Project: Re-tune credit card churn model

Business problem

CCPL want to use credit card churn model to **pick up most potential churn** customers to do **SMS marketing.** Need to review the model after 2 years implementation

Solution

EDA **re-tune** the model by:

- Updating **newest data**
- Adding **more features** regarding spending status, hard-rock (non-spending) , request to close (rtc) status, etc.

Result

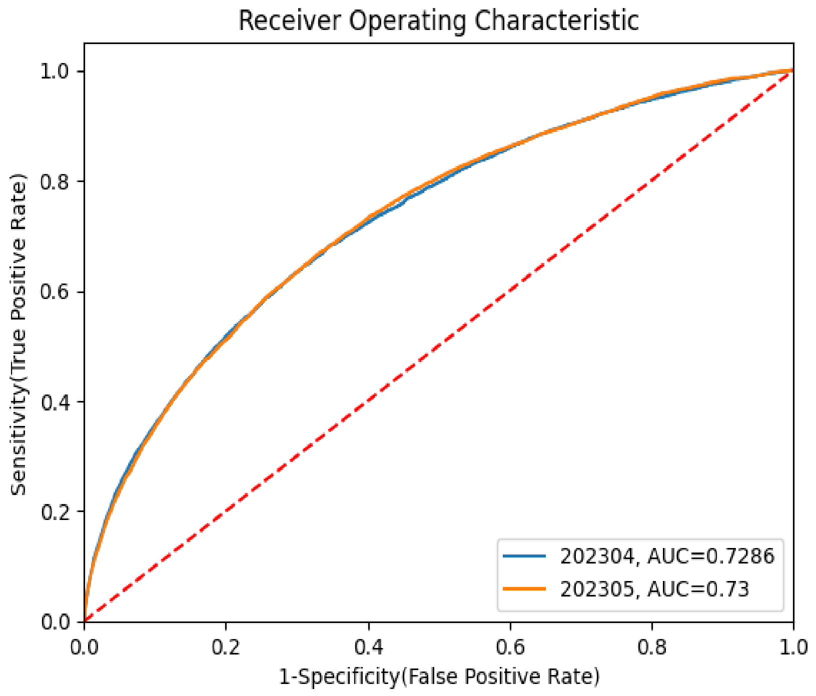
Top lift of new model (re-tuned) **increase 10%** compared to the old model. New model is implemented since 4/2023. **AUC** of new model is **consistent** at about **73%** in recent 3 months

Key outputs & insights

Top lift comparison

Model	Potential_churn_group	Nov.2022	Dec.2022	Jan.2023
Current Model	10	2.7	2.8	2.9
New Model	10	2.9	3.2	3.2

AUC of new model since implemented



Project: Impact measurement of model churn

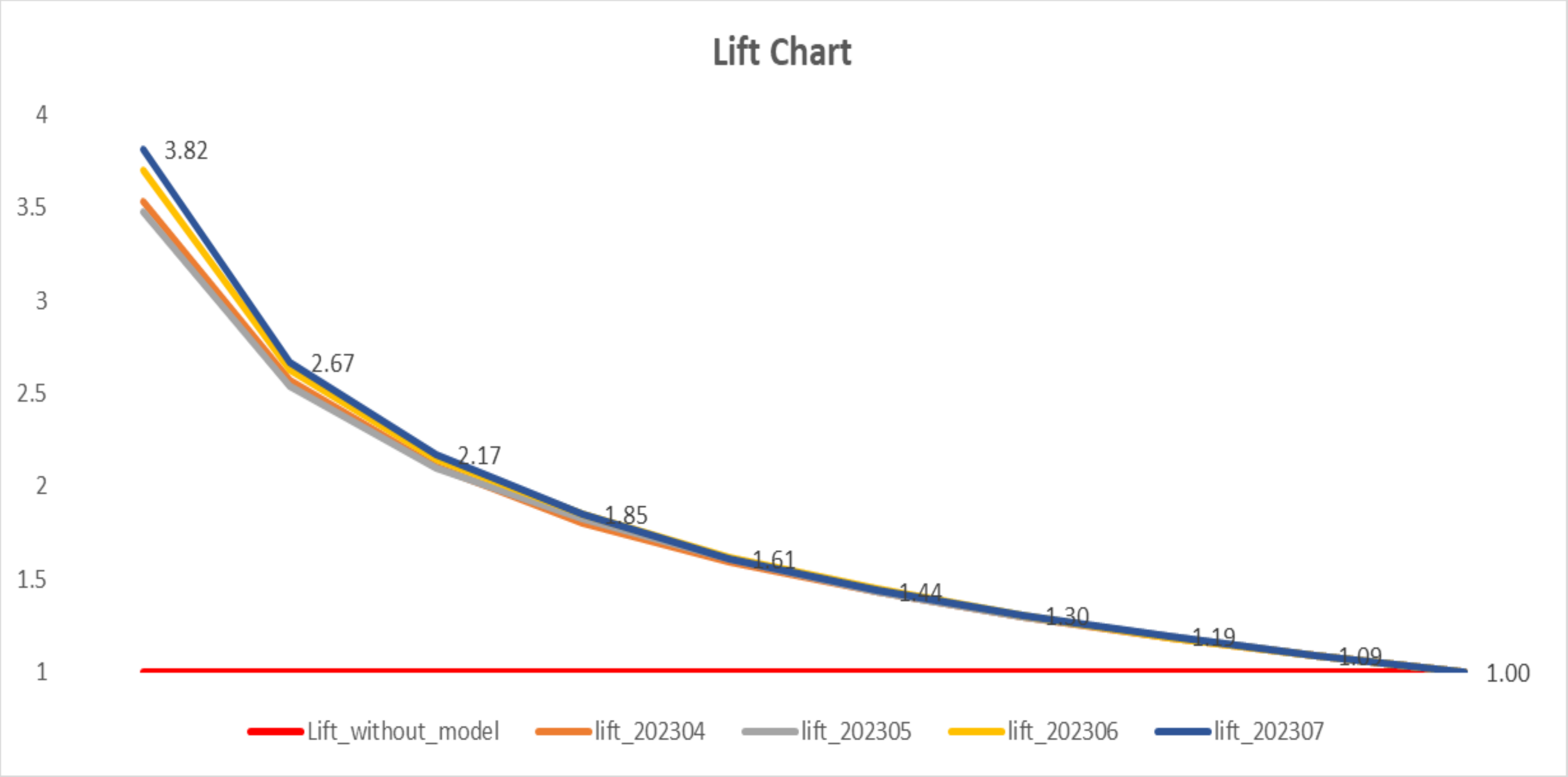
Business problem

Currently, CCPL use churn model by select highest potential churn group (9,10) , combined with multiple conditions like customers decrease spending, customers in request to close or "hard-rock" list. However, they find out that the rate of customer spending is still higher than other data groups. They concern that these group is not right target customers that potential to churn.

