Predictive Analytics: Clustering

</talentlabs>

Overview

This case study aims to give an idea of applying Machine Learning to customer segmentation. Market segmentation is the process of dividing a broad consumer or business market, usually consisting of existing and potential customers, into subsets of consumers based on common characteristics. Clustering Algorithms are used to tackle such segmentation.

Problem Statement

As a junior data analyst intern in a firm. For one of the clients XYZ marketings, your manager asked you to do a customer segmentation ML model that can, later on, be developed to give recommendations like saving plans, loans, wealth management, etc. on target customer groups of XYZ Marketings. For this, you have been provided with a dataset summarising the usage behavior of active credit card holders during the last 6 months in XYZ Marketings.

Dataset

Download the dataset

S.No	Attribute Names
1	Customer ID
2	Balance
3	Balance Frequency
4	Purchases
5	One-off Purchases
6	Installment Purchases
7	Cash Advance
8	Purchases Frequency
9	One-off Purchases Frequency
10	Purchases Installments Frequency
11	Cash Advance Frequency
12	Cash Advance TRX
13	Purchases TRX
14	Credit Limit
15	Payments
16	Minimum Payments
17	PRC Full payment
18	Tenure

The dataset summarizes the usage behaviour for 8950 active credit card holders for the past 6 months. The dataset comprises data at the customer level with 18 variables.

Expected Outcomes

Create a machine learning clustering model that group people based on their behaviours and usage data.

- Understand the given problem and explore the data set, and do the necessary data pre-processing.
- Do exploratory data analysis to understand and visualize the data.
- Work on various versions of clustering models, with parameter tuning and adjustments to select an ML model that performs best for the data at hand.
- Evaluate your results and find the best algorithm that segregates the customers. (i.e. Explain why you selected the final model, and what makes it the best one)

Note: You need to submit one/two Ipython notebooks that clearly explain the thought process behind your ML model (either in comments or markdown text), code, and relevant plots. Also, all the visualizations and plots must be done in Python(should be in the Ipython notebook).

Thank you.

All The Best :)