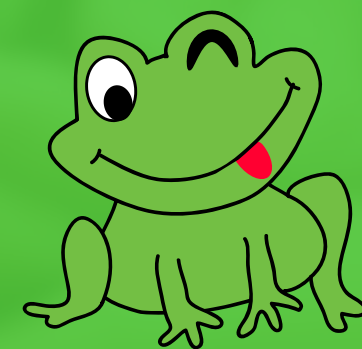




# DỰ ĐOÁN RỦI RO TÍN DỤNG

ĐỘI THI: FROG



Data Science Talent  
Competition 2024



# MỤC LỤC



**01 LÀM SẠCH DỮ LIỆU**

**02 TỔNG QUAN**

**03 PHÂN TÍCH**

**04 XÂY DỰNG MÔ HÌNH**

**05 KẾT LUẬN & ĐỀ XUẤT**

# Xử lý giá trị null

Trung bình mỗi cột có khoảng 2000 giá trị null cần được xử lý. Nhóm đã xử lý bằng cách tiếp cận từng cột và sử dụng mối liên hệ giữa cột cần xử lý và các cột trong cùng một nhóm.

Tên nhóm cột	Mô tả	Biện pháp
NUMBER_OF_LOANS_	Tổng tất cả khoản vay (không phân biệt loại) từ ngân hàng - tổ chức phi ngân hàng	Dựa trên tổng số khoản vay từ ngân hàng và tổ chức phi ngân hàng, lấp đầy các khoản vay ngắn, trung và dài hạn.
NUM_NEW_LOAN_TAKEN	Tổng số khoản vay mới từ ngân hàng - tổ chức phi ngân hàng mà khách hàng thực hiện	Độ tương quan giữa các cột số khoản vay mới bằng 1 nên dùng giá trị tương ứng tại các cột khác để điền giá trị còn thiếu
OUTSTANDING_BAL_LOAN_	Số dư nợ của các khoản vay	Điền dữ liệu dựa vào sự tương quan với dữ liệu của cột số dư của thẻ tín dụng và các khoản vay hiện tại
ENQUIRIES_FROM_FOR_xM	Số lượt tra cứu tín dụng liên quan đến các khoản vay - thẻ tín dụng từ ngân hàng - tổ chức phi ngân hàng	Sử dụng tổng số tra cứu từ ngân hàng và tổ chức phi ngân hàng để lấp đầy.

Với các giá trị null còn lại không theo các quy tắc trên, nhóm sử dụng giá trị trung bình của cột đó để điền.

# Xử lý giá trị ngoại lai và giảm số lượng cột



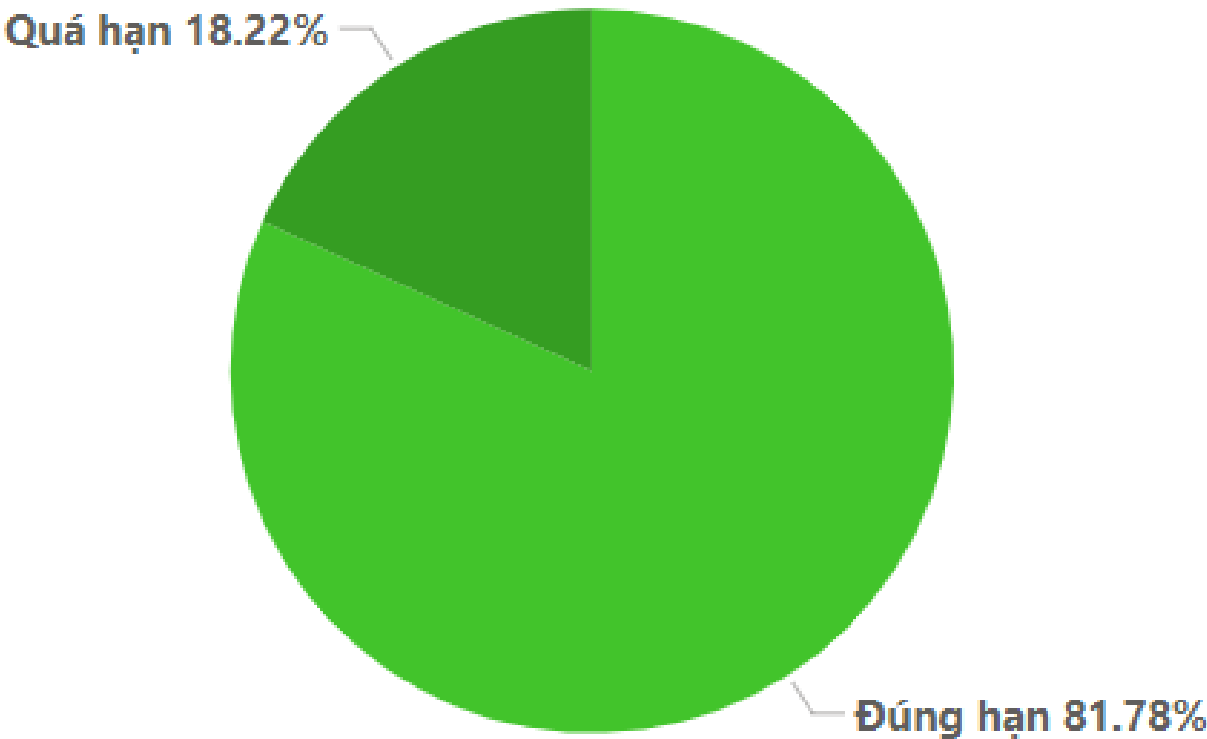
Việc loại bỏ giá trị ngoại lai rải rác có thể làm giảm độ chính xác của mô hình dự đoán. Nhóm đã chọn chuẩn hóa dữ liệu bằng StandardScaler().

