Santander Customer Transaction Prediction Project: find the target customers for marketing campaign

CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

Weizhong Yao, Yuanzhe Wang
Xiaowen Tang, Xiaoshuo Liu, Jane Jian

Agenda

- 01 | Key Takeaways
- 02 | Project Introduction
- 03 | Analysis Process
  - Feature Engineering
  - Model Selection
  - Cost Analysis: threshold selection
- 04 | Business Value

Key Takeaways

- Predicting which customers will make a transaction helps to better campaign
- Feature engineering techniques can improve prediction accuracy
- Based on Area under ROC Curve (AUC), LightGBM algorithm has the highest performance(AUC = 0.92282)
- The best model, LightGB algorithm, produces **15.4%** higher average customer profit than baseline model, and **200%** higher than without using any algorithm

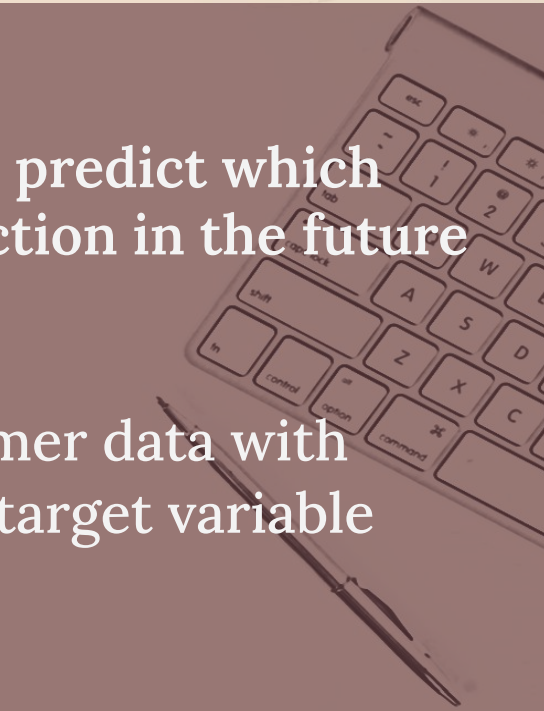
Project Overview

Goal

Discover the best approach to predict which customers will make a transaction in the future

Data

De-identified historical customer data with 200 predictor variables and 1 target variable



Analysis Process

Transforming the raw data into features that better represent the underlying problem to the predictive models

Feature Engineering

Designing a predictive model that has the best performance

Model Selection

Converting a predicted probability into a class label

Threshold Selection

Analysis Process

Transforming the raw data into features that better represent the underlying problem to the predictive models

Feature Engineering

Designing a predictive model that has the best performance

Model Selection

Converting a predicted probability into a class label

Threshold Selection

Feature Engineering helps prediction to perform better

Fake Test Data

- Synthetic data from real testing data
- Intentionally disturbs final prediction performance.
- Found 100,000 fake records

Magic Features

- Unique value that only appears once among a feature
- Provides more useful information to the prediction model.
- Added 200 magic features

Data Augmentation

- Resampling data
- Helps the training algorithm to learn about the data better.
- Added new information to the data set

LightGBM:

- Raw data: AUC = 0.8979
- After feature engineering : AUC = 0.9228

Compared to raw data, model built with data after feature engineering has better performance, with AUC increased by 2.5%

Analysis Process

Transforming the raw data into features that better represent the underlying problem to the predictive models

