Machine Learning Engineer Nanodegree
Capstone Proposal

Rohan Digambar Gawade

Targeted Campaigning for client of Arvato Financial Solutions, a Bertelsmann subsidiary

Domain Background

Customer Segmentation and acquisition are the important components for any company to improve its marketing strategy and increase customer based for the product. In order to acquire new customers it is important to know about the existing customers demographics and rather than campaigning on all the general population, we can find the target audience using these existing customer demographics and have potential customers to target for campaigning. This will help our client, the mail-order company to focus and allocate its precious resources efficiently.

Problem Statement

We have two problems at hand, one is customer segmentation for existing customer and using it to identify potential customers from the people in Germany. The other is creating a model to predict the which individuals are likely to be responsive to marketing campaign of our client, the mail-order company. The end goal is to increase the efficiency of the customer acquisition process.

Dataset and Inputs

- **Metadata**

File	Description
DIAS Information Levels - Attributes 2017.xlsx	A top-level list of attributes and descriptions, organized by informational category.
DIAS Attributes - Values 2017.xlsx	A detailed mapping of data values for each feature

- **Dataset for Customer Segmentation**

File	Description	Shape
Udacity_AZDIAS_052018.csv	Demographics data for the general population of Germany	891 211 persons (rows) x 366 features (columns).

Udacity_CUSTOMERS_052018.csv	Demographics data for customers of a mail-order company	191 652 persons (rows) x 369 features (columns).

- **Dataset for Model Training and Testing**

File	Description	Shape
Udacity_MAILOUT_052018_TRAIN.csv	Demographics data for individuals who were targets of a marketing campaign	42 982 persons (rows) x 367 (columns).
Udacity_MAILOUT_052018_TEST.csv	Demographics data for individuals who were targets of a marketing campaign	42 833 persons (rows) x 366 (columns)

Solution Statement

We are going to address two things.

1. Customer Segmentation
2. Identify whether to include a person in campaign using prediction model.

Customer Segmentation

I will be using clustering to analyze the demographic relationship between the mail-order company's existing customers and general population of Germany.

Identify whether to include a person in campaign using prediction model

I will be building supervised machine learning model to predict whether we will be including a person in the campaign or not and will be evaluating the model on the test dataset and submitting the result on Kaggle.

Benchmark Model

I will be using Random Forest model for benchmarking and will compare the other models against it.

Evaluation Metrics

As there is class imbalance and also in the competition we are told to use AUC to evaluate the model performance, I will also be using AUC as evaluation metric.

Project Design

The below flow depicts the design I will be using for this project.

1. Problem Definition – Defining the problem statement for customer segmentation and prediction for mail-order company
2. Data Cleaning – Cleaning the input files by removing Null values or imputing them
3. Data Exploration – Do preliminary data exploration and visualization.
4. Customer Segmentation – Use Dimensionality reduction and clustering to segment the customer data and apply it on general population to know the relationship between existing customers and general population
5. Build Model – Using the Mailout train data build model for prediction
6. Classify Potential Customers – Using the trained model classify and predict target customers for campaigning.
7. Evaluation - Evaluate on test data and submit the results on Kaggle.