Nhóm đã áp dụng **PCA**, **RFE**, và **KBest** để giảm số lượng cột. Và sử dụng mô hình **Random Forest Classifier** để đánh giá các phương pháp này

Kết quả cho thấy **RFE** cho độ chính xác cao nhất, do đó nhóm quyết định giảm số lượng cột xuống 50.

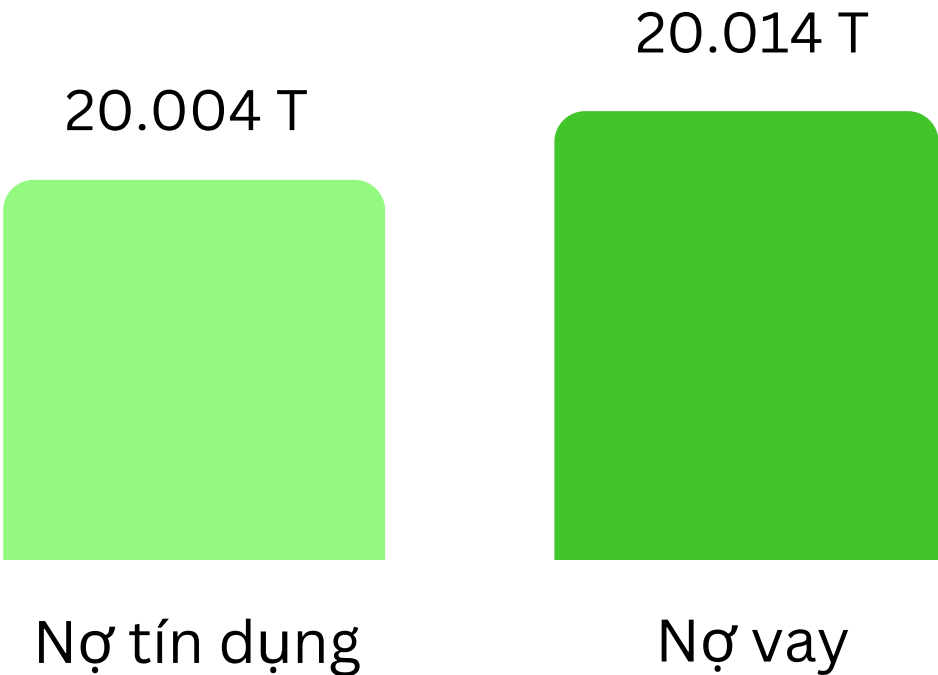
Mô hình đánh giá	Random Forest Classifier		
Phương pháp	PCA	RFE	KBest
Tổng phương sai các cột còn lại	0.972	1.0	1.0
Accuracy	0.8835	0.8877	0.8850