Solution

EDA **validate model** in recent months, access current implementation as well as propose a more reasonable implementation.

Key outputs & insights



Key insights

- Model still has moderate performance with top lift around **3.5 to 3.8**
- Group of customers decrease in spending **30-50%** should not be targeted since it's not a strong signal . Customers might decrease in spending but it's just because they spend much higher in the past , not because they stop spending and churn
- EDA suggests way to implementation: select top **highest churn score customers** , number of customers should be targeted base on marketing budget

Project: Credit Card user TOI Analysis

Business problem

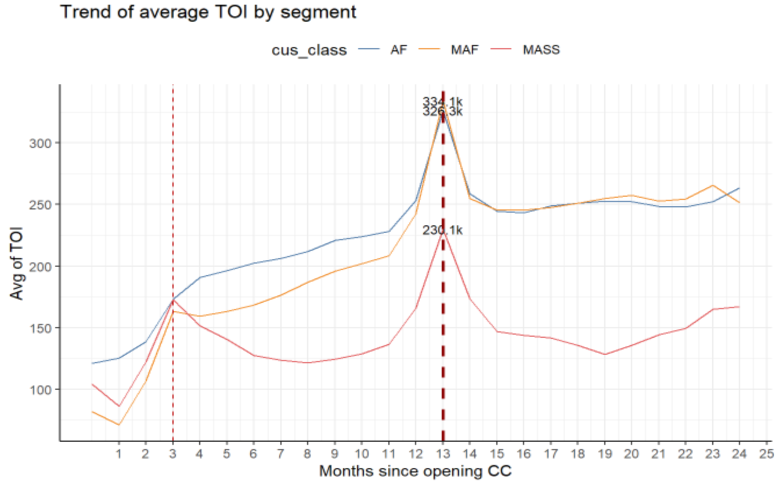
CCPL wanted to **analyze TOI of credit card customers** in order to have appropriate actions to **increase TOI** of credit card customers

Solution

EDA used **Exploratory Data Analysis** focus on different aspects:

- **How long from card opening customer can** reach the highest TOI?
- **Correlation** between **spending and TOI**
- **TOI** & **Cost** of current portfolio

Key outputs & insights



Average TOI balance in the last 2 years of customer (thousand VND)

	cus_class	no_card	min	q25	q50	q75	q90	q95	max	mean	mean_trimed
1	AF	124,898	-32,139	-2.6	34.4	342.2	968	1,630	747,428	312	156.7
2	MAF	280,122	-22,667	-4.5	41.4	347.7	770	1,121	74,924	239	151.0
3	MASS	54,807	-22,320	-33.4	5.6	281.4	726	1,180	43,670	209	107.4

Key insights

- **Time** for the cardholder to **reach** the **highest TOI**: **MOB 3** & **MOB 13** (both NFI & NII reach the highest level at *month 3 and month 13 after card opening*)
- **Higher spending, higher TOI** generated
- Cards bring highest proportion of positive TOI: **Titanium Lady, Step Up ~ 95%,** highest proportion of negative TOI: **Super Shopee ~ 36%**
- Mean monthly TOI AF & MAF **~ 150k**, MASS **107k**

Profile of request to close CC customers

Business problem

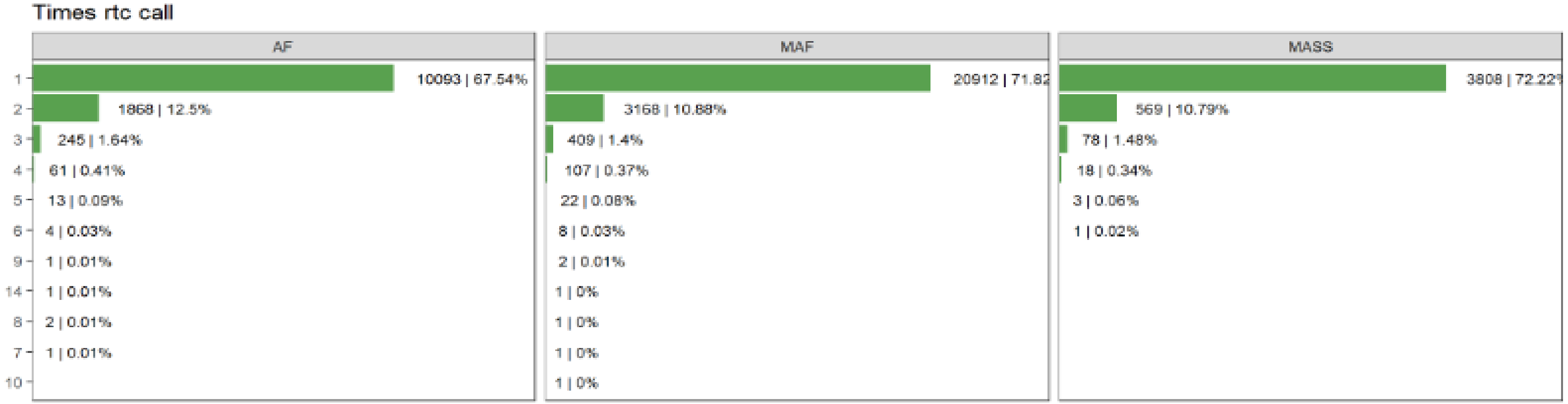
CCPL wanted to **analyze profile of request to close** customers to have suitable action to retain

Solution

EDA used **Exploratory Data Analysis** focus on different aspects:

- Reason for **request to close card**
- **Times and number of call for** request to close
- **Percent of offer** after request to close

Key outputs & insights



Offer Customer Receive	8/31/2023	9/30/2023	10/31/2023
CASHBACK OR VOUCHER	3.86%	3.55%	3.99%
CHURN	26.44%	32.31%	23.39%
EXCLUDE	9.51%	9.75%	9.13%
LOP or Free Reissue Card	1.09%	1.18%	0.73%
MPTN	24.72%	22.20%	28.53%
NO-OFFER	34.39%	31.01%	34.23%

