

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

Introduction the dataset

- **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:** Xây dựng mô hình học máy để dự đoán xem KH “tốt”, “xấu”
- Gồm 2 file csv:
 - + **application_record:** 438557 rows x 18 columns
 - + **credit_record:** 1048575 rows × 3 columns

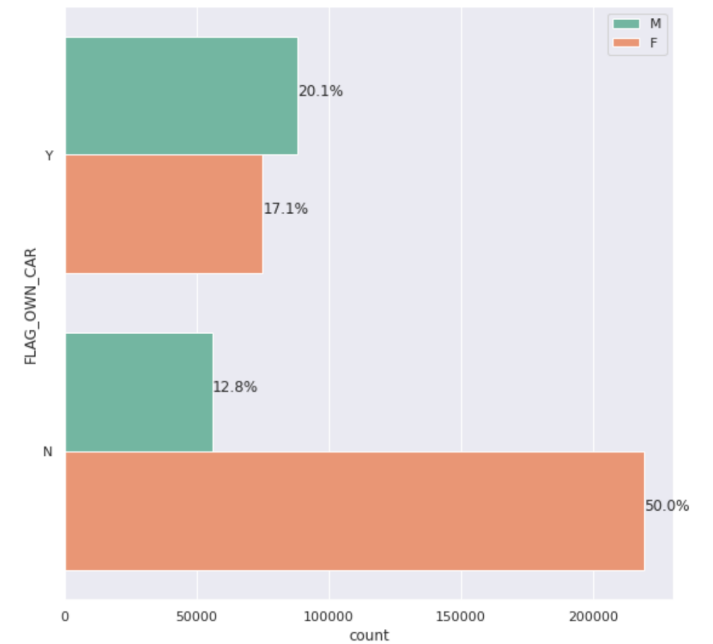
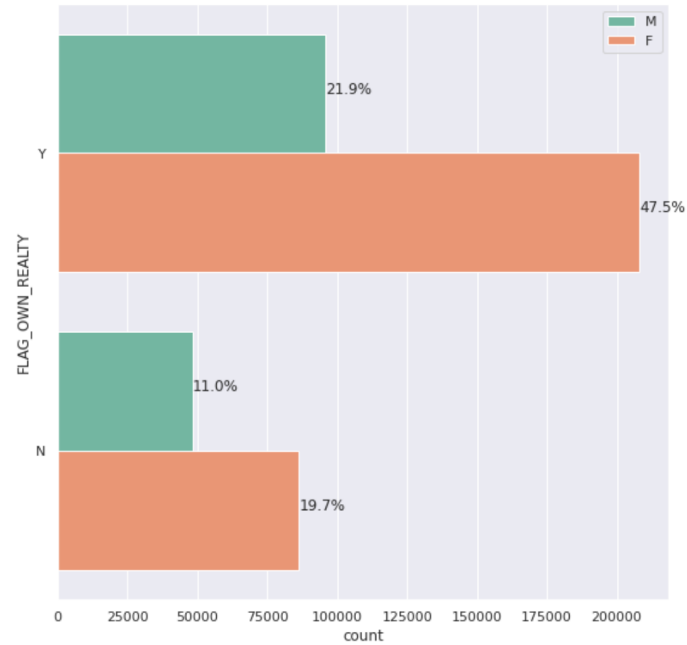
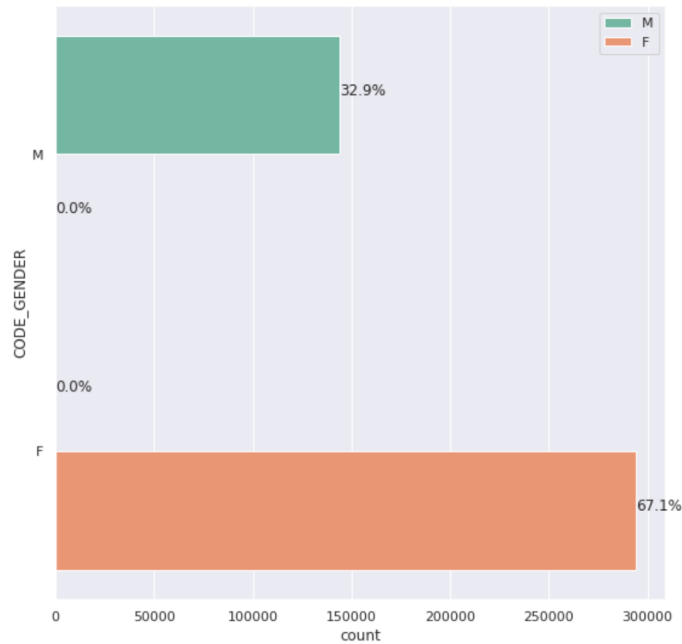
application_record.csv