### Trạng thái trả nợ của khách hàng



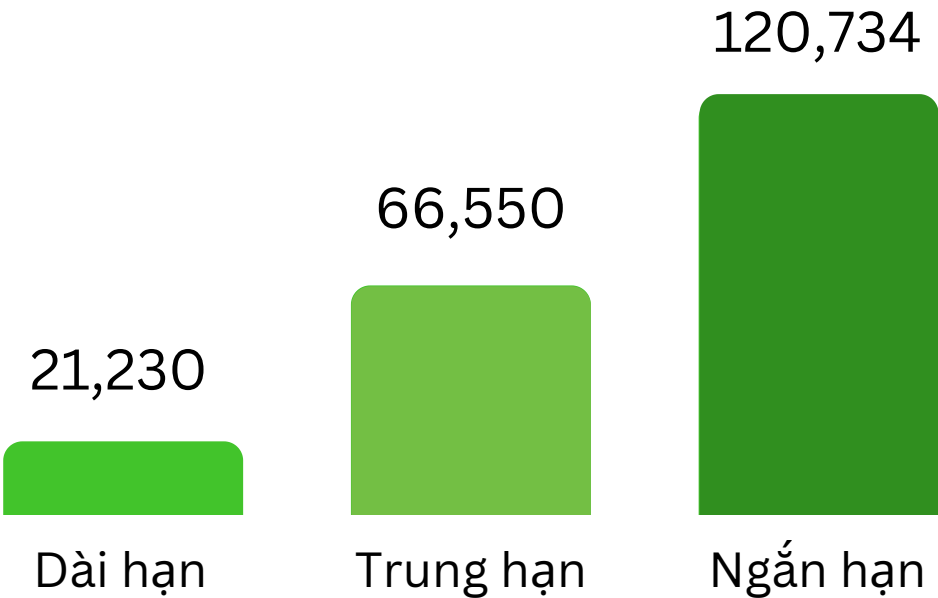
Dữ liệu có 20000 khách hàng, trong đó **3644** (18.2%) khách hàng trả nợ **quá hạn** và **16356** (81,8%) khách hàng trả nợ **đúng hạn**.

### Tổng dư nợ (đvtt)



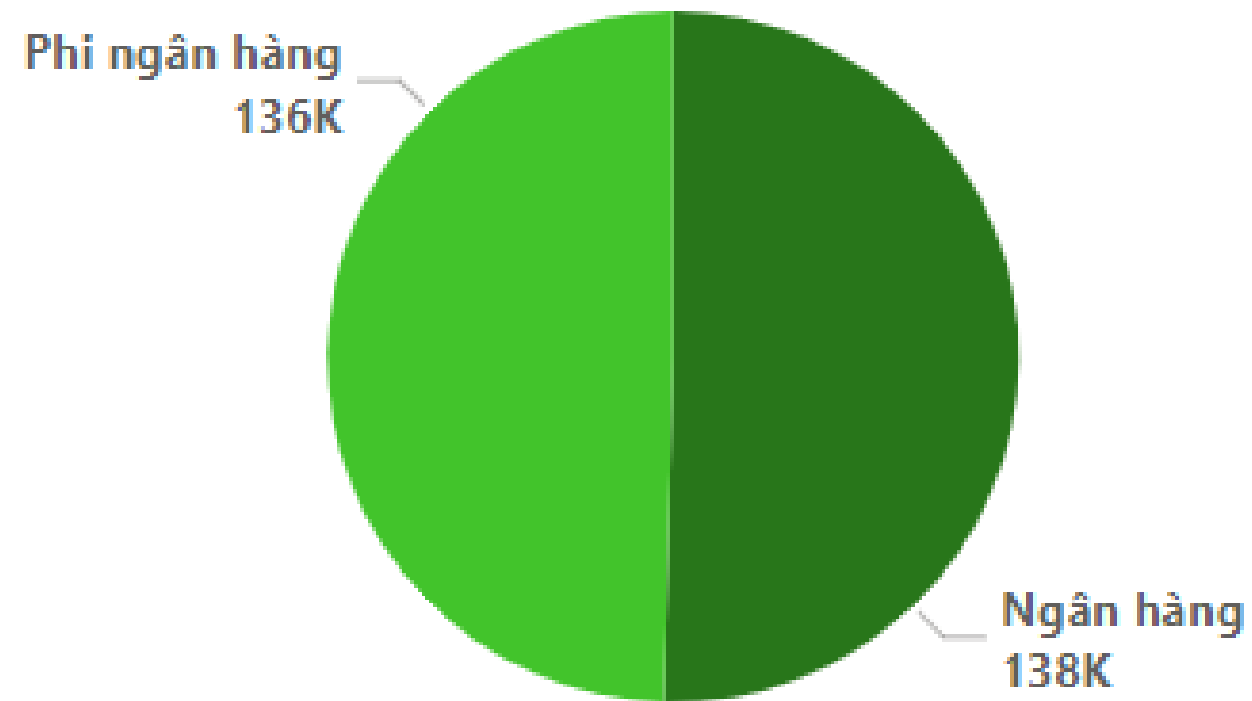
**Tổng dư nợ** là 18,78 tỷ đơn vị tiền tệ (đvtt), trong đó nợ vay là 18,778 tỷ, nợ tín dụng là 18,775 tỷ.