Key insights

- **50%** customers don't share reason to close their card. Top second reason is "There are no POS / card user locations nearby, cash is preferred "
- **~80%** of customers call to close their cards **1 to 2 times**. In average, after 7 months from opening the card, the customer will request to close the card for the first time, then 3 months later continue to call to close the card a second time.
- Type of card that has highest percentage of request to close call : **Super Shopee (30%) and Titanium Lady (18%)**
- In recent months, besides successfully retain customers with no-offer **~30%,** customers receive free annually fee offer ~25%

Project: Recommend suitable offer for customers who request to close credit card

Business problem

Currently, all customers that request to close card will be offered a **'free next annual fee',** but it is **not effective as expected** approach, so CCPL wanted to give suitable offer (annual fee/cashback) to each customer group for whom call 247 to request to close their credit card

Solution

Based on customers request to close data, EDA use **exploratory data analysis** to analyze behaviors of these customers. Features used in analysis including: **MOB**, **product holding**, **revolver status**, **card type**, **RTC history**, **cc spending**, **TOI**

Key outputs & insights

TOOL CHECK OFFER

Please key in Contract_number

Contract_number	324-P-073xxx

contract_number	acnt_contract_id	hard_rock	campaign_group	open_date	ovd_days	offer_1st	offer_2nd
324-P-073xxx		1	casa	15/2/2023	0	No-offfer	Cashback

Key insights

1. Customers have following 4 of 7 **characteristics can retain without any offer:**

- **MOB:** < 9
- **Product Holding :** AF (BANCA, UPL, OD, Investment); MAF **(**IDC, Banca, UPL,OD); MASS (IDC,BANCA, UPL)
- **Revolver** customers
- **Card:** World, World Lady, Shopee
- Used to call **RTC**
- **CC Spending**: AF> 4.8 mil, MAF > 3mil, MASS >1 mil
- **TOI:** >150k

2. Customers that offered **cashback:**

- **Spending** AF>= 3.6 mil, MAF >=3 mil
- **TOI:** AF >= 168k, MAF >= 160k, MASS > 130k

Project: Analytics to push CC spending

Business problem

CCPL want to develop some strategies with **objective** is to **push spending** of CC customers

Solution

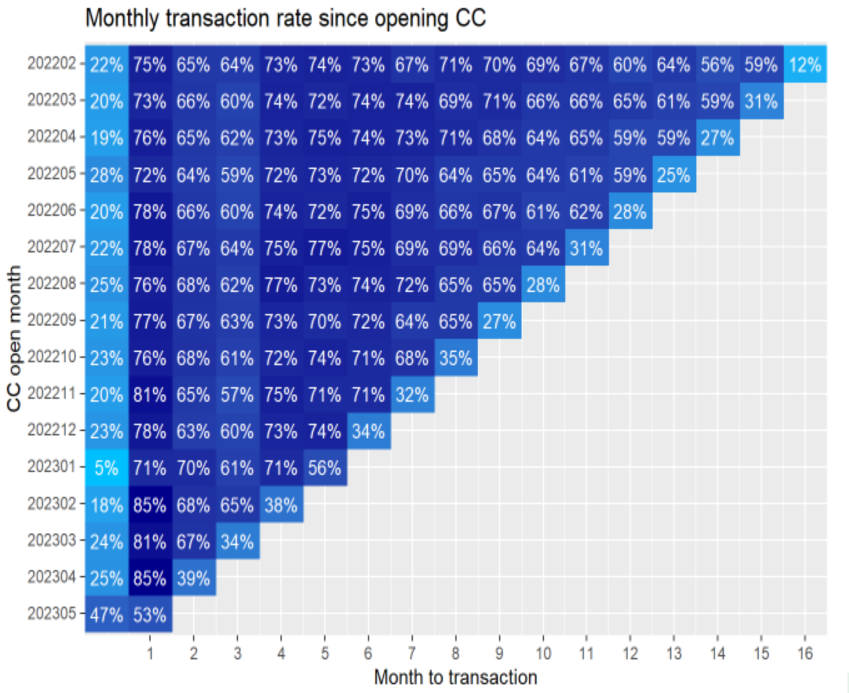
EDA used **Exploratory Data Analysis** to analyze data, focus **on time** that customers spend on CC the most:

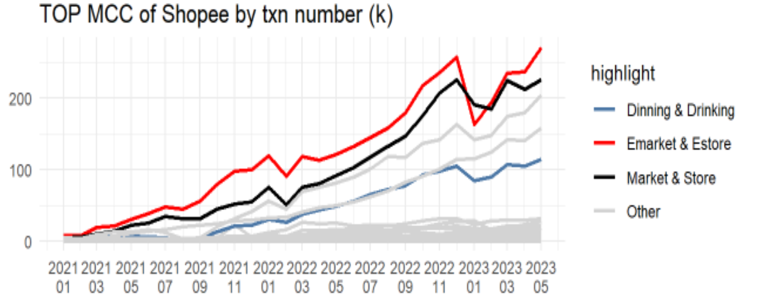
(1)Highest Spending **Day in Month**

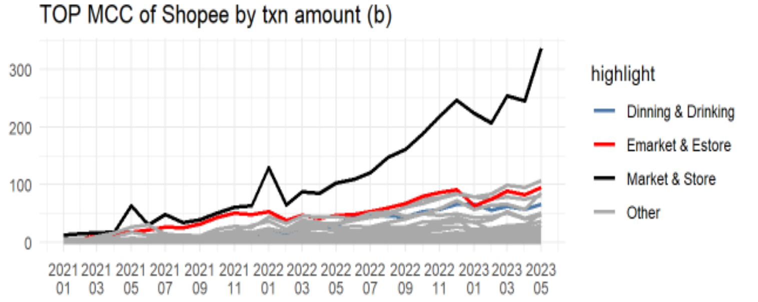
(2)Highest spending day after on book **(MOB)**

(3)Highest spending **Month in Year**

Key outputs & insights







Key insights

- Customers spend more after **monthly statement date** and **payroll date**
- The **first** and **fourth** month are the months with the highest percentage of customers using cards to spend, 70% spend from the first month
- Customer spend more **on holiday (4 quarter) and Tet season**
- **Market & Store** is the MCC that spent the most in terms of transaction volume and value, at all months of the year, also in most popular card types: Shopee, Lady, World, Stepup
- The top 3 MCCs used in each card type are also **different from** the MCC that gets the most cashback

