also help to improve performance of prediction

4.1 Reusability

The code used for the project can be easily reused with no errors.

4.2 Application compatibility

The whole project will be done using Python as the main interface and each component and library in it is used to perform various tasks related to the project and ensure the proper completion of the project.

4.3 Deployment

The best machine learning model will be deployed as an API using FastAPI.

5. Conclusion

K-means clustering is the primary model used to group shopping customers based on a number of underlying factors. The goal is to demarcate groups that share common behavior.