### Tổng số khoản vay theo kỳ hạn



Các khoản vay **ngắn hạn** (dưới 1 năm) chiếm số lượng nhiều nhất, theo sau là các khoản vay **trung hạn** (1 đến 5 năm) và **dài hạn** (> 5 năm).

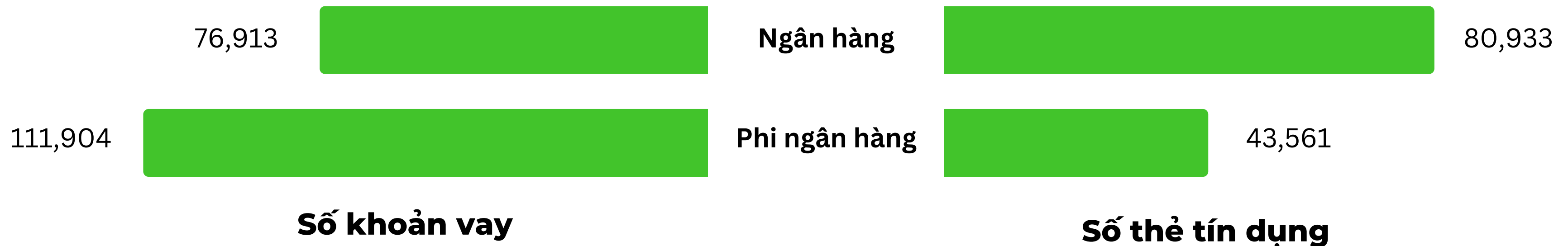
## Tổng số mối quan hệ tài chính



Số mối quan hệ tài chính mà khách hàng có với các tổ chức ngân hàng và phi ngân hàng khá tương đồng nhau.

Như vậy, trung bình mỗi khách hàng có 6.9 mối quan hệ với ngân hàng và 6.8 mối quan hệ với tổ chức phi ngân hàng.

Tuy nhiên, khách hàng lại có xu hướng mở thẻ tín dụng nhiều hơn tại ngân hàng và vay nhiều khoản vay hơn tại các tổ chức phi ngân hàng