Key recommendations

- Develop features on apps **to summary** most spending MCC => **Cross Sell** on Lady and World card (5% cashback on Market Store)
- **Flexible Cashback Option** based on customer demand
- **Customize cashback amount** based on transaction history

Project: Comparison of CC spending group

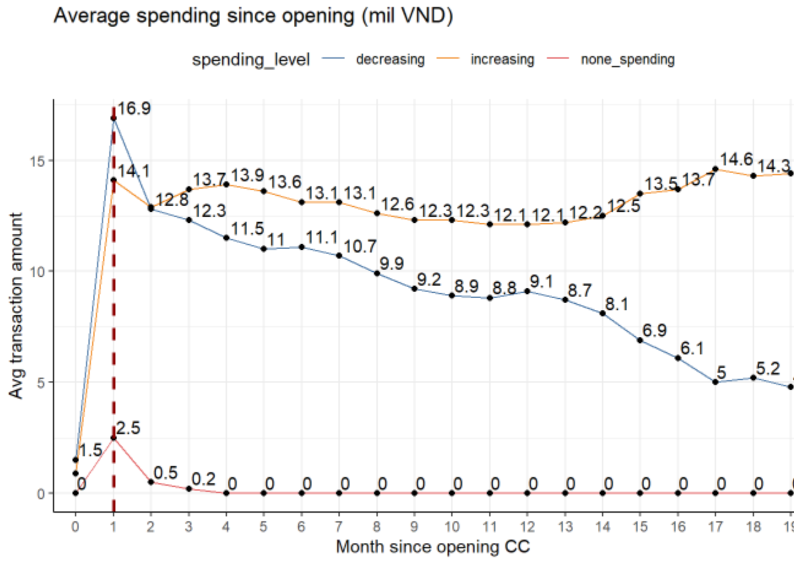
Business problem

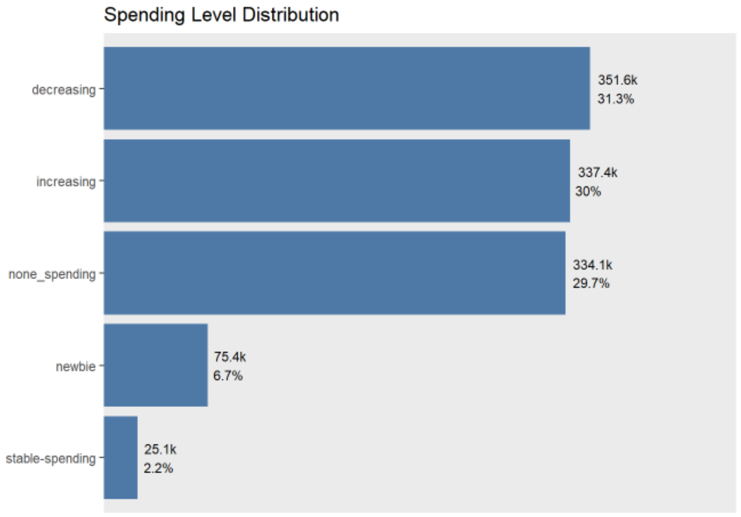
CCPL want to analyze behavior of decrease/ increase and none spending group to have appropriate actions

Solution

EDA used **Exploratory Data Analysis** to analyze data, focus **on time** that customers spend on CC the most:
(1) Comparison of
(2) Customer journey **of three grou3 spending groups:** time to spend after onboarding each group

Key outputs & insights





Key insights

- Notably, in the none spending group **(~31%),** there are 100k customers who have not made any transaction since opening card, accounting for ~31% of the non-spending group, mostly own MC2/ No1 card
- Only take cards opening after January 2022 into account , in 38k customers who have already churned, 50% of them stopping spending for **2 consecutive months.**
- The current none spending group has ~14% spending on MCC group **Other Retail** (Advertising Services / Business Services / Direct Marketing Remaining/ Vehicles) slightly higher than the other 2 groups (1.3 times more).
- Customers in all 3 groups spend most in **first month** after open their card, the group that reduces and stops spending sharply reduces spending from the **second month after opening the card**, while the group that increases spending increases evenly and continuously after this time point.

Project: IDC Card Closed Prediction Model

Business problem

Predict IDC cards **closed in next 3 months**

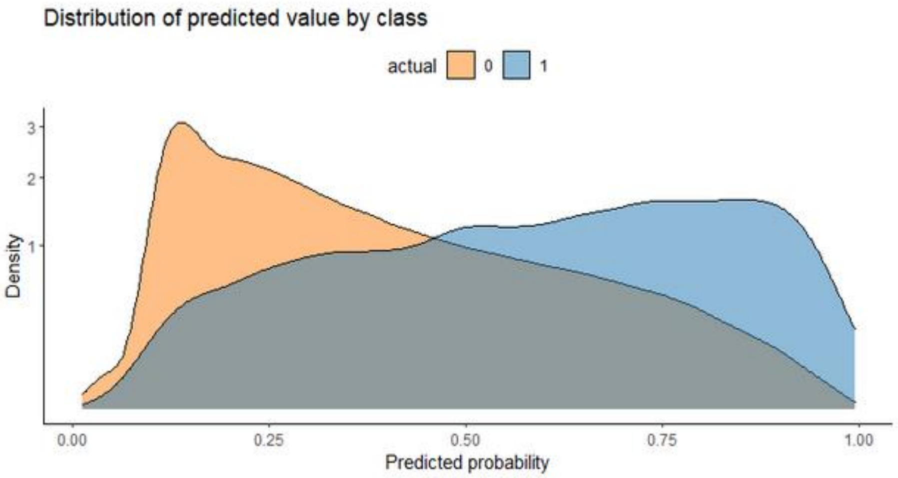
Solution

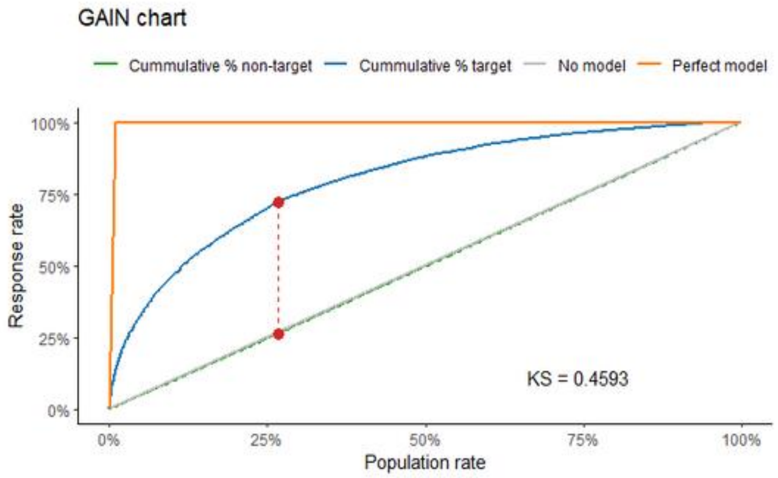
- **Train** model on active IDC cards (202211), card level
- **Validate** model on active IDC cards (202205, 202208, 202210)
- Use **undersampling & stratified sampling (1 % churn)**