Feature name	Explanation	Remarks
ID	Client number	
CODE_GENDER	Gender	
FLAG_OWN_CAR	Is there a car	
FLAG_OWN_REALTY	Is there a property	
CNT_CHILDREN	Number of children	
AMT_INCOME_TOTAL	Annual income	
NAME_INCOME_TYPE	Income category	
NAME_EDUCATION_TYPE	Education level	
NAME_FAMILY_STATUS	Marital status	
NAME_HOUSING_TYPE	Way of living	
DAYS_BIRTH	Birthday	Count backwards from current day (0), -1 means yesterday
DAYS_EMPLOYED	Start date of employment	Count backwards from current day(0). If positive, it means the person currently unemployed.
FLAG_MOBIL	Is there a mobile phone	
FLAG_WORK_PHONE	Is there a work phone	
FLAG_PHONE	Is there a phone	
FLAG_EMAIL	Is there an email	
OCCUPATION_TYPE	Occupation	
CNT_FAM_MEMBERS	Family size	

credit_record.csv

Feature name	Explanation	Remarks
ID	Client number	
MONTHS_BALANCE	Record month	The month of the extracted data is the starting point, backwards, 0 is the current month, -1 is the previous month, and so on
STATUS	Status	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 C: paid off that month X: No loan for the month

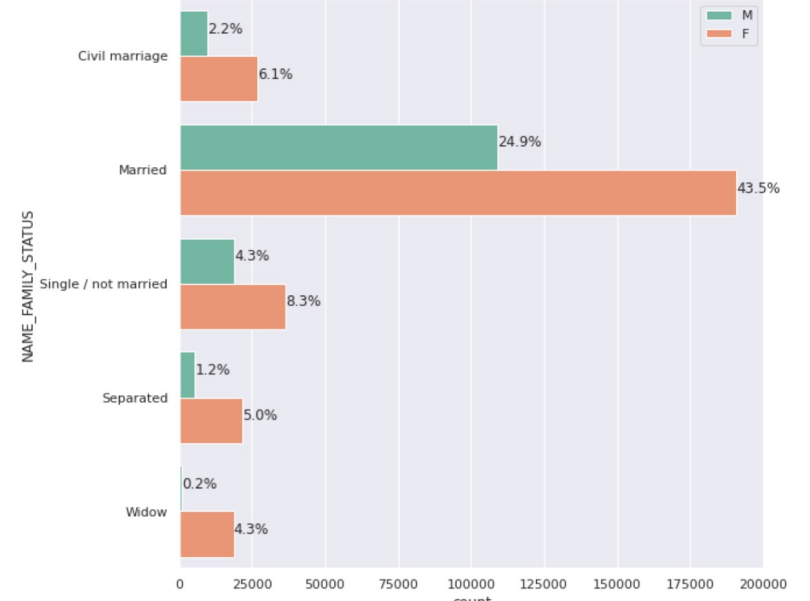
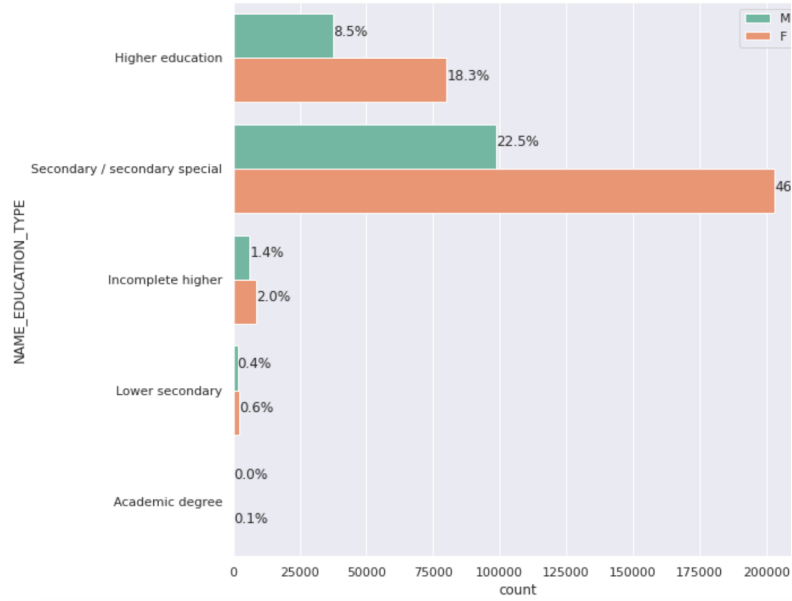
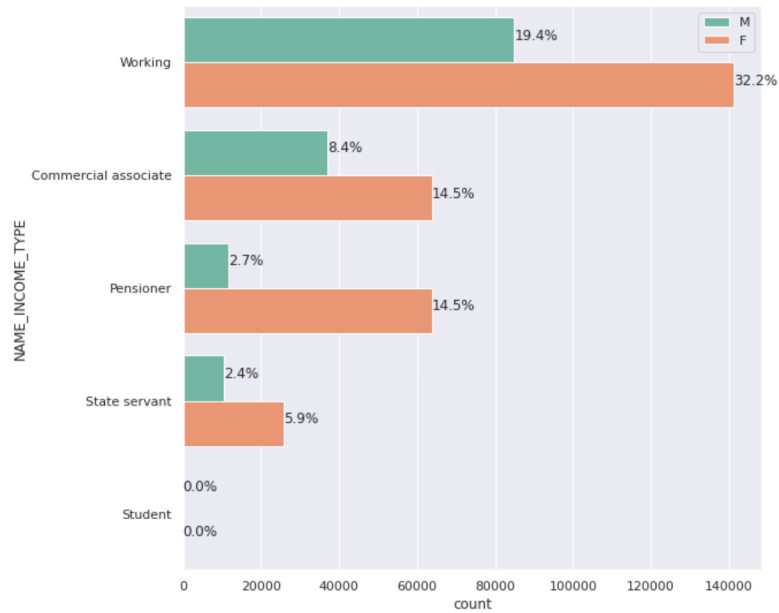
Graphic