Làm sạch dữ liệu

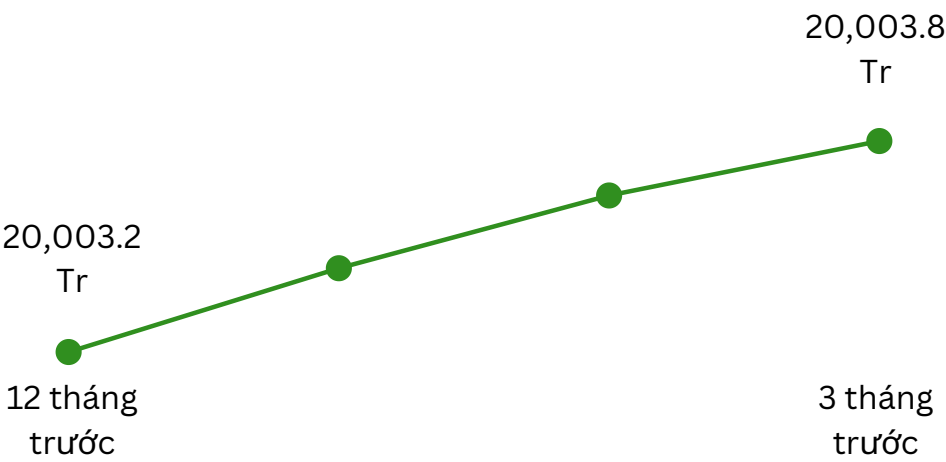
Tổng quan

Phân tích

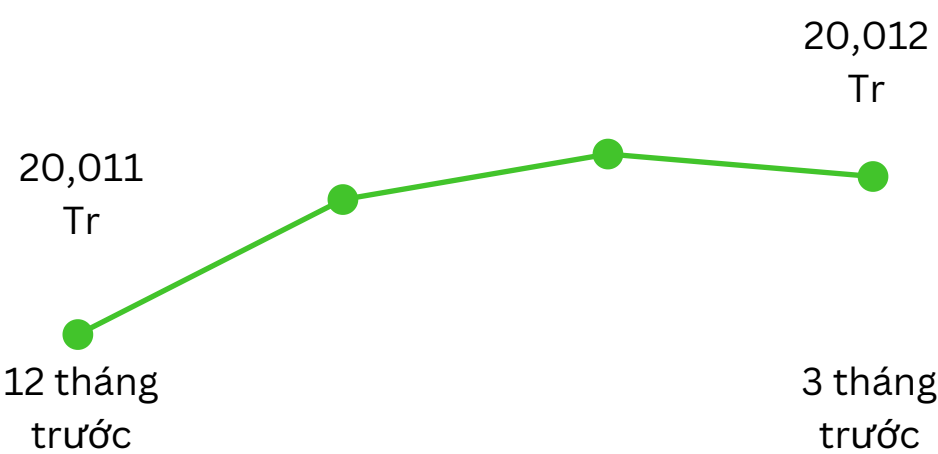
Xây dựng mô hình

Kết luận & Đề xuất

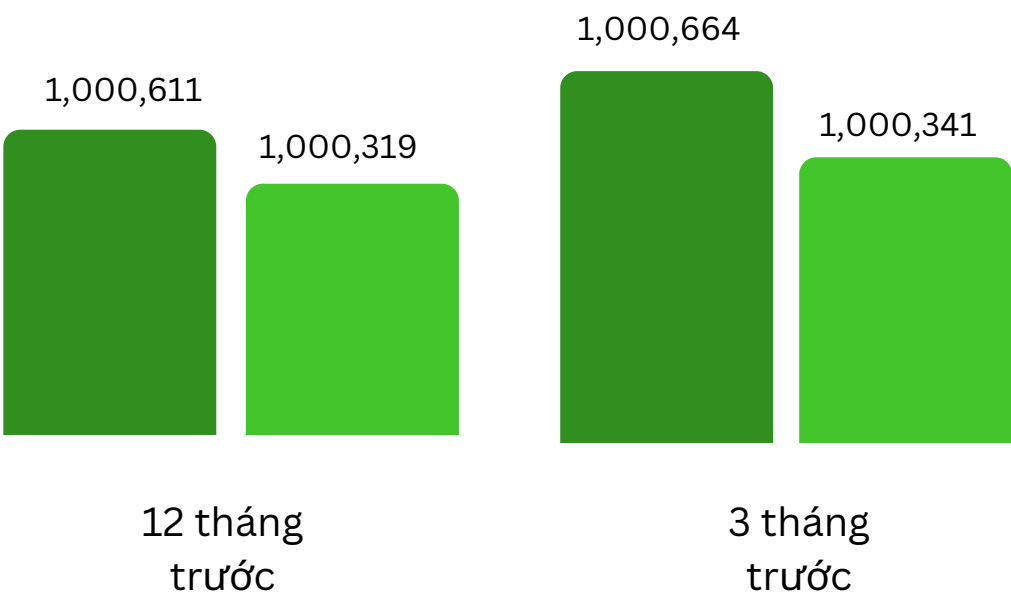
Tổng dư nợ tín dụng trong giai đoạn 12 tháng - 3 tháng trước



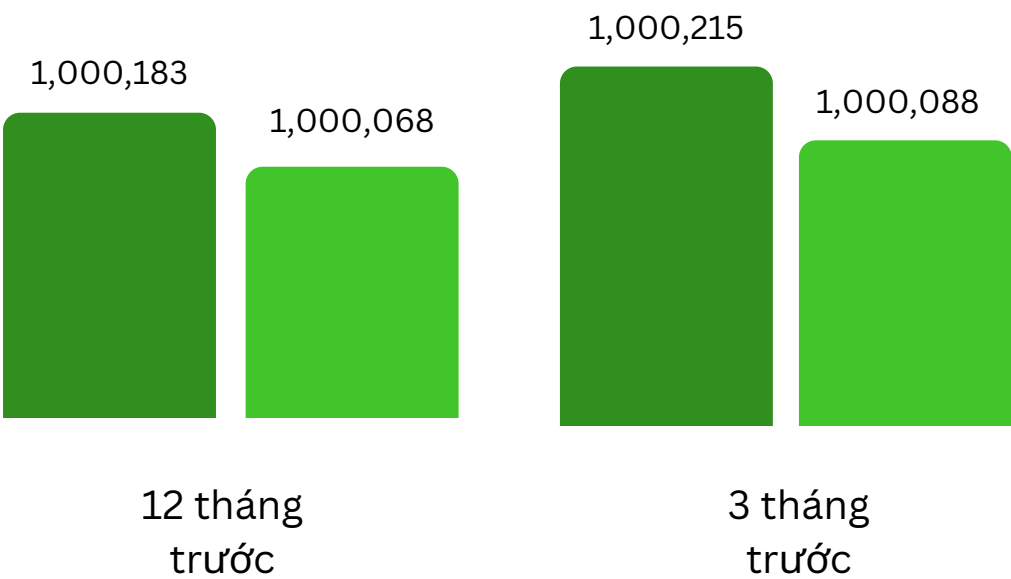
Tổng dư nợ vay trong giai đoạn 12 tháng - 3 tháng trước



Trung bình nợ vay theo nhóm khách hàng



Trung bình nợ tín dụng theo nhóm khách hàng



Khách hàng trả nợ đúng hạn      Khách hàng quá hạn trả nợ

Tổng số dư nợ đang có **xu hướng tăng lên** trong giai đoạn 12 tháng đến 3 tháng trở lại, đặc biệt là số dư nợ tín dụng.

