Direct Marketing Optimisation

Fang Zhou, Ph.D.

Overview

- Business Context and Objective
- Data Understanding and Preparation
- EDA and Visualization
- Model Development and Evaluation
- Model Prediction
- Target Strategy Optimisation

Business Context and Objective

In financial industry, institutions frequently use direct marketing campaigns to promote their products, for example, consumer loans, credit cards, and mutual funds. However, such campaigns are constrained by operational limitations, such as the number of clients that can be contacted and the capacity to tailor offers to individual needs. Effective targeting strategy ensures that the right clients receive the most relevant product offers, resulting in higher conversion rates, improved client satisfaction, and increased revenue.

In our particular use case, the key objective is to

- Predict the client's likelihood of purchase for a consumer loan, credit card, and mutual fund, respectively.
- Optimise targeting strategy and allocation of marketing offers to maximise revenue.

Data Understanding and Preparation

- The given excel file provides multiple datasets, including binary indicators of product sales status and revenue from the sales, as well as several potential factors, covering social-demographic information, product holdings and account balances, aggregated financial transaction data for each client.

Data Sheet	Description
Soc_Dem	Social demographic information, like sex, age, bank tenure
Products_ActBalance	Product holdings and account balances - Count_*: count of specific product types - ActBal_*: account balance for the corresponding product types
Inflow_Outflow	Aggregated financial transaction data, like volume and transaction statistics
Sales_Revenues	- Binary indicators of product sales status (1=sold, 0=not sold) for consumer loan, credit card, and mutual fund - Revenue generated from the sales

Data Understanding and Preparation

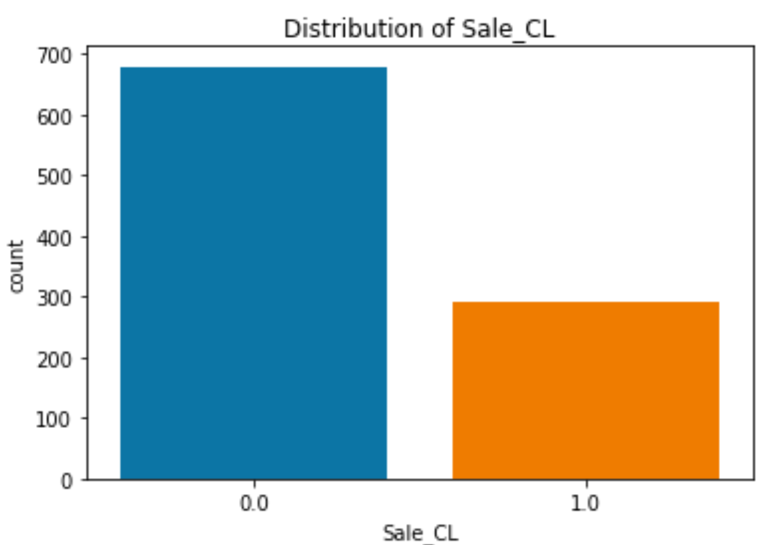
The merged dataset contains 1615 entries and 36 columns. Here are some key observations.

