

# Credit Card Approval

Dương Minh Hiếu - DA18



# Credit Card Approval

---

01

## **INTRODUCTION THE DATASET**

Describe the dataset, the purpose of the model

02

## **DATA PREPARATION**

Explore, analyze and scale data

03

## **MODEL BUILDING**

Build and select the best model for prediction

04

## **PREDICT & EVALUATE**

Predict and evaluate the performance of the model

## Datasets problem

- **Điểm tín dụng** là phương pháp sử dụng thông tin và dữ liệu cá nhân của người đăng ký thẻ tín dụng để dự đoán khả năng trả nợ trong tương lai. Ngân hàng có thể quyết định có cấp thẻ tín dụng cho người nộp đơn.
- **Mục đích:** cần xây dựng mô hình học máy để dự đoán xem KH **“tốt”, “bình thường”, “xấu”**

## Datasets Description

### Application record

- + Gồm các thông tin cơ bản của KH như applicant gender, DOB, education type, assets that applicant had, etc.
- + Có 18 biến: 12 biến phân loại, 5 biến liên tục và 1 biến ID

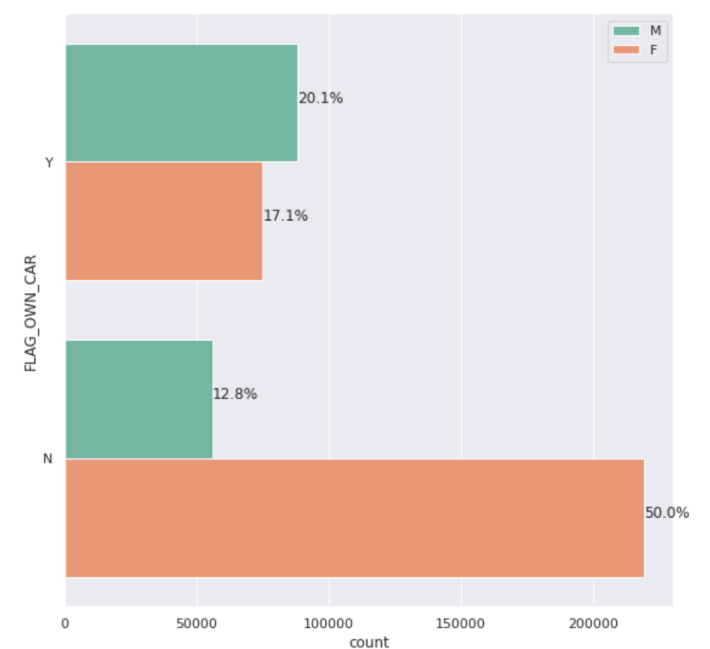
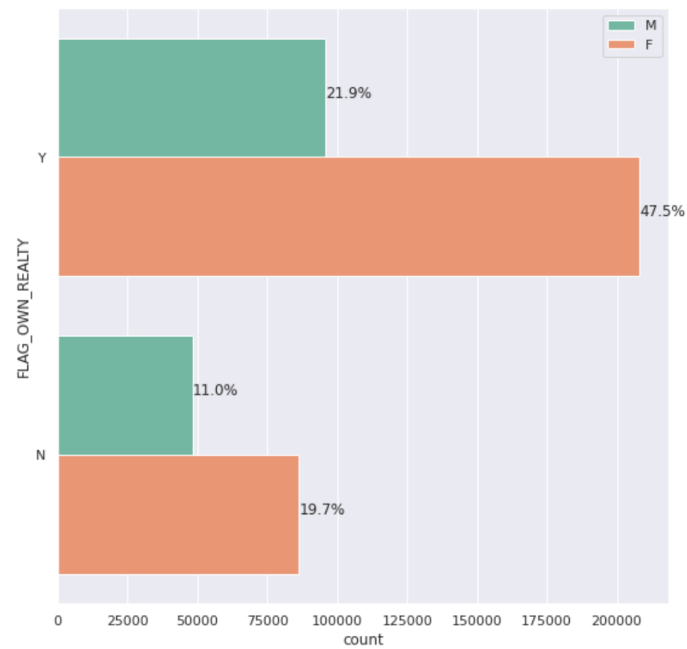
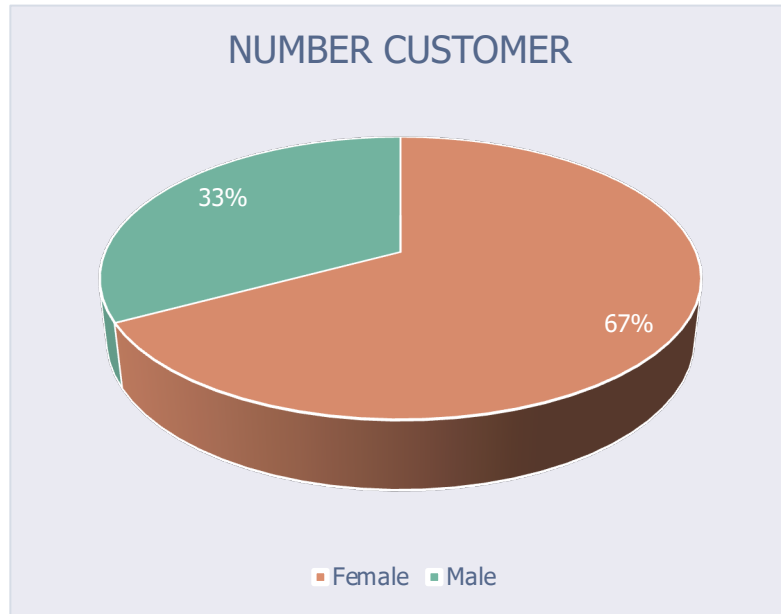
### Credit record

- + Gồm các thông tin thanh toán khoản vay của KH
- + 1 biến phân loại, 1 biến liên tục và 1 biến ID

## Action

1. Từ 2 datasets, áp dụng vào **build models**:
  - Logistic Regression
  - Naive Bayes
  - SVM
  - Decision Tree
  - Random Forest
  - KNN
  - Boosting: AdaBoost, Gradient Boosting, XGBoost
2. So sánh các chỉ số Accuracy, F1 score, Precision, Recall để **chọn mô hình** phù hợp (ưu tiên phân loại KH “xấu” chính xác nhất)

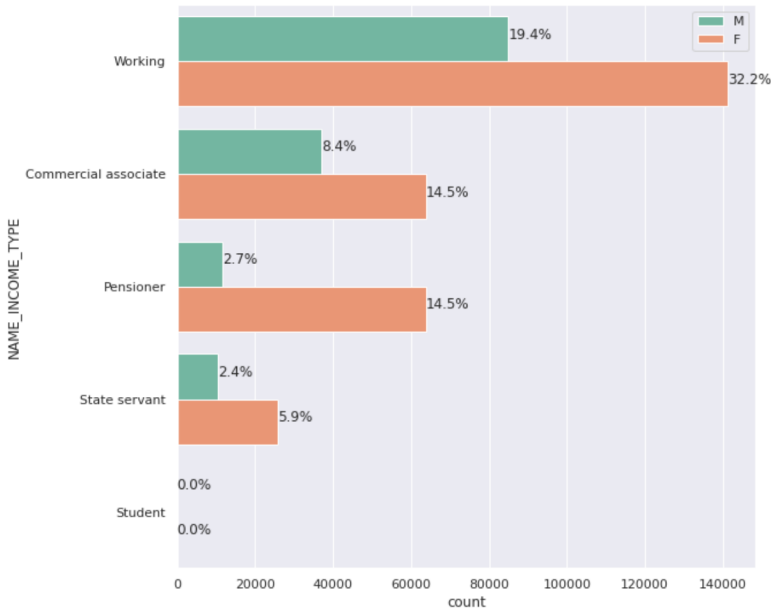
# Graphic



## Insights from application details record:

- Đa số KH là nữ, số lượng KH nữ gấp đôi số lượng KH là nam
- KH có sở hữu BDS, KH nữ chiếm đa số, 47.5% gấp hơn 2 lần KH nam 21.9%
- KH sở hữu ô tô, KH nam và nữ xấp xỉ nhau, 20.1% KH nam và 17.1% KH nữ

# Graphic

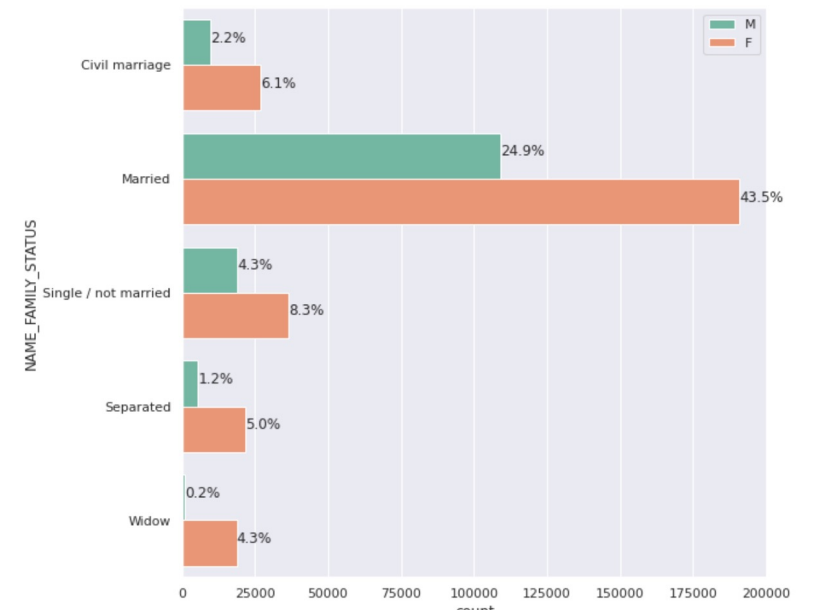
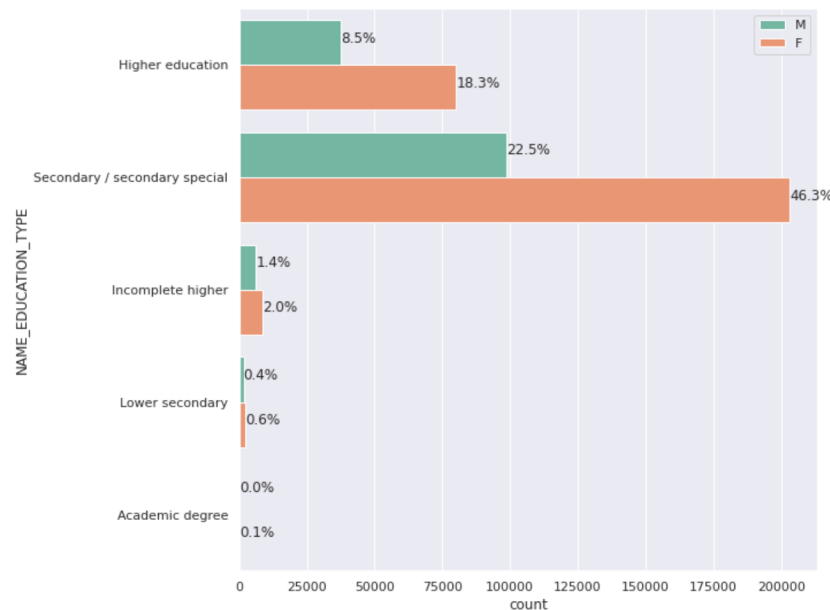


## Income type

Đại đa số KH đang đi làm hoặc tự kinh doanh: 19.4% KH nam và 32.2% KH nữ đang đi làm; 8.4% KH nam và 14.5% KH nữ đang tự kinh doanh.

## Education type

Phần lớn KH có trình độ ĐH và Trung học: 8.5% KH nam và 18.3% KH nữ đã học ĐH; 22.5% KH nam và 46.3% KH nữ đã học trung học

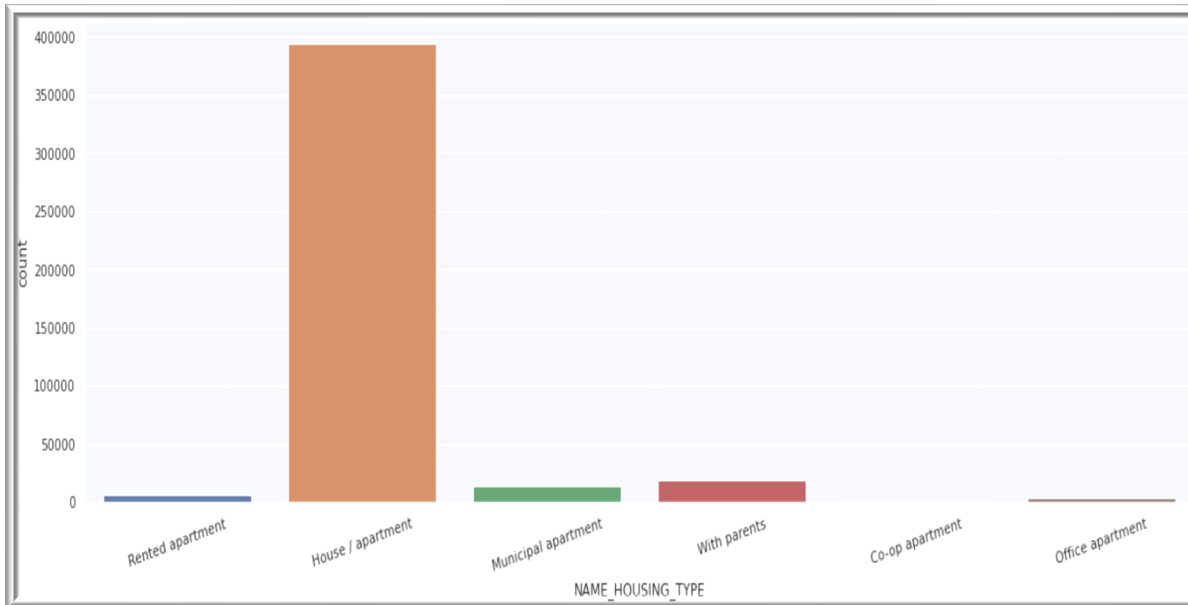
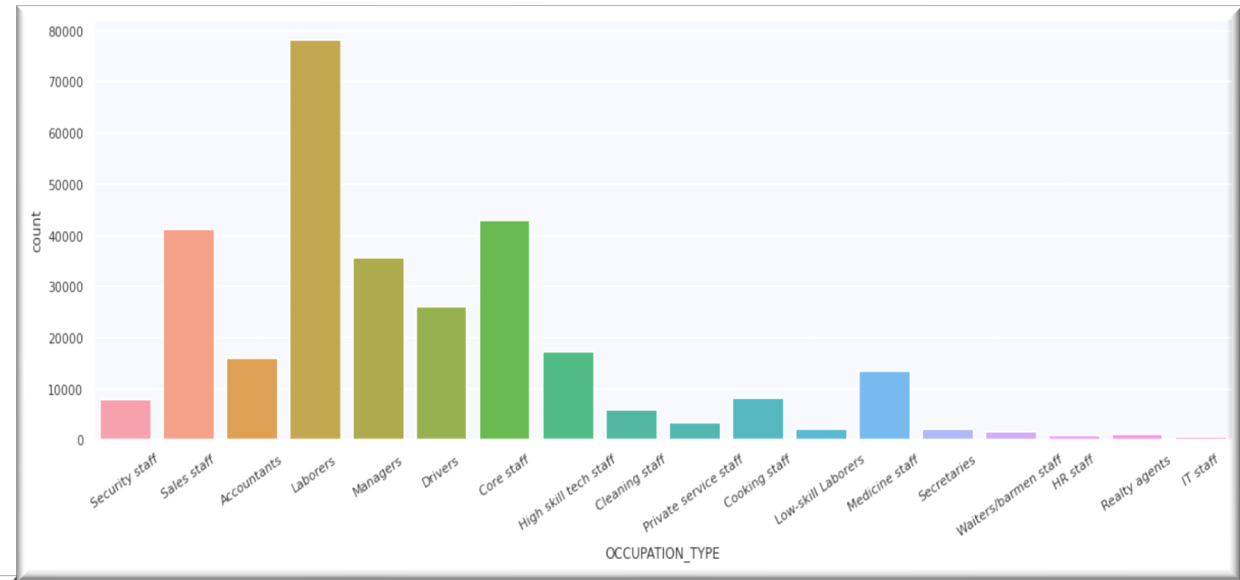


## Family status

Đa số KH đã kết hôn: 24.9% KH nam và 43.5% KH nữ

# Graphic

**Nghề nghiệp** phổ biến là  
Laborers, core staff, sales staff,  
manager, driver, high skill tech  
staff, accountant



**Nơi ở** phổ biến:  
House/apartment

# Xử lý file credit\_card

## Quy định NNNH

**Nhóm 1:** Dư nợ đủ chuẩn (Các khoản nợ được thanh toán trong hạn + Các khoản nợ quá hạn dưới 10 ngày)

**Nhóm 2:** Dư nợ cần chú ý (Các khoản nợ quá hạn từ 10 – 90 ngày)

**Nhóm 3:** Dư nợ dưới tiêu chuẩn (Các khoản nợ quá hạn từ 30 – 90 ngày)

**Nhóm 4:** Nợ nghi ngờ mất vốn (Các khoản nợ quá hạn từ 90 – 180 ngày)

**Nhóm 5:** Nợ có khả năng mất vốn (Các khoản nợ quá hạn hơn 180 ngày)

## Credit\_record [Status]

**C:** paid off that month

**X:** No loan for the month

**0:** 1-29 days past due

**1:** 30-59 days past due

**2:** 60-89 days overdue

**3:** 90-119 days overdue

**4:** 120-149 days overdue


**5:** Overdue or bad debts, write-offs for more than 150 days

## Phân nhóm

Nhóm - **Good\_Debt:** status **C, X, 0**

Nhóm - **Neutral\_Debt:** status **1, 2**

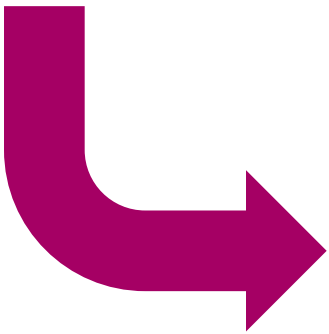
Nhóm - **Bad\_Debt:** status **3, 4, 5**

 `credit_record.head()`



	ID	MONTHS_BALANCE	STATUS
0	5001711	0	X
1	5001711	-1	0
2	5001711	-2	0
3	5001711	-3	
4	5001712	0	

```
[ ] # Xếp hạng tín dụng của KH theo đa số của loại nhóm nợ
credit_record.loc[(credit_record['Good_Debt'] > credit_record['Neutral_Debt']), 'CREDIT_APPROVAL_STATUS'] = 2
credit_record.loc[(credit_record['Good_Debt'] > credit_record['Bad_Debt']), 'CREDIT_APPROVAL_STATUS'] = 2
credit_record.loc[(credit_record['Neutral_Debt'] > credit_record['Good_Debt']), 'CREDIT_APPROVAL_STATUS'] = 1
credit_record.loc[(credit_record['Neutral_Debt'] > credit_record['Bad_Debt']), 'CREDIT_APPROVAL_STATUS'] = 1
credit_record.loc[(credit_record['Bad_Debt'] > credit_record['Good_Debt']), 'CREDIT_APPROVAL_STATUS'] = 0
credit_record.loc[(credit_record['Bad_Debt'] > credit_record['Neutral_Debt']), 'CREDIT_APPROVAL_STATUS'] = 0
```

