Project Design Phase-I
Solution Architecture

Date	07 November 2023
Team ID	PNT2022TMID591582
Project Name	Project - Understanding Audience: A Machine Learning Approach to Customer Segmentation
Maximum Marks	4 Marks

Solution Architecture:

- **Data Ingestion:** Data is collected from the provided sources such as Kaggle/ Amazon Datasets and its structure is analyzed.
- **Data Preprocessing:** Data is cleaned, normalized, and transformed to handle missing values and outliers. Feature engineering is performed to extract relevant customer attributes.
- **K-Means Clustering**: The preprocessed data is fed into a K-Means clustering algorithm to group customers with similar characteristics into clusters.
- **Cluster Analysis**: Analyze the resulting clusters and classify them into different groups for supervised machine learning algorithm.
- **Feature Selection:** Identify key features that contribute to the differences between customer clusters.
- **Data Classification:** Random Forest classifier, Adaboost classifier, Decision Tree classifier and XGBoost classifier are trained on the clustered data to predict customer behaviors or preferences within each cluster.
- **Model Evaluation:** Evaluate different model classification performance based on their accuracy score.
- **Saving the Model**: Consider the best solution among the four classification techniques based on the evaluation and save the model with the most considerate results.
- **User Interface:** Develop a user-friendly interface for business users to interact with the system, allowing them to explore customer segments and plan marketing strategies using Flask.

Solution Architecture Diagram: