

BANK CUSTOMER SEGMENTATION



Project Proposal

Name : Mukesh Sirvi

Roll No. : 21ME61R14

INTRODUCTION

Almost all bank marketing team are leveraging artificial intelligence or machine learning to launch a marketing ad campaign that is tailored to specific group of customers. In order the campaign to be successful they need to know their target market and take advantage of that information. The process is known as "Marketing Segmentation" and it is crucial for maximizing marketing campaign conversion rate.

Market segmentation is a marketing concept which divides the complete market set up into smaller subsets comprising of consumers with a similar taste, demand and preference.

In this project, an unsupervised machine learning will be implemented to distinguish the different customer classes.

The idea of the project and the dataset has been taken from kaggle.

PROBLEM STATEMENT

Draw insight from the data and create customer classes that will help the marketing team to define marketing strategy.

DATASET

The Dataset summarizes the usage behavior of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioral variables.

- CUSTID : Identification of Credit Card holder (Categorical)
- BALANCE : Balance amount left in their account to make purchases
- BALANCEFREQUENCY : How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)
- PURCHASES : Amount of purchases made from account
- ONEOFFPURCHASES : Maximum purchase amount done in one-go
- INSTALLMENTSPURCHASES : Amount of purchase done in installment
- CASHADVANCE : Cash in advance given by the user
- PURCHASESFREQUENCY : How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)
- ONEOFFPURCHASESFREQUENCY : How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)
- PURCHASESINSTALLMENTSFREQUENCY : How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)
- CASHADVANCEFREQUENCY : How frequently the cash in advance being paid
- CASHADVANCETRANSACTIONS : Number of Transactions made with "Cash in Advanced"
- PURCHASESTRANSACTIONS : Number of purchase transactions made
- CREDITLIMIT : Limit of Credit Card for user
- PAYMENTS : Amount of Payment done by user
- MINIMUM_PAYMENTS : Minimum amount of payments made by user
- PRCFULLPAYMENT : Percent of full payment paid by user
- TENURE : Tenure of credit card service for user

SOLUTION STATEMENT

There can be multiple ways to approach this unsupervised learning problem such as using-

❖ K- Means Clustering and Hierarchical Clustering

❖ DBSCAN and Gaussian Mixture Models

My aim will be to analyse K-Means Clustering model and determine their performance using performance metrics.

EVALUATION METRICS

INERTIA : To select best model we need to evaluate K-Means model performance but unsupervised task we don't have any targets.

But we can measure the distance between each instance and its centroid. This is the idea behind inertia metric

SILHOUETTE SCORE :

$$\text{silhouettescore} = (p - q) / \max(p, q)$$

p = mean distance to the points in the nearest cluster

q = mean intra-cluster distance to all the points

The range of **Silhouette score** is [-1, 1]. Its analysis is as follows –

- **+1 Score** – near +1 **Silhouette score** indicates that the sample is far away from its neighboring cluster.
- **0 Score** – 0 **Silhouette score** indicates that the sample is on or very close to the decision boundary separating two neighboring clusters.
- **-1 Score** – -1 **Silhouette score** indicates that the samples have been assigned to the wrong clusters.

PROJECT DESIGN AND WORKFLOW

The project will comprise broadly the following steps –

STEP 1 : PREPROCESSING

STEP2 : EXPLORATORY DATA ANALYSIS

STEP 3 : IMPLEMENTATION OF K-MEANS CLUSTERING

STEP 4 : BASED ON EVALUATION METRICS FINDING BEST NUMBER OF CLUSTER

STEP 5 : COMPARING RESULTS