➢ Model chose: **StackedEnsemble**

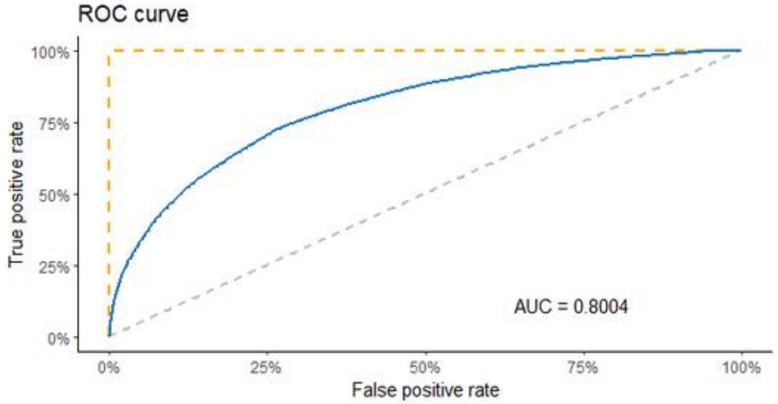
- Number of Base Models: 5
- Base Models (count by algorithm type): drf gbm glm (1 / 3 / 1)

Key outputs & insights





Decile	Total	Positives	Response rate	Cummulative positives	Cummulative response rate	Lift
1	123664	5116	4.14%	46%	4.14%	4.6
2	123664	1865	1.51%	63%	2.82%	3.2
3	123664	1287	1.04%	75%	2.23%	2.5
4	123664	800	0.65%	82%	1.83%	2.1
5	123664	645	0.52%	88%	1.57%	1.8
6	123664	462	0.37%	92%	1.37%	1.5
7	123664	321	0.26%	95%	1.21%	1.4
8	123664	238	0.19%	97%	1.08%	1.2
9	123665	186	0.15%	99%	0.98%	1.1
10	123665	115	0.09%	100%	0.89%	1.0



Key takeaways:

- Better Density Curve if train model **on card level.**
- Chose suitable timeframe to predict model base on flag predictor (in this case next 3 months).

Project: COF Customers Prediction Model

Business problem

Predict COF (card on file) users in next 1 month

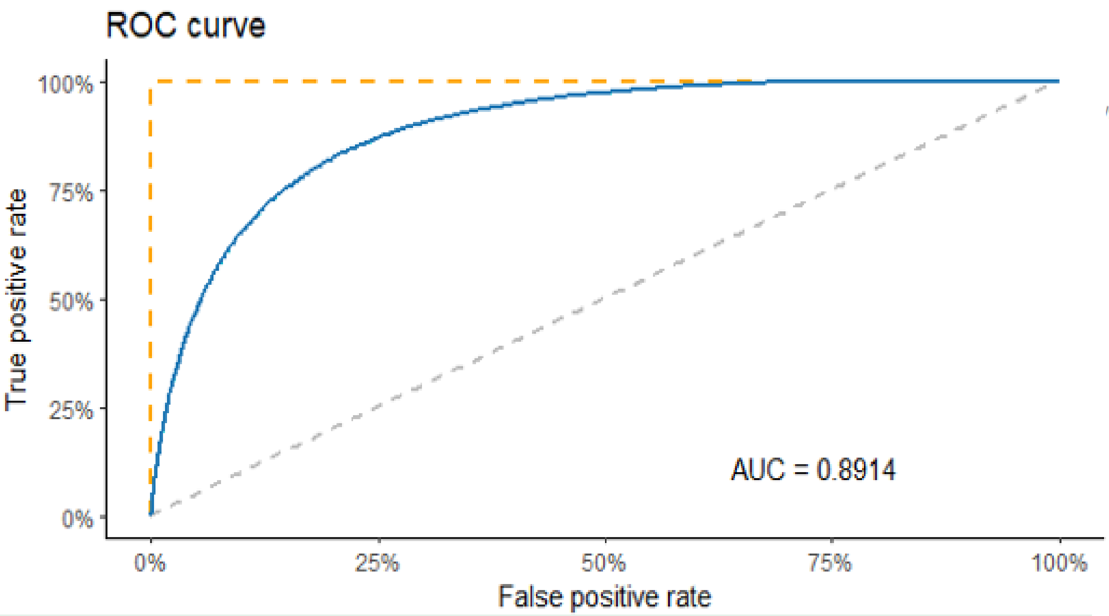
Solution

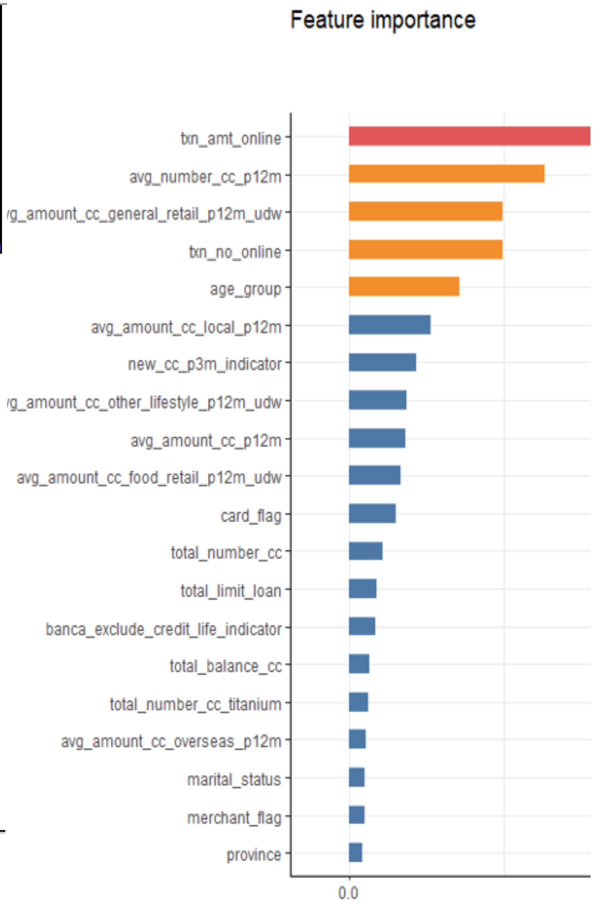
- **Train** model on customers **have never use COF** on any online payment method before (202203+202206+202209+202212), %COF**: 4%**
- **Validate** model customer 202304, 202304