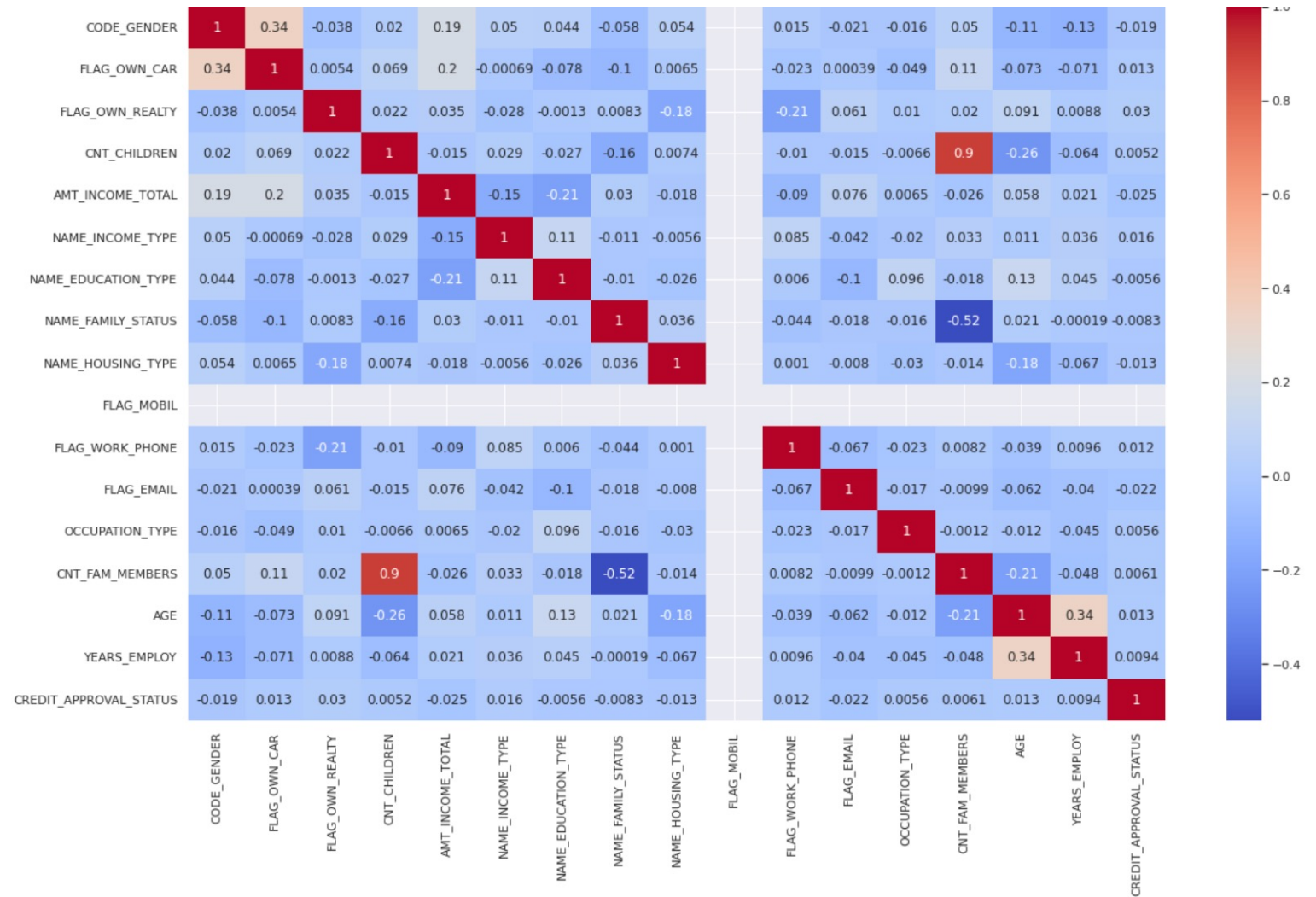


	STATUS	Bad_Debt	Good_Debt	Neutral_Debt	CREDIT_APPROVAL_STATUS
ID					
5001711		0	4	0	2.0
5001712		0	19	0	2.0
5001713		0	22	0	2.0
5001714		0	15	0	2.0
5001715		0	60	0	2.0
...		...	...	...	...
5150482		0	18	0	2.0



# Data preparation

- **Thêm 2 cột:** Age (quy đổi từ Day of birth), Years\_employ (quy đổi từ Days\_employ)
  - **Normalize:** amt\_income\_total
  - **Mã hoá:** code\_gender, flag\_own\_car...
  - **Merge 2 file** theo ID number
  - **Drop:** ID, flag\_phone, days\_birth, days\_employ
- => Dataset còn 17 features để phân tích



# Model Building

RF

	precision	recall	f1-score	support
Bad_Debt	0.39	0.25	0.31	36
Neutral_Debt	0.41	0.43	0.42	881
Good_Debt	0.92	0.92	0.92	6624

KNN

	precision	recall	f1-score	support
0.0	0.12	0.19	0.15	36
1.0	0.34	0.40	0.37	881
2.0	0.92	0.89	0.90	6624

Adaboost

	precision	recall	f1-score	support
0.0	0.44	0.33	0.38	36
1.0	0.24	0.36	0.29	881
2.0	0.91	0.85	0.88	6624

STT	Model	Accuracy rate
1	Random Forest	86%
2	KNN	83%
3	Adaboost	79%
4	XGBoost	69%
5	Decision Tree	63%
6	GradientBoost	61%
7	LR	55%
8	MNB	52%
9	GNB	48%
10	SVM	38%