Trong đó, nhóm khách hàng **trả nợ đúng hạn** có xu hướng **vay nhiều hơn** nhóm quá hạn trả nợ. Đồng thời, đây cũng là nhóm có trung bình **số dư nợ cao hơn**.

## Khi phân tích thêm, ta thấy được:



Trung bình số  
khoản vay tăng  
thêm

84,5  Trung bình số  
lượt tìm kiếm

Là số khoản vay tăng thêm và số lượt tìm kiếm trung bình của một khách hàng thuộc **nhóm trả nợ đúng hạn** trong khoảng thời gian 12 - 3 tháng trước.



Trung bình số  
khoản vay tăng  
thêm

65,2  Trung bình số  
lượt tìm kiếm

Là các con số tương ứng của **nhóm trả nợ quá hạn**. Cả 2 đều thấp hơn nhóm trả nợ đúng hạn.

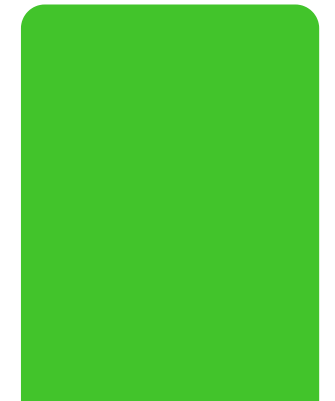
## Trung bình số lần nợ tín dụng quá hạn trong quá khứ

1.72



Đúng hạn

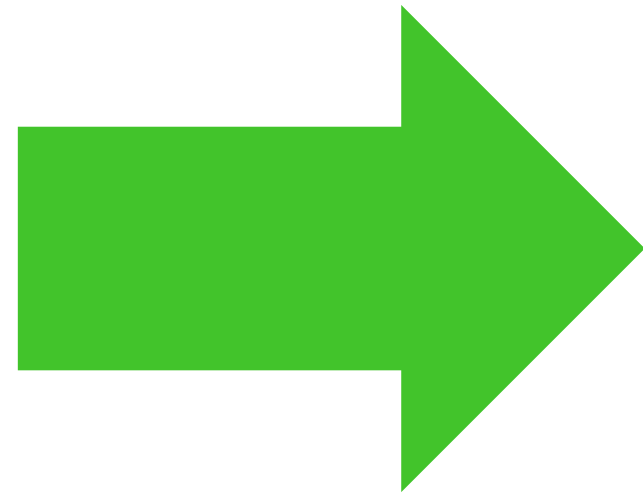
1.50



Quá hạn

Một điểm đáng chú ý khác đó là nhóm **khách hàng trả nợ đúng hạn** ở hiện tại lại có số lần **nợ tín dụng quá hạn nhiều hơn** trong quá khứ.



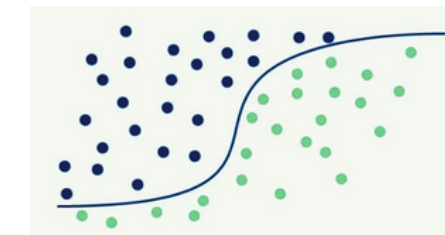
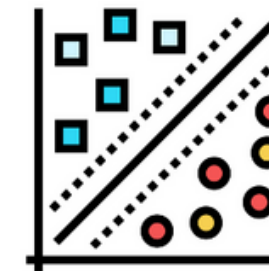


Như vậy, rủi ro tín dụng không chỉ nằm ở nhóm khách hàng trả nợ quá hạn mà **có xu hướng tăng lên ở nhóm khách hàng trả nợ đúng hạn** trong khoảng thời gian 12 tháng trở lại đây khi nhóm này cho thấy sự tăng lên trong các khoản vay và tín dụng, tiềm ẩn **nguy cơ chậm trả nợ**.



Nhóm đã nghiên cứu và đề xuất **mô hình dự đoán khả năng trả nợ** dựa trên ứng dụng của 10 thuật toán Machine Learning khác nhau, từ đó lựa chọn ra mô hình phù hợp nhất để áp dụng thực tiễn.