➢ Model chose: **Xgboost**

Key outputs & insights

	Train				Validation	
	20220331	20220630	20220930	20221231	20230331	20230430
COF	23,1	25,287	30,137	27,411	7108	8,368
NON-COF	607,9	624,968	661,846	732,593	1,165,392	1,177,280
% COF	3.70%	3.90%	4.40%	3.60%	0.60%	0.70%
				AUC	0.91	0.9





Key insights:

- Most **importance features** include total transaction amount online, number of cc , total transaction numbers , merchant flag, age group of customer,.. those variables also show significant impacts to COF flag via **exploratory data analysis insights.**

Project: Predict CC spending next month for each customers

Business problem

CCPL want to push spending of customers by using cashback/ voucher, so they need model to predict how much money customers will spend next month to estimate how much cashback they need to spend for those customers

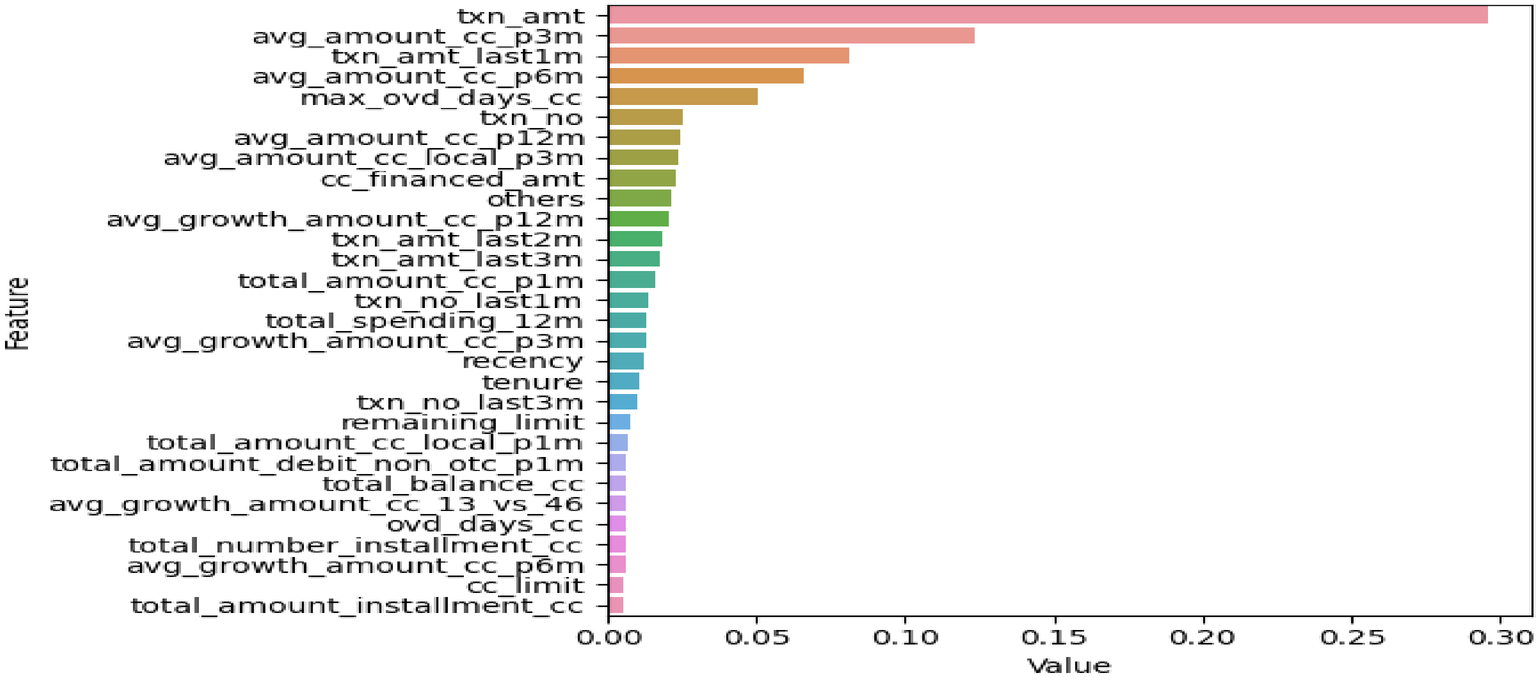
Solution

EDA **use regression model** using 50 most important features related to creditcard, demographic, transaction, deposit.
The output includes:
-Identify increases/decreases/remains unchanged spending group , compared to the previous month
-Predict how much money customers will spend next month

Results

Model is piloted and implemented on data of December 2023. In this pilot, 19k customers who was predicted to decrease in spending is chosen , combined with result of next MCC model to get most suitable offer to push spending

Top important features of model



Project: Predict CC spending next month for all portfolio

Business problem

CCPL want to do budgeting and planning as well as response to seasonal and trend (if any) so they ask to predict next month spending of portfolio