# AdaBoost - Confusion matrix

	precision	recall	f1-score	support
Bad_Debt	0.44	0.33	0.38	36
Neutral_Debt	0.24	0.36	0.29	881
Good_Debt	0.91	0.85	0.88	6624
accuracy			0.79	7541
macro avg	0.53	0.51	0.52	7541
weighted avg	0.83	0.79	0.80	7541

## Bad Debt

TP = 11

FN = 5 + 20 = 25

FP = 3 + 10 = 13

TN = 7492

## Neutral Debt

TP = 323

FN = 3 + 555 = 558

FP = 5 + 1003 = 1008

TN = 5652

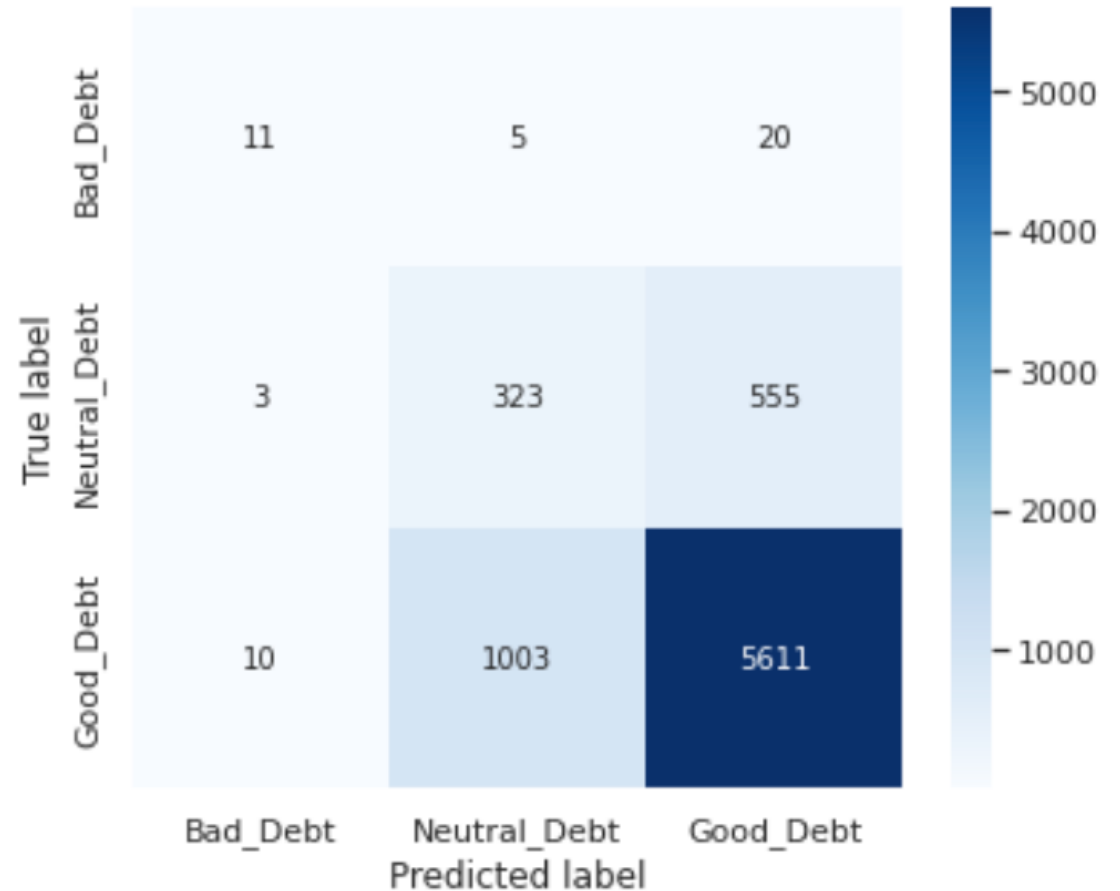
## Good Debt

TP = 5611

FN = 10 + 1003 = 1013

FP = 20 + 555 = 575

TN = 342



The background is a vibrant magenta color. In the top-left corner, there are several overlapping geometric shapes: a yellow triangle, a pink parallelogram, a dark blue parallelogram, and a light green parallelogram. In the top-right corner, there is a teal shape with diagonal white stripes. In the bottom-left corner, there is a yellow shape with diagonal white stripes. In the bottom-right corner, there is a white grid of dots and a white plus sign next to a white circle.

**Thank you for your great support!**