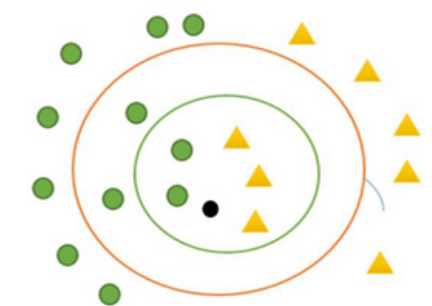
ADA BOOST  
CLASSIFICATION



NEURAL NETWORK

RANDOM  
FORESTS

XGBoost



Làm sạch dữ liệu

Tổng quan

Phân tích

Xây dựng mô hình

Kết luận & Đề xuất

# So sánh tổng quát các mô hình

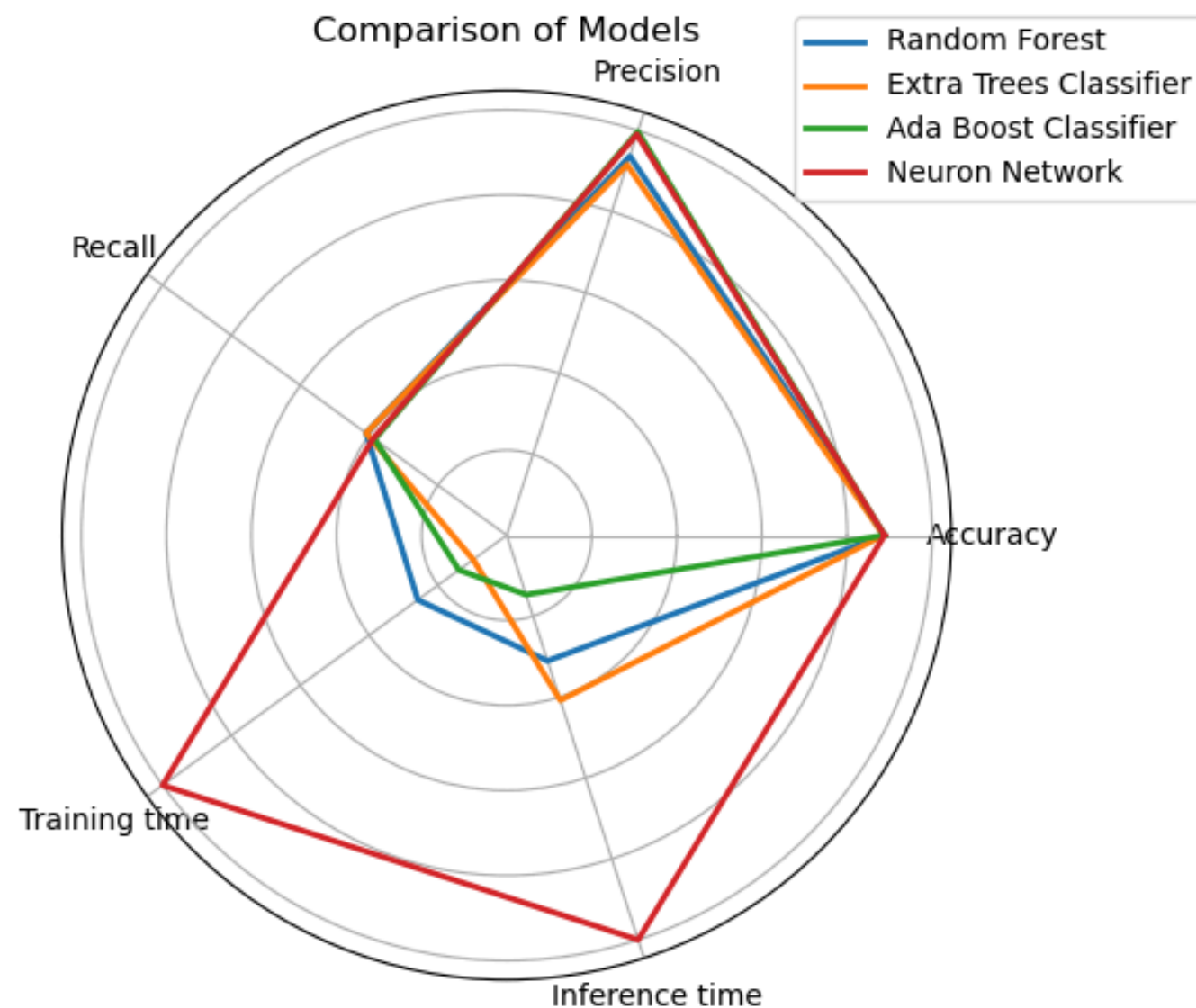
Mô hình **Ada Boost Classifier** có độ chính xác cao nhất, cho thấy nó có hiệu suất tốt nhất trong việc phân loại dữ liệu trong các mô hình được so sánh.