Insights from application details record:

- 36.9% KH là Nam và 67.1% là nữ
- 20.1% KH nam và 17.1% KH nữ sở hữu oto
- 21.9% KH nam và 47.5% KH nữ sở hữu BĐS

Graphic



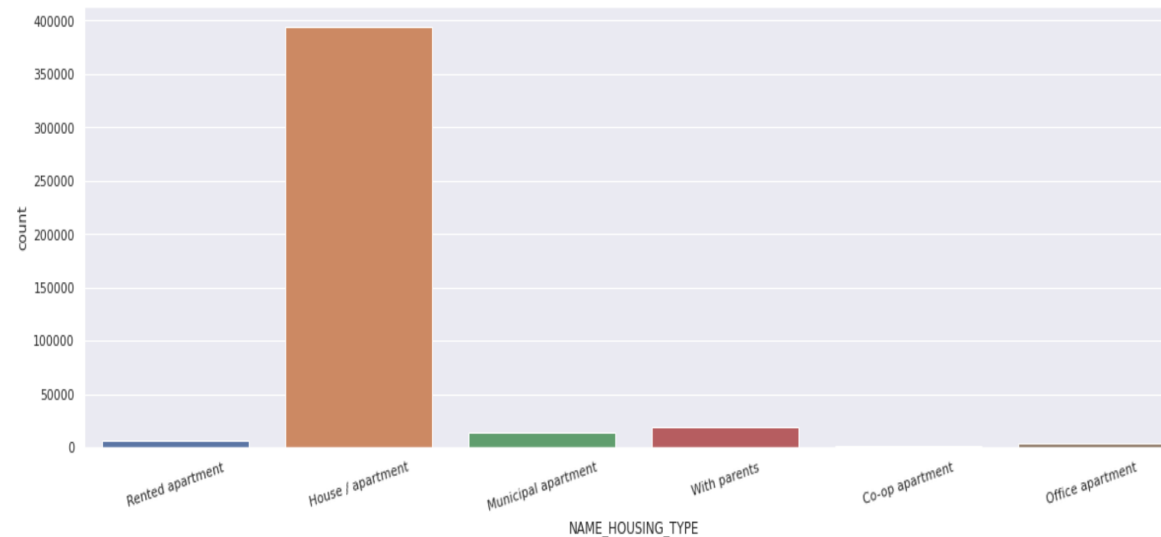
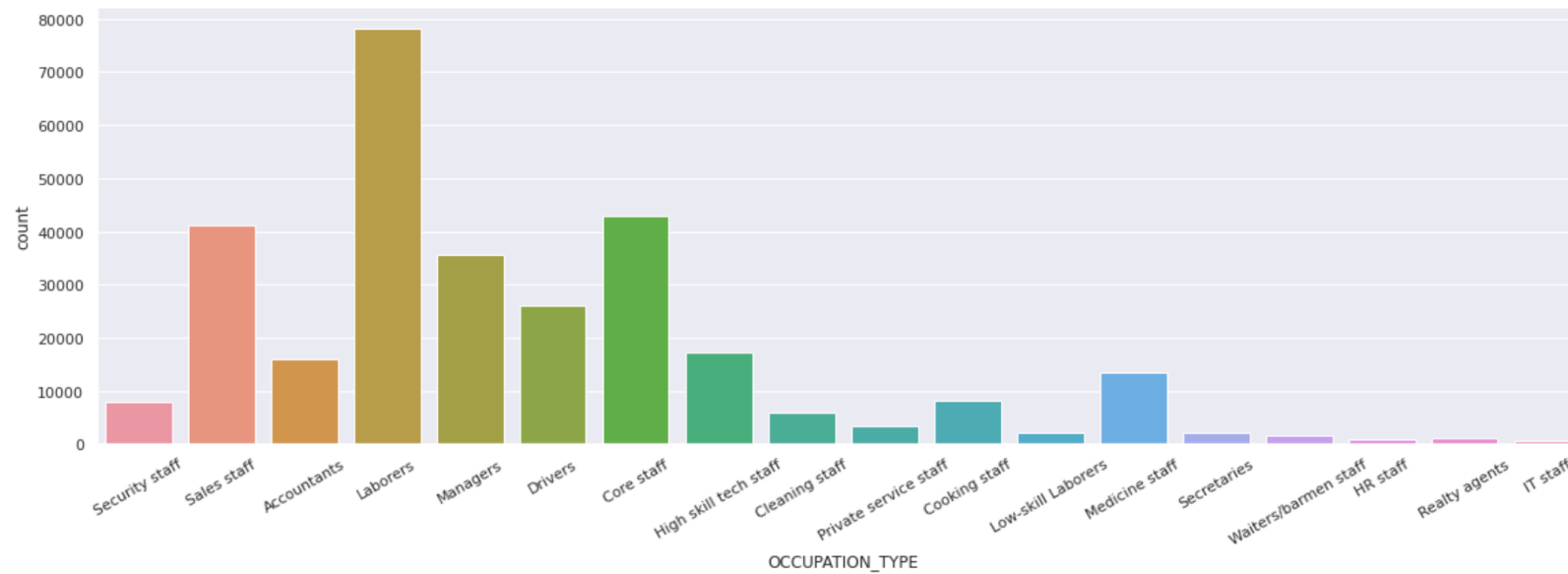
19.4% KH nam và 32.2% KH nữ đang làm KD

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

24.9% KH nam và 43.5% KH nữ đã kết hôn

Graphic



- Nghề nghiệp phổ biến là Laborers, sales staff, Manager, core staff, 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 (Nợ xấu)(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 1 - "**Good_Debt**": status **C, X, 0**

Nhóm 2 - "**Neutral_Debt**": status **1, 2**

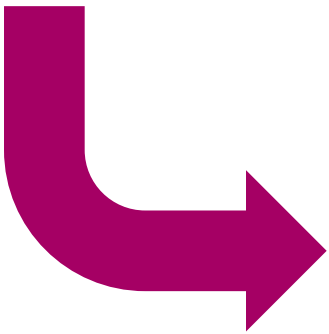
Nhóm 3 - "**Bad_Debt**": status **3, 4, 5**