Designing a predictive model that has the best performance

Converting a predicted probability into a class label

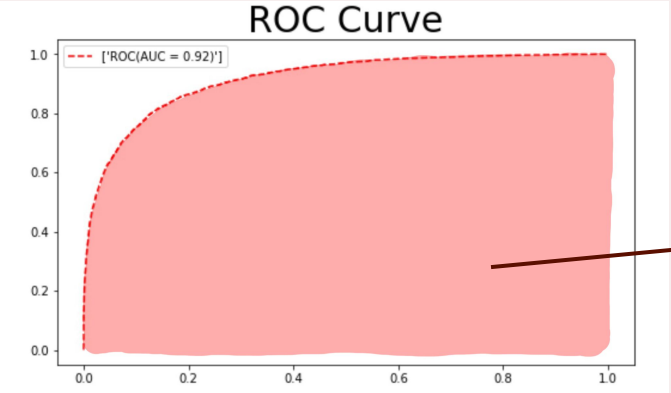
Feature Engineering

Model Selection

Threshold Selection

Model Selection: Introducing AUC

AUC - Area Under Curve



AUC

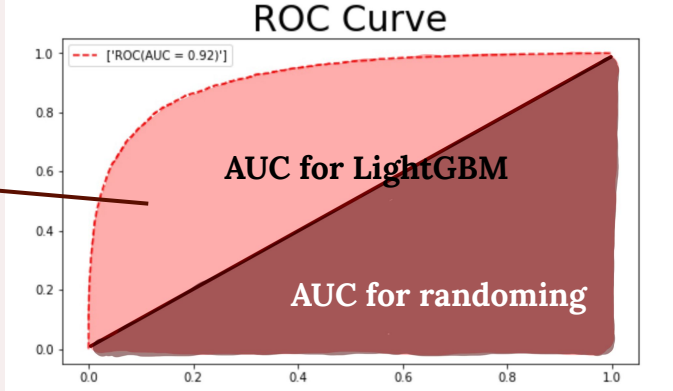
AUC represents the probability that a model can distinguish between a positive and negative outcome when given a random record from the dataset. The higher the AUC, the better.

Model Selection: LightGBM has the best performance

Best Performance

Model	AUC
LightGBM	0.92282
XGBoost	0.91745
Neural Network	0.88149
Naive Bayes	0.888

Value that LightGBM brings in compared to randoming



Analysis Process

Transforming the raw data into features that better represent the underlying problem to the predictive models

Designing a predictive model that has the best performance

Converting a predicted probability into a class label

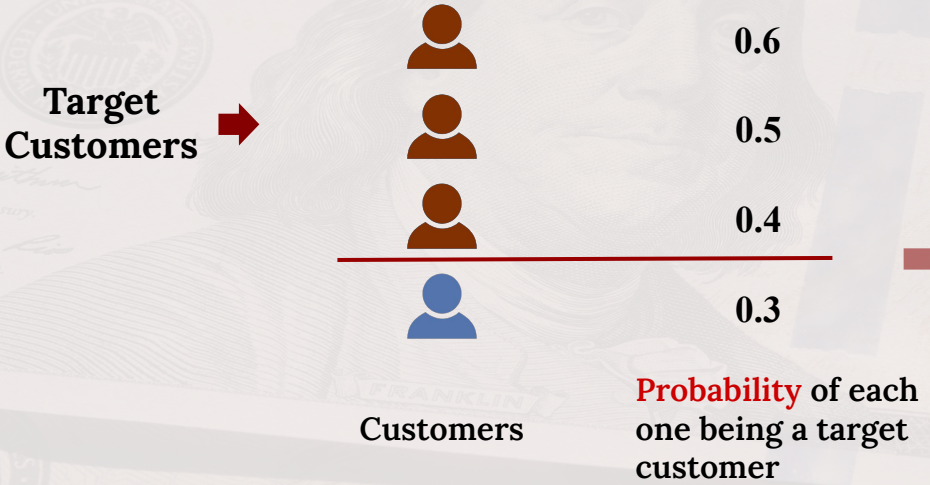
Feature Engineering

Model Selection

Threshold Selection

Output of Prediction Model and Threshold Selection

Example of output of our model:



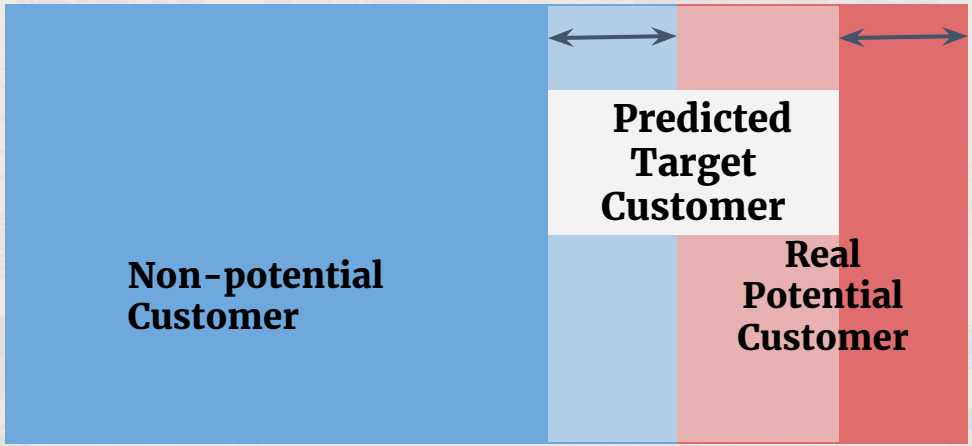
Threshold: lowest probability of a customer to define as a target customer

Threshold Selection - Minimize misclassification cost

How to choose the optimal threshold?

Threshold Selection - Minimize misclassification cost

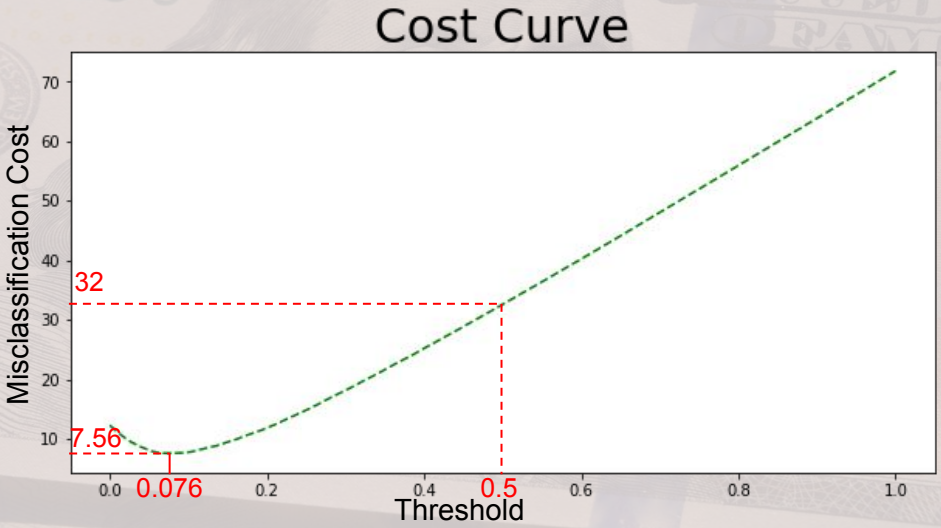
Sunk Cost of wrongly targeting a non-potential customer + Opportunity Cost of missing a potential customer = **Misclassification Cost**



Goal:
Cover more total potential customers and avoid covering non-potential customers to **minimize the misclassification cost**

Threshold Selection - Minimize misclassification cost

Select optimal threshold



Optimal threshold is 7.6%, the corresponding average misclassification cost is 7.56, compared with 32 using the naive threshold of 50%.

Business Value

Business Value: our model leads to better marketing campaign strategy

$$\text{Revenue} = \#\text{ of Real Potential Customers} \times \text{Revenue Per Person}^{[1]}$$

$$\text{Cost} = \#\text{ of Target Customers} \times \text{Cost Per Person}^{[1]}$$

$$\text{Avg Profit Per Customer} = \frac{\text{Revenue - Cost}}{\#\text{ of Target Customer}}$$

[1] What is the average customer acquisition cost that a bank pays to acquire a credit card customer?, Parvathy P. and 2 others
https://askwonder.com/research/average-customer-acquisition-cost-bank-pays-acquire-credit-card-customer-i-m-fy512dtei

Business Value: our model leads to better marketing campaign strategy

Average profit per target customer

Method	Average profit per target customer ($)
Randomly Select Target Customer	-59.51
Using Baseline Model	53.90
Our Optimized Model	62.19

Conclusion

Feature Engineering , Model Selection, Threshold Selection

- Choose LightGBM as the algorithm to help Santander to predict which customers will make a transaction in the future
- Achieve AUC of 0.92282
- 7.6% threshold

Business Value

- Achieve an average profit of $62.19 per customer