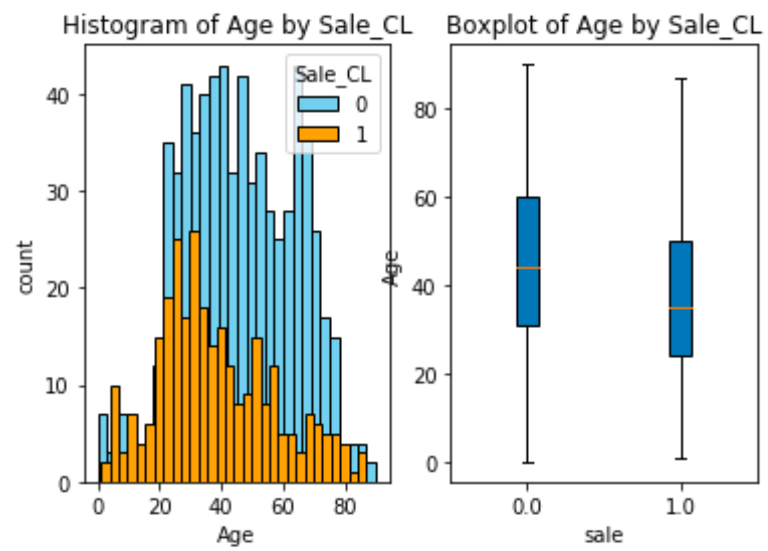
- The dataset has a mix of integer, float, and categorical values.
- Some columns named as Count_*, ActBal_*, Volume_*, and Transaction_* contain missing values. The categorical column of Sex contains a few missing values.
- The binary indicators of product sales status and revenue information for clients, contain 60% non-missing values (served as a model development set) and 40% missing values (served as a model prediction set).
- The columns, like Count_OVD and Count_CC, have a standard deviation of zero initially. After filling missing values with zero, the standard deviation becomes non-zero.

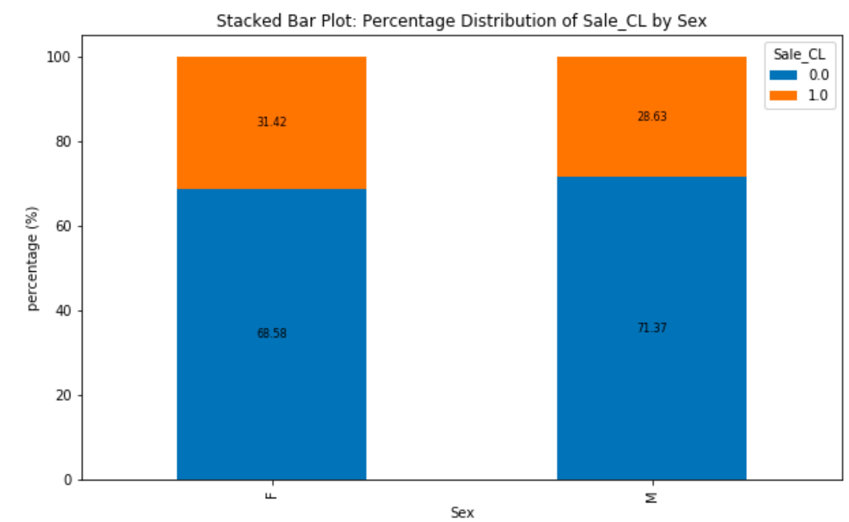
Data preprocessing (e.g., reformat data types, handle missing values) is conducted to prepare and clean the merged data. Then, the processed dataset is split into the model development set (969 entries) and the model prediction set (646 entries) based on the presence of product sales indicator. (See details in jupyter notebook)

EDA and Visualization

EDA and Visualization is conducted on the model development set to understand the data distribution and patterns/relationships between the product sales indicator and other potential factors. Below are some sample graphs.







From those graphs, we could get a rough idea about which factors are more likely to contribute to product purchase by clients. For example, the above graphs show that clients with a younger age and a sex of female, are more likely to purchase a consumer loan.