 `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	

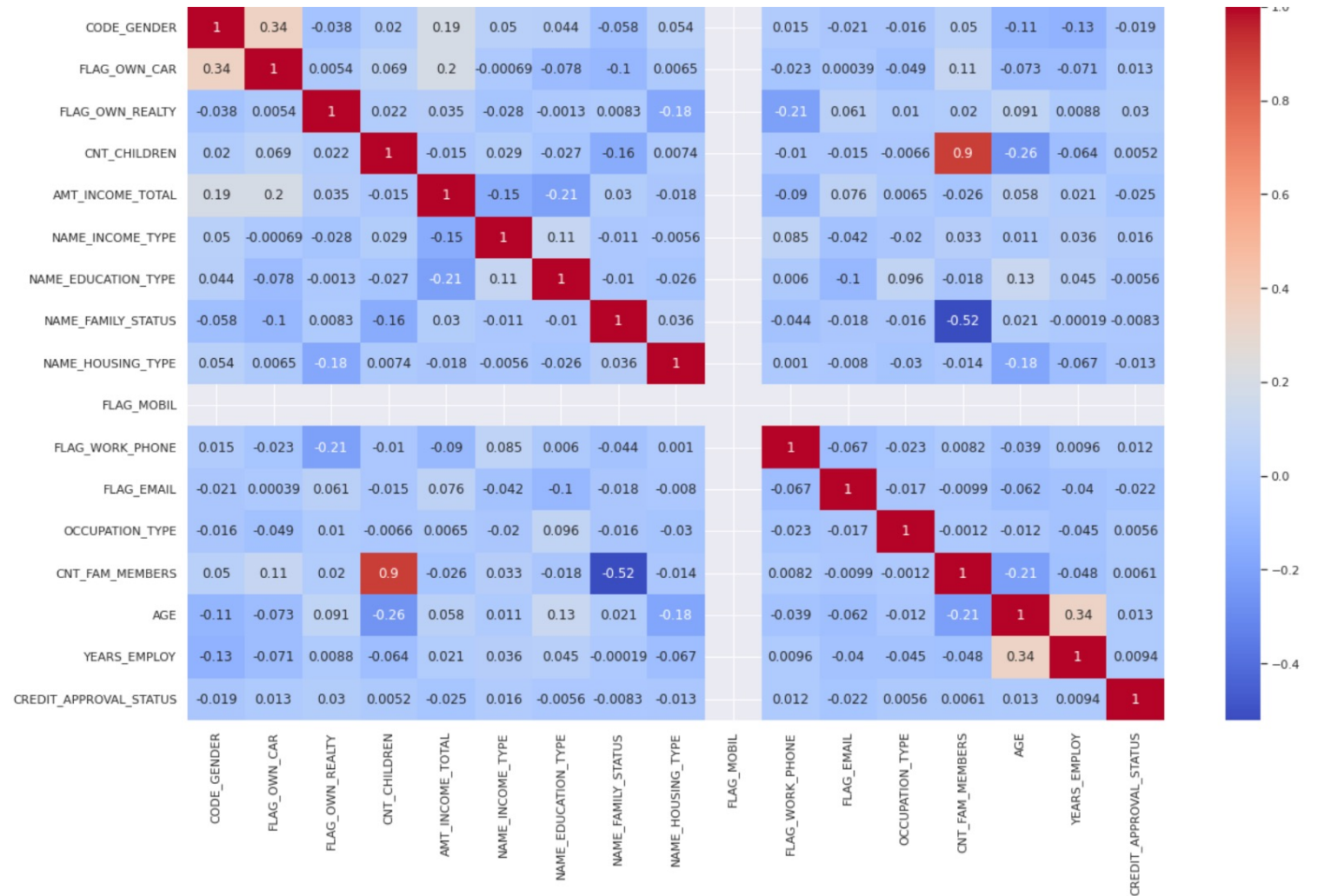
```
[ ] # Xếp hạng tín dụng của KH theo đa số của loại nhóm nợ
credit_record.loc[(credit_record['Good_Debt'] > credit_record['Neutral_Debt']), 'CREDIT_APPROVAL_STATUS'] = 2
credit_record.loc[(credit_record['Good_Debt'] > credit_record['Bad_Debt']), 'CREDIT_APPROVAL_STATUS'] = 2
credit_record.loc[(credit_record['Neutral_Debt'] > 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)
 - **Nomalize:** amt_income_total
 - **Mã hoá:** code_gender, flag_own_car...
 - **Merge 2 file** theo ID number
 - **Drop:** ID, FLAG_PHONE, DAYS_BIRTH, DAYS_EMPLOYED
- => Dataset có 17 features để phân tích



Model Building

10 Model: Random forest, KNN, Adaboost có Accuracy rate cao nhất (86%, 83%, 79%)

