CLASSIFICATION OF STAR LEADS FOR MARKETING OPTIMIZATION OF GETMONEY.COM

Amir Khoeilar

Abstract

GetMoney.com is loan service provider website that connects costumers to lenders by simplifying the process of fast loans. They provide different types of loans unsecured personal loans, payday loans, and installment loans and also business loans. The website has approached me in order to help them classify between their revenue making clients and the regular clients. With using classification models, I'm going to detect the most important features impacting the classification of star clients from regular clients in order to broaden their understanding of their clients and optimize their marketing outreach and process and decrease marketing wast.

Design

I began cleaning the data with google sheets with the provided data from the board of GetMoney.com. Narrowing down the the data into useful features while trying to get an understanding of the data and the relationships between features. Uploading the data into pandas, I started playing around with some more cleaning and also doing some EDA, plating with Seaborn and recognizing the imbalanced target data.

First, I modeled a baseline classification model with logistic regression, knowing that I would be mostly interested in F! Score and how accuracy wouldn't helpful for me do to target being imbalanced in my dataset.

Second, I split my data for into 60-20-20 in order to perform train-validation-test. I used oversampling method on my training set in order to balance my data and preform models on them. I used 4 models of "Logistic Regression","KNN","Decision Tree" and " Random Forest" and found out that LR model out performs all the other from the ROC-AUC curve. With that, I tested my model with the test data set and changed the threshold in order to receive a much better F1 score.

Lastly, using feature importance with "Logistic Regression", I graphed a bar chart for all the coefficients and chose the best features that have the most impact for website.

Data

The data provided has 6500 clients with 15 features for each client with about 5 numerical features and other being categorical. With the use of pandas, seaboard and also scikit-learn packages I calculated different metrics for each models in order to perform classifications and finally get to important features.

Algorithms

Feature Engineering

1. Oversampling in order to balance the target data

2. Scikit-learn libraries in order to perform different classification models and also feature importance

Model Evaluation and Selection

In the conclusion, with the help of the cleaned data and also pinpointing the main useful features, GetMoney.com can collect better data with respect to their clients. Also from the data collected, they have a better understanding of the demographic of their website visitors.

I have provided a recommendation to the company with help classification modeling. The recommendation to the board is to invest in their marketing in order to purchase data with leads that hold the characteristics with the recommended features from the classification model.

Tools

- Google Sheets
- Pandas
- Seaborn
- Scikit-learn

Communication

There is going to be the code for the data operations and also PDF slides of powerpoint presentation available on my GitHub account.

https://github.com/rezxkoi/getmoney.com