Solution

EDA **use multiple algorithm to predict timeseries**

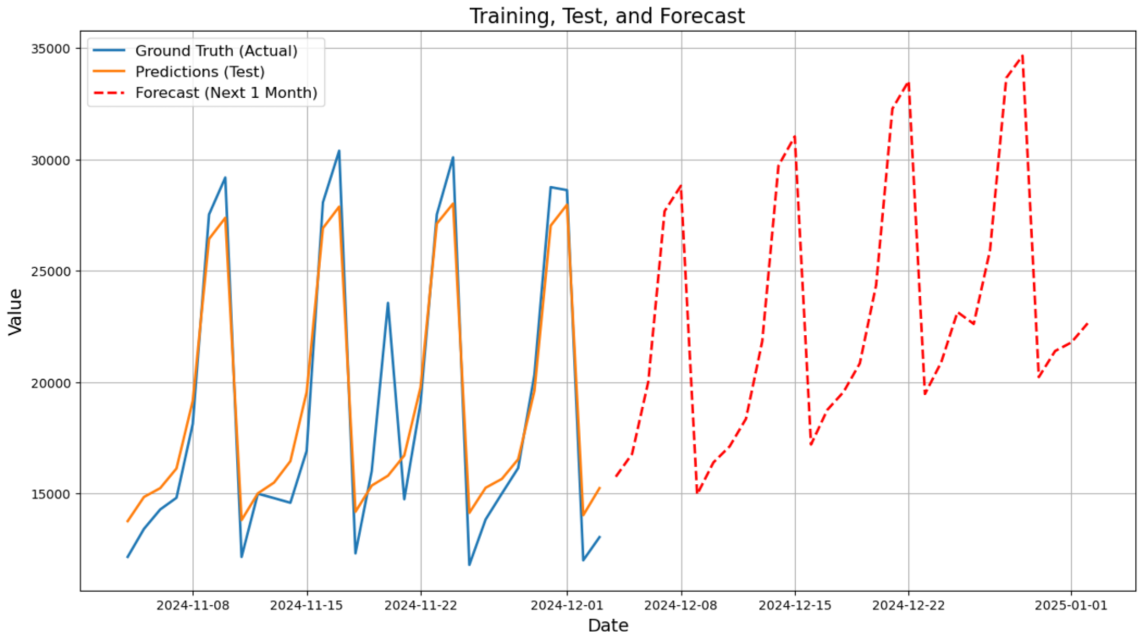
- Prophet
- Supervised learning using regression model
- SARIMA, ARIMA….

And choose prophet since it return lowest MAPE

Results

Model is piloted and implemented on data of December 2024

```
# Create and configure the Prophet model
m = Prophet(
    yearly_seasonality=True,
    weekly_seasonality=True,
    holidays=holidays_daily_vietnam,
    seasonality_mode="multiplicative",  # This should be fine now
    **best_params  # Ensure best_params does not include seasonality_mode
)
```



Project: MCC recommendation for credit card user

Business problem

CC want to analyze behavior of customers used CC and recommend next MCC that they can probably use base on past purchase behavior

Solution

EDA used **recommendation model to predict next MCC :**

- **Item based**
- **User based**
- **Matrix Factorization**

Key outputs & insights

precision@K: trong số top K MCC gợi ý cho 1 KH thì có bao nhiêu MCC là đúng

recall@K: trong số n MCC KH thực sự chi tiêu thì mình gợi ý được bao nhiêu MCC

$$precision@k = \frac{\text{number of recommended relevant items among top k}}{\text{number of recommended items k}}$$

$$recall@k = \frac{\text{number of recommended relevant items among top k}}{\text{number of all relevant items in the system}}$$

MRR: lấy giá trị từ 0 (worst) đến 1 (best), dùng để đánh giá chất lượng các gợi ý mà model đề xuất. Chỉ số này quan tâm đến việc MCC đúng đầu tiên xuất hiện thứ bao nhiêu trong top n MCC model đề xuất

Hit ratio: có bao nhiêu KH được dự đoán đúng ít nhất 1 MCC

Item based collaborative filtering result

Key insights

Chỉ số	**Giá trị**	**Meaning**
Hit Ratio	0.461	46.15% of users received **at least one correct recommendation**
Precision@K	0.426	42.6% of the top-K recommendations were relevant
Recall@K	0.504	50.4% of all relevant items were successfully recommended

Project: IDC spending on advertising service Analytics

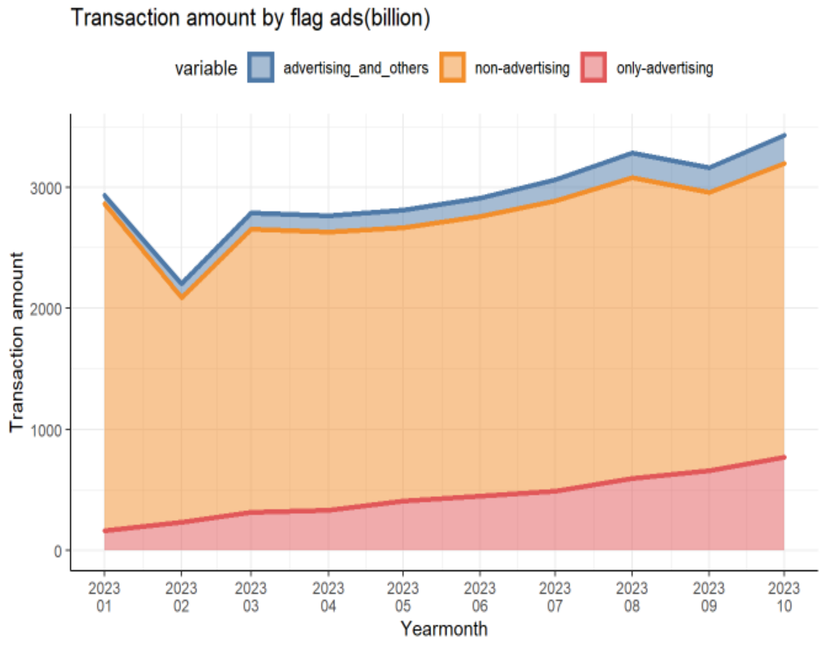
Business problem

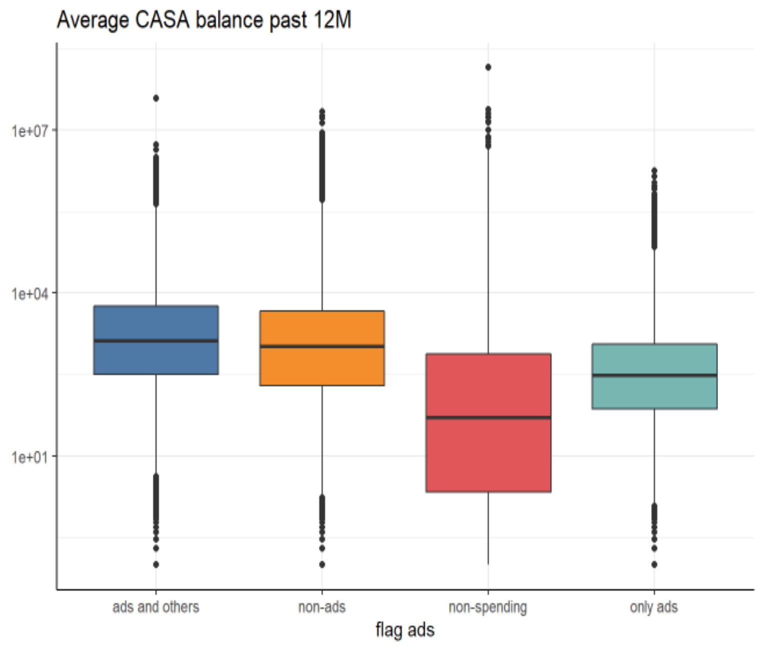
IDC want to analyze behavior of customers use IDC for advertising services, regarding's spending level compared to the entire IDC portfolio , they also want to calculate proportion of customers who have suspicious behaviors that highly potential to cheat