Model Development and Evaluation

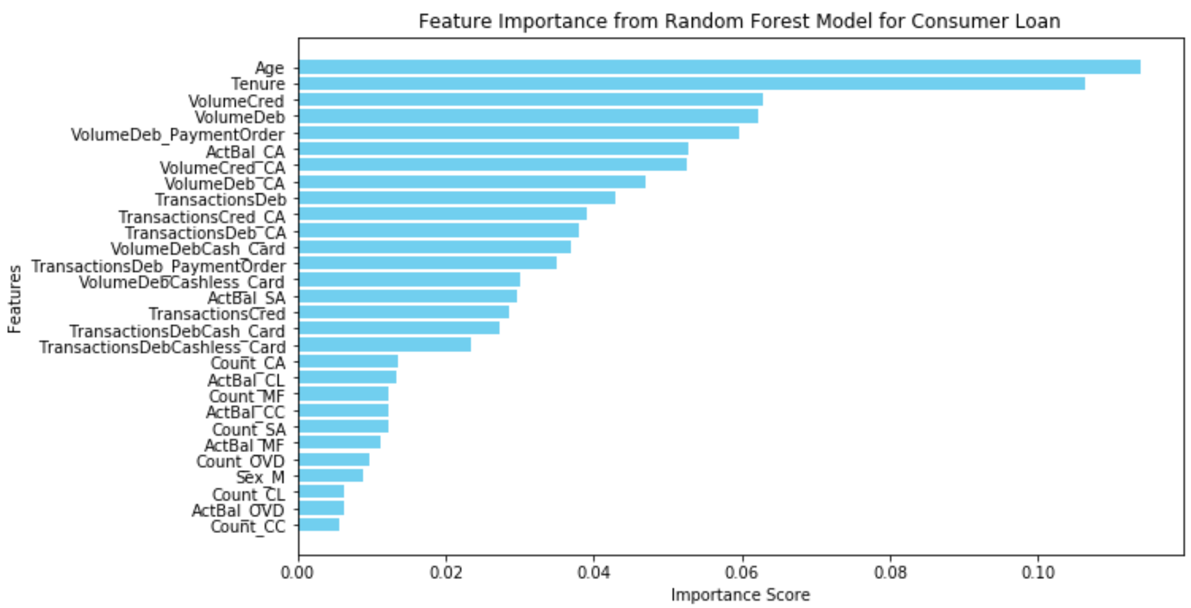
To be more quantitative and scientific, machine learning model is used to further detect the patterns. Before building the model, we need to preprocess the data, including but not limited to feature scaling and categorical variable encoding. Then, we build three machine learning models for binary classification using algorithm like random forest to estimate the likelihood of purchase for a consumer loan, credit card, mutual fund, respectively.