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	Decision Tree	62.67 %
2	GNB	47.82 %
3	MNB	52.23 %
4	LR	54.58 %
5	SVM	37.78 %
6	Random Forest	86 %
7	KNN	83 %
8	Adaboost	79 %
9	GradientBoost	60.58 %
10	XGBoost	69.45 %

Confusion matrix

	precision	recall	f1-score	support
Bad_Debt	0.39	0.25	0.31	36
Neutral_Debt	0.40	0.42	0.41	881
Good_Debt	0.92	0.91	0.92	6624
accuracy			0.85	7541
macro avg	0.57	0.53	0.54	7541
weighted avg	0.86	0.85	0.85	7541

Bad Debt

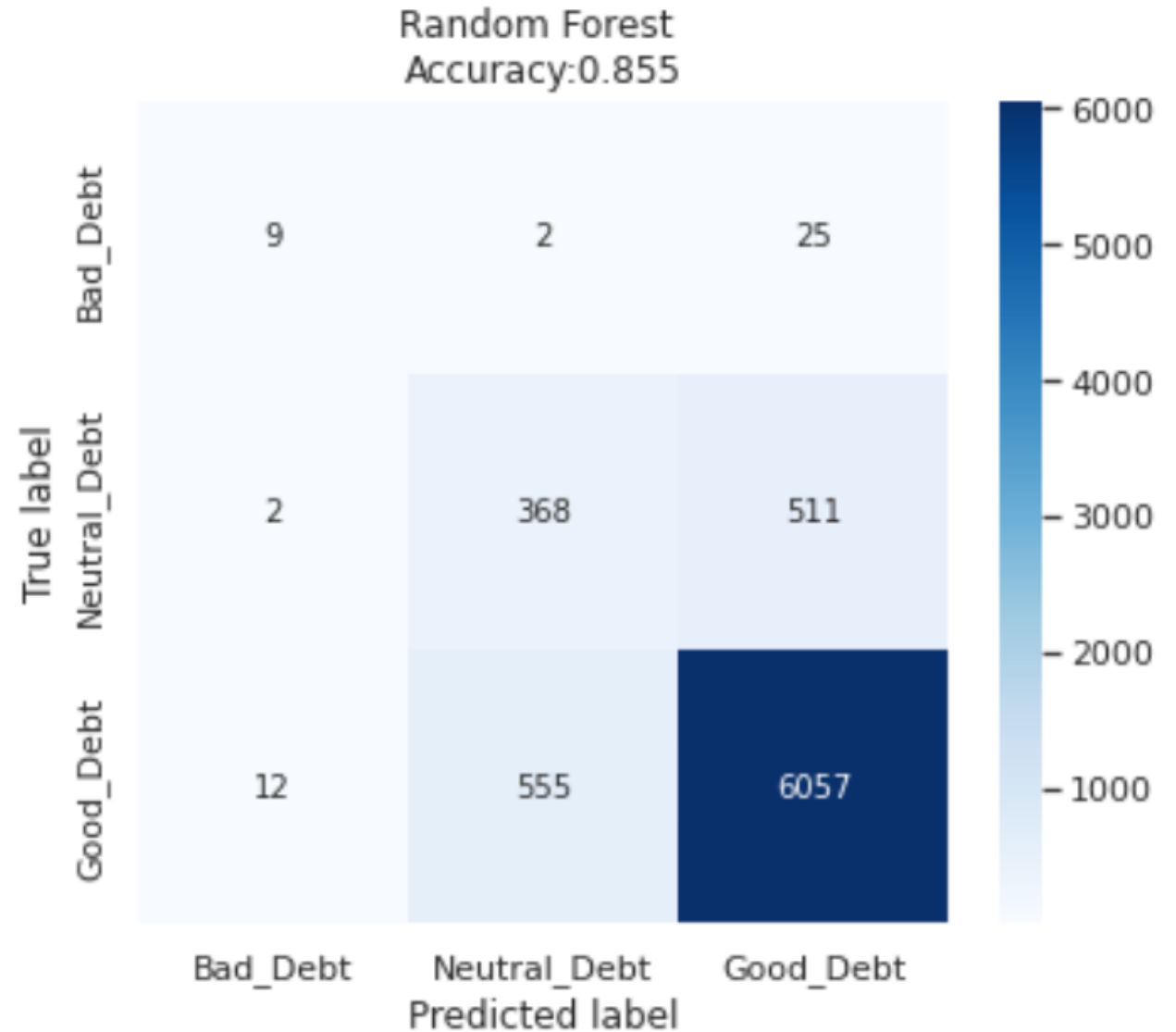
TP = 9
FN = 2 + 25 = 27
FP = 2 + 12 = 14
TN = 7491

Neutral Debt

TP = 368
FN = 2 + 511 = 513
FP = 2 + 555 = 557
TN = 6103

Good Debt

TP = 6057
FN = 12 + 555 = 567
FP = 25 + 511 = 536
TN = 381



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 pink plus sign next to a pink circle.

Thank you for your great support!