Propensity To Convert Models - Predicting the the likelihood of Customer accepting Personal Loan Offer

Introduction

Banks spend millions of dollars on marketing teams and the campaigns in targeting right customer, right channel where the response rate is higher when offer is made. Based on various studies made and the observation I have at my work, the response rate is very low, response rate is below 5% in personal loans space. The low acceptance rate means that randomly choosing customers to market to is incredibly wasteful in terms of both time and money. Thus, it is in the interest of the banks to determine which customers are the most likely to accept this type of loan based on the customer's information. Also the creatives that are being sent are looked at being more personalized so that message in the offer resonates with their situation and make the customers to accept the rate.

For my project, I am focusing only the persona/customer segmentation. Creative design and propensity rate of each creative is something would be further investigated outside of this work. For this project, I am aiming to apply the K-Means, Support Vector Machine, Logistic Regression, Perceptron, and Multi-layer Perceptron algorithms to predict whether customers will be willing to take on a personal loan. The input into our algorithms will be a customer's age, years of professional experience, annual income measured in thousands of dollars, home address, ZIP code, family size, average spending on credit cards per month measured in thousands, education level, and the value of a house mortgage (if the customer has any). Our algorithm will use this information to output a prediction on whether a customer will take a personal loan (1) or not (0).

In practice, marketing teams can optimize their spend on campaigns by targeting these customers who are likely to accept the offer.

Project Plan

I will be working on the term project individually as my work schedules are pretty tight for group project at this point. I have created high level plan with all the critical tasks and aiming to use that to track the progress and keep myself focused. For peer reviews, I will be partnering with Victoria Hall for each milestone throughout the entirety of the course. I have chosen the dataset and since I am in same domain so hoping I could leverage my domain experience to understand the problem and features better. The dataset is small and I wish if I could get better volumed dataset with at least 100K observations. I would continue to explore and merge the datasets if I found any better.

Dataset and Features

I am choosing the dataset from Kaggle as titled "Personal Loan Modeling" . It contains a list of 5000 bank customer's data resulting from a marketing campaign. For each customer, there are 12 features and a [0,1] label of whether they accepted a personal loan offered in the bank's marketing campaign. The specific features for each customer are described in detail below:

- **id** : Customer ID
- **age** : Customer's age in completed years
- **experience** : years of professional experience
- **income** : Annual income of the customer
- **zip_code** : Home Address ZIP code.
- **family** : Family size of the customer

- **ccavg** : Avg. spending on credit cards per month

- **education** : Education Level.

    - Undergrad

    - Graduate

    - Advanced/Professional

- **mortgage** : Value of house mortgage if any.

- **personal_loan** : Did this customer accept the personal loan offered in the last campaign?

- **securities_account** : Does the customer have a securities account with the bank?

- **cd_account** : Does the customer have a certificate of deposit (CD) account with the bank?

- **online** : Does the customer use internet banking facilities?

- **creditcard** : Does the customer use a credit card issued by UniversalBank?

Modeling Methodology

For this project, as of now I am aiming to apply the K-Means, Support Vector Machine, Logistic Regression, Perceptron, and Multi-layer Perceptron algorithms to predict whether customers will be willing to take on a personal loan or not. On the model evaluation, I plan to calculate the accuracy, precision, recall and F1 score as well as visualize a confusion matrix and an ROC, AUC curve. I am also interested to explore the use of confidence intervals as thats what I have been hearing at my work but I am yet to understand how to apply them to this problem.

Risks / Contingencies

Since the dataset is small and simulated dataset, it is highly possible that few models may perform poorly on the data hence as a contingency plan, I have aimed to apply multiple models and compare the results but still results may not be aligned with real world scenario but I take this

opportunity in validating my understanding with end to end flow of applying predictive analytics to a problem.

In general such datasets would be imbalanced so I am keeping closed tab on sampling (under/over) and also hyper parameter tuning because it could help the algorithms distinguish the majority class from the minority data group in our dataset.

References:

1. Datta, Anirban. "Personal Loan Modeling." Kaggle, 2020.

https://www.kaggle.com/teertha/personal-loan-modeling/activity

2. Tuscano, Hugo. "Creating a Model to Predict if a Bank Customer accepts Personal Loans".

https://toscano84.github.io/2018/11/creating-a-model-to-increase-the-personal-loans-in-banks/