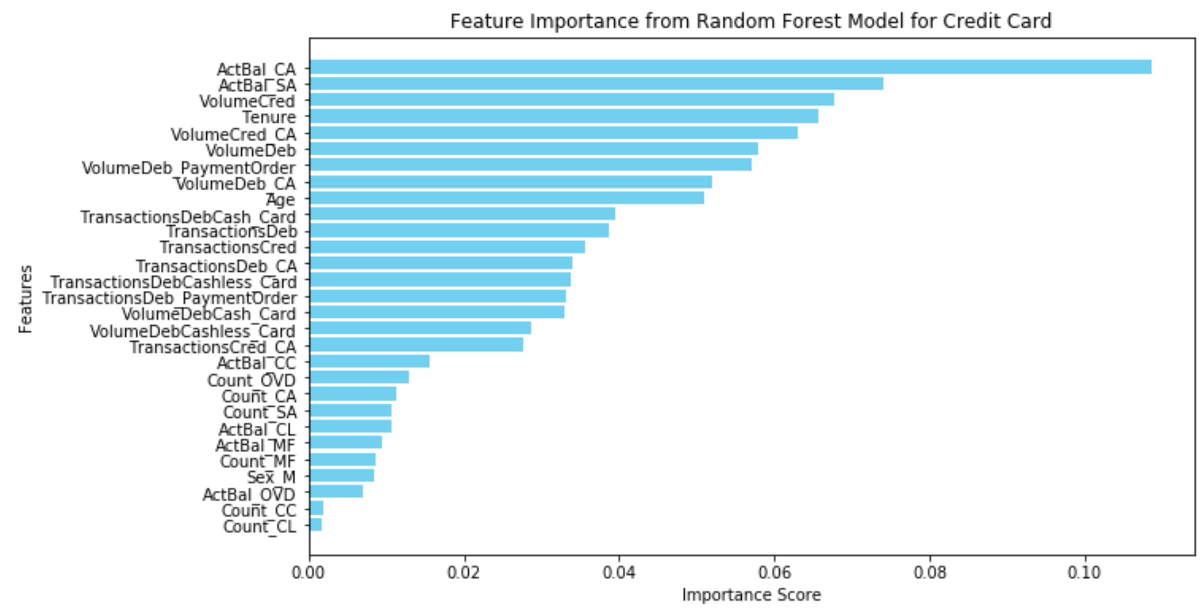
Model Evaluation

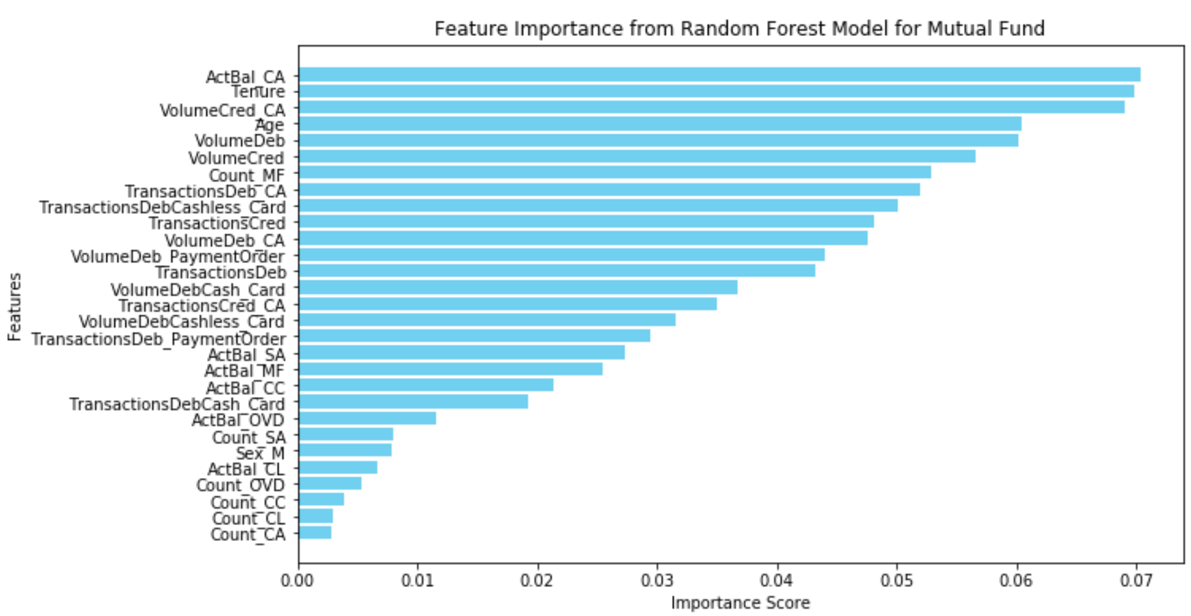
Model	AUC	Accuracy	Precision	Recall	F1-Score
rf_cl	0.55	0.71	0.60	0.55	0.55
rf_cc	0.47	0.68	0.46	0.47	0.46
rf_mf	0.51	0.73	0.54	0.51	0.46

Model Development and Evaluation

Feature Importance







The top 5 most important factors contributing to the purchase of consumer loan, credit card and mutual fund, respectively are:

- **CL:** Age, Tenure, VolumeCred, VolumeDeb, VolumeDeb_PaymentOrder
- **CC:** ActBal_CA, ActBal_SA, VolumeCred, Tenure, VolumeCred_CA
- **MF:** ActBal_CA, Tenture, VolumeCred_CA, Age, VolumeDeb

Model Prediction

After building the three models on the model development set, we will utilise them to predict the likelihood of purchase for each product on the model prediction set, respectively. From those prediction probabilities, we could further answer two questions:

- Which product is more likely to be purchased by each client across the three products?
  - Obtain the maximum probability of purchase for a product and the corresponding top product for each client.
- Which clients have a higher propensity of purchase for a consumer loan, credit card and mutual fund, respectively? (Under a constraint of contact limitation)
  - Rank and select the top 15% of clients based on the probability of purchase for a consumer loan, credit card and mutual fund, respectively.

