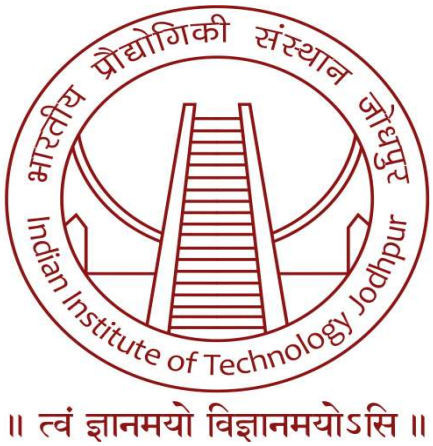
INDIAN INSTITUTE OF TECHNOLOGY JODHPUR

NH 62 Nagaur Road, Karwar, Jodhpur, Rajasthan



Trimester 2 – Post Graduate Diploma in Data Engineering

Virtualization and Cloud Computing
Design Document

Submitted By:
Aneerban Chowdhury (G23AI2059)
Vanshika Gupta (G23AI2050)
Shikha Soni (G23AI2075)

Under the Guidance of
Dr. Sumit Kalra
Associate Professor, Dept. of CSE

1. Introduction

- The primary goal of this project is to enhance customer segmentation by leveraging advanced data analytics.
- Businesses face challenges in accurately segmenting customers so that they can enhance their marketing strategies.
- In this project we have used ML model- K-Means clustering for the data classification task.
- A Flask web app presents the segmentation results.
- The solution is then deployed on Google Cloud Platform for reliability and scalability.
- We have also later established the benefits of deploying a web application over cloud platform rather than on local machines.

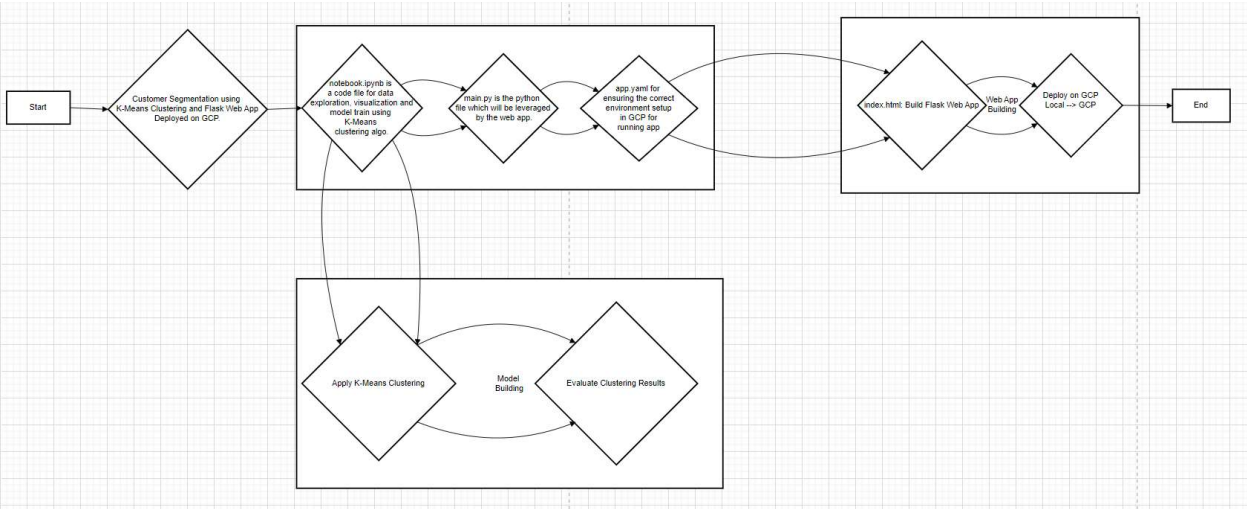
2. Background

- As the customer base continues to grow, personalized marketing becomes essential for driving engagement and increasing conversions.
- By segmenting customers into distinct clusters, the marketing team can target each group with highly tailored campaigns, leading to more effective outreach and better customer experiences.
- The K-means clustering algorithm has been selected for this project due to its proven simplicity and efficiency in handling unsupervised classification tasks, making it an ideal solution for creating actionable customer segments.
- This approach will allow for more precise targeting and optimized marketing efforts across the customer base.

3. Requirements

- **Data Ingestion**: System must accept customer data in CSV format, including features like annual income and spending score.
- **Clustering**: Apply K-means algorithm to classify customers into distinct segments.
- **API for Predictions**: Deploy an API for external applications to submit data and receive cluster predictions.
- **Model Training**: Enable periodic retraining of the model to adapt to changes in customer behavior.

4. Architecture and Design



- **Google Cloud Platform (GCP)**: The application will use GCP for deployment, leveraging Google App Engine (GAE) for hosting.

  Components:

  - ✓ **Frontend**: A simple web interface.
  - ✓ **Backend**: Python/Flask-based API providing access to the K-means clustering model.
  - ✓ **Model**: K-means clustering model implemented in Python using scikit-learn.
  - ✓ **Monitoring**: Stackdriver for logging and monitoring.

  Architecture Diagram:

  - ✓ Client → App Engine (API) → K-means Model (in App Engine) → CSV fiel (for data).

- **Google Cloud Storage**: For storing customer data as CSV files, which the model will periodically ingest for retraining.

5. Dependencies:

- **Modeling**: Python, Scikit-learn for building the K-means clustering model.
- **API Development**: Python with Flask for API services.
- **Deployment**: Google App Engine for scalable deployment.
- **Monitoring**: Stackdriver for logging and performance monitoring.