Capstone Project Proposal

Project Name: Create a Customer Segmentation Report for Arvato Financial Solutions

Domain Background

Arvato is a financial solutions company with solutions such as credit management and fraud management.

Arvato uses analytics to help them understand their customers better and find new customers.

Prediction of customer churn has been done in papers such as: Customer churn prediction system: a machine learning approach by Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh and Chadha Pratyush Sethi in 2021.

And it describes a study that was conducted to develop a machine learning model to predict customer churn in a telecommunications company. The authors used data from a sample of customers to train and test several machines learning algorithms, including logistic regression, decision tree, random forest, and support vector machine.

The study found that the random forest algorithm outperformed the other algorithms in terms of accuracy and F1-score, and was able to identify the most important predictors of customer churn, including monthly charges, tenure, and the presence of tech support and online security services. The authors suggest that the model could be used by the company to identify customers who are at high risk of churn and to implement targeted retention strategies.

Overall, the study demonstrates the potential of machine learning algorithms in predicting customer churn and highlights the importance of identifying and addressing the factors that contribute to churn in order to improve customer retention and business performance.

Problem Statement

Arvato aims to improve their customer service, attract new clients, and enhance their product offerings. They plan to achieve this by creating customer segments from their existing clientele. This task involves utilizing classification algorithms to classify customers from the general demographic population. By doing so, Arvato hopes to gain a better understanding of their diverse client base, which will help them expand their business and maximize revenue through targeted marketing and advertising efforts.

Datasets and Inputs

The goal of this project is to analyze demographic data of customers of a German mail-order sales company and compare it with data from the general population. Unsupervised learning techniques will be used to perform customer segmentation to identify the key customer groups for the company. AWS Sagemaker will be utilized for EDA and to create appropriate customer segments as well as a model for predicting a customer's segment based on certain characteristics. A dataset from a mail campaign will be used to create a model to predict if a customer is likely to respond to the campaign. This model will then be applied to a third data set containing demographic information for the company's marketing campaign targets to predict which individuals are most likely to become customers. The data for this project has been provided by Bertelsmann Arvato Analytics.

Solution Statement

The goal is to help the company to know who will be a good candidate to be a customer in there marketing.

Benchmark Model

For this project, the benchmark model will accurately predict customer conversion. The K-Nearest Neighbor (KNN) model will be used as the benchmark for this task.

Evaluation Metrics

the performance of unsupervised algorithms USING the silhouette method will be used to determine the optimal number of clusters and will be visualized to ensure distinctiveness. In contrast, the supervised model will be evaluated using the confusion matrix and domain knowledge will be applied to assess the model's performance.

Project Design

1- Loading dependencies
2- Loading the data
3- Processing the data
4- EDA
5- Creating supervised model
6- Clustering
7- Predictions