Solution

EDA used **exploratory data analysis** method to analyze spending behaviors of customers on advertising services, combined with hypothesis from business and data insights to detect customers group who highly potential to cheat

Key outputs & insights





Key insights

- **Profile** of customers having suspicious behaviors:
  - Spending on **advertising services**
  - Number of IDC **card closed>=3**
  - Number **of card open >=10**
  - Number of different **IDC card type :>=4**
  - Number of **Card Ok not unlock >=2**
  - Number of card open recently (**MOB<=3M) >=4**
  - Has at least 1 card open and close **within 1 month**
  - Mean casa one month <=300k
  - Mean spending on advertising > **125 mil monthly, 68 transaction monthly**
  - In **chargeback list**
  - Trans amount cancel/ Total amount >10%

Project: Model to predict cheating customers who use Facebook advertising

Business problem

IDC want to identify and promptly prevent customers with suspicious behavior like request to refund continuously their spending amount on Facebook advertising

Solution

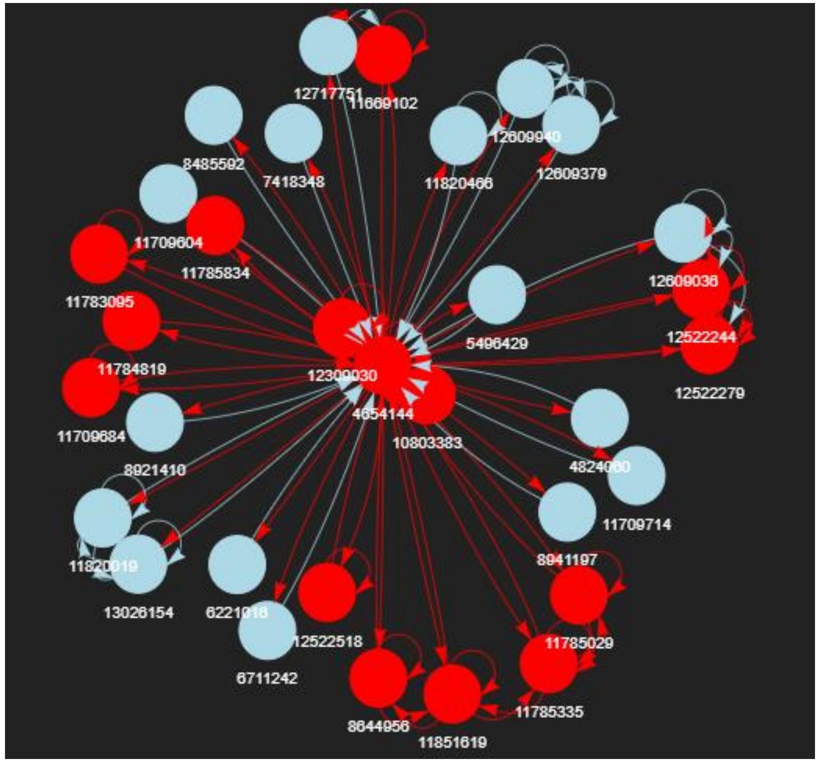
EDA **develop** the model by:

- Using graph analytic to identify relationship between cheating customers
- Features used related to : graph features, suspicious behaviors, spending on IDC , demographics…

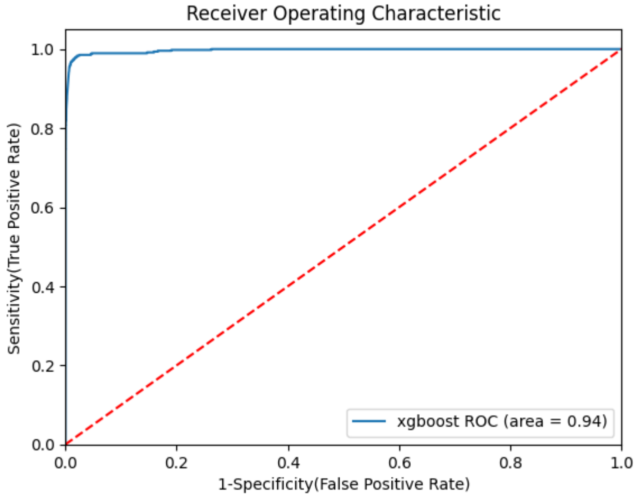
Result

EDA **develop xgboost model** to predict cheating customers and validate on test sample of 9,10/2023

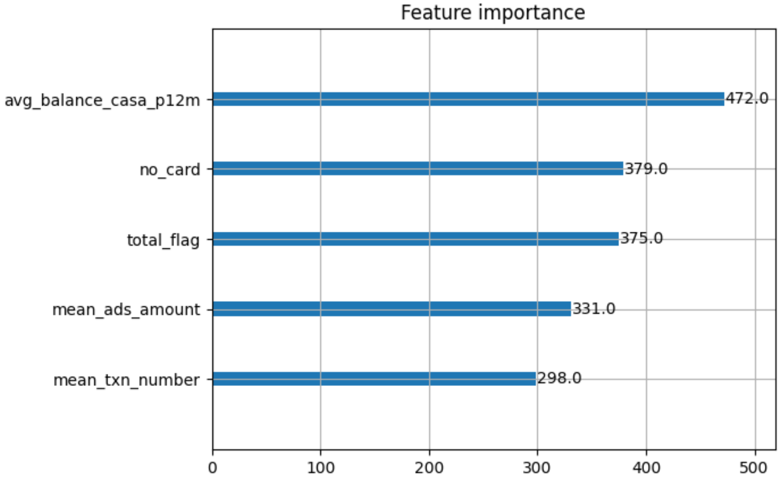
Key outputs & insights



Customers have NEO transaction to each others. **Red**: cheating, **blue**: not cheating



AUC: 0.94



Top features importance include casa, number of card, total flag of suspicious, mean advertising amount and transaction number