Model Prediction

- Most likely to be purchased product across the three products for each client. **(top_product_for_each_client.csv)**

Client	Max_Prob	Top_Product
909	0.4	Prob_CL
699	0.3	Prob_CL
528	0.5	Prob_CL
1145	0.5	Prob_CC
517	0.3	Prob_CL
1475	0.2	Prob_CL
175	0.5	Prob_CL
1134	0.5	Prob_CL
1218	0.7	Prob_MF
7	0.5	Prob_CL

- Top 10 clients who have a higher propensity of purchase for a consumer loan, credit card and mutual fund, respectively. **(top_client_prob_cl.csv, top_client_prob_cc.csv, top_client_prob_mf.csv)**

Client	Prob_CL	Client	Prob_CC	Client	Prob_MF
674	0.9	19	0.9	1480	1
240	0.8	382	0.9	1095	0.9
126	0.8	1410	0.9	1093	0.8
342	0.8	701	0.9	1007	0.8
595	0.8	1280	0.9	1218	0.7
350	0.8	1129	0.9	109	0.7
490	0.8	727	0.9	1289	0.7
498	0.8	401	0.9	211	0.7
1458	0.7	1331	0.8	389	0.7
471	0.7	332	0.8	769	0.7

Target Strategy Optimisation

We aim to maximise revenue through optimising targeting strategy and allocation of product marketing offers, under the constraints of contact limitation and single offer per client, to finally answer the follow two questions:

- Which clients are to be targeted with which offer?
    - We calculate the expected revenues by multiplying the probability of purchase for each product with the average revenue for the corresponding product sale.
    - By ranking expected revenues generated from each product sale for a client, we can obtain the maximum expected revenue and the corresponding product offer for that client.
    - Under the constraint that only 15% of clients can be targetted, we select the top 15% of clients who have a higher maximum expected revenue for targeting with their corresponding product offer.
- What would be the expected revenue based on your strategy?
    - By summing up the maximum expected revenues of the top 15% of clients, we obtain the expected revenue based on our strategy.

Target Strategy Optimisation

- List of sample clients and their optimal targeting offers to maximise expected revenue. **(top_client_revenue_offer.csv)**

Client	Prob_CL	Prob_CC	Prob_MF	Exp_Revenue_CC	Exp_Revenue_CL	Exp_Revenue_MF	Max_Exp_Revenue	Target_Offer
674	0.9	0.1	0.3	0.27128265516733	3.24379809818664	0.577504643962849	3.24379809818664	Exp_Revenue_CL
240	0.8	0.2	0.2	0.54256531033466	2.88337608727701	0.385003095975232	2.88337608727701	Exp_Revenue_CL
126	0.8	0	0.2	0	2.88337608727701	0.385003095975232	2.88337608727701	Exp_Revenue_CL
342	0.8	0.3	0.3	0.81384796550199	2.88337608727701	0.577504643962849	2.88337608727701	Exp_Revenue_CL
595	0.8	0.6	0.2	1.62769593100398	2.88337608727701	0.385003095975232	2.88337608727701	Exp_Revenue_CL
350	0.8	0.4	0.2	1.08513062066932	2.88337608727701	0.385003095975232	2.88337608727701	Exp_Revenue_CL
490	0.8	0.5	0.2	1.35641327583665	2.88337608727701	0.385003095975232	2.88337608727701	Exp_Revenue_CL
498	0.8	0.2	0.1	0.54256531033466	2.88337608727701	0.192501547987616	2.88337608727701	Exp_Revenue_CL
1458	0.7	0.3	0.3	0.81384796550199	2.52295407636739	0.577504643962849	2.52295407636739	Exp_Revenue_CL
471	0.7	0.1	0.1	0.27128265516733	2.52295407636739	0.192501547987616	2.52295407636739	Exp_Revenue_CL

- Under the constraints of contact limitation and single offer per client, the total maximum expected revenue based on our target strategy is **217.57**. **(total_expected_revenue_on_strategy.csv)**