**Random Forest, Neuron Network, và Extra Trees Classifier** đều có độ chính xác gần nhau, cho thấy chúng là những lựa chọn tiềm năng, không kém cạnh so với Ada Boost Classifier.



Model	Accuracy
Ada Boost Classifier	0.89
Random Forest	0.89
Neuron Network	0.89
Extra Tree Classifier	0.89
Support Vector Machine	0.88
XG Boost	0.88
Logistic Regression	0.87
K-nearest Neighbors	0.87
Decision Trees	0.81
Naive Bayes	0.52

# Phân tích sâu hơn 4 mô hình nhận thấy



## Độ chính xác (accuracy)

**Neural Network (0.8882)** và **AdaBoost (0.8880)** có độ chính xác cao, nhưng không phải yếu tố chính do phân bố khách hàng không đều.



## Độ Chính Xác Dự Đoán (Precision):

**AdaBoost (0.9976)** và **Neural Network (0.9906)** cho kết quả cao, giảm dương tính giả. **Random Forest (0.9368)** và **Extra Trees (0.9156)** có nhiều khách hàng đúng hạn bị dự đoán quá hạn.



## Độ Nhạy (Recall):

**Random Forest** và **Extra Trees** chỉ phát hiện **40.86%** khách hàng quá hạn. **AdaBoost (0.3838)** và **Neural Network (0.3875)** có Recall thấp hơn.



## Thời Gian Huấn Luyện:

**Extra Trees (4.13 giây)** nhanh nhất, Mạng **Nơ-ron (42.45 giây)** chậm nhất.



## Thời Gian Suy Diễn:

**AdaBoost (0.11 giây)** nhanh nhất, phù hợp cho ứng dụng cần phản hồi nhanh.

Làm sạch dữ liệu

Tổng quan

Phân tích

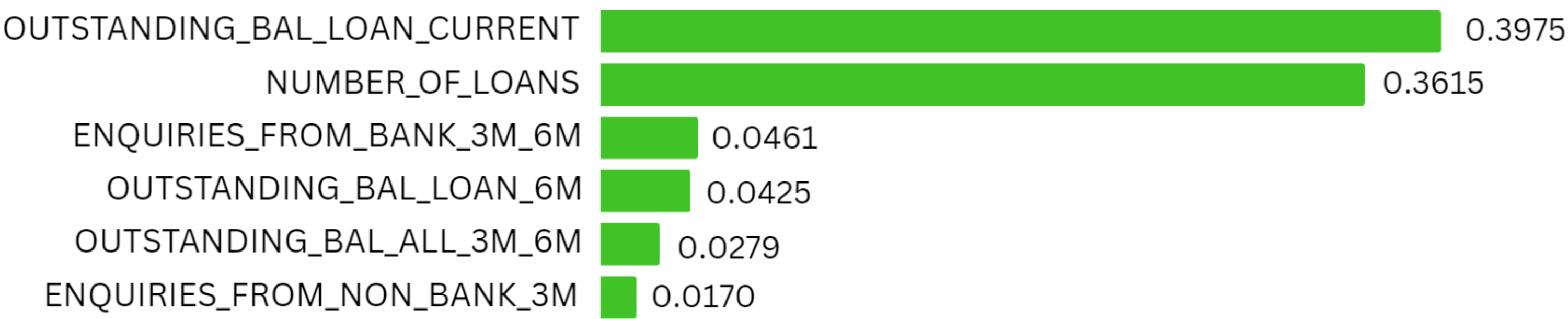
Xây dựng mô hình

Kết luận & Đề xuất

# Quyết Định Chọn Mô Hình AdaBoost Classifier

Với ý tưởng về mô hình có thể áp dụng tốt trong thực tiễn, nhóm đã quyết định lựa chọn mô hình AdaBoost Classifier, với thời gian suy đoán nhanh, độ chính xác cao và ít nhầm lẫn.

## Biểu đồ tầm quan trọng của các biến



Số dư nợ vay ở thời điểm hiện tại và số khoản vay có tác động lớn tới việc khách hàng trả nợ đúng hạn, theo sau là biến số lượt tra cứu tín dụng trong 3-6 tháng và số dư nợ vay 6 tháng trước có mức tác động trung bình.

Biến chênh lệch số dư nợ trong khoảng 3-6 tháng và số truy vấn từ các tổ chức phi ngân hàng có tác động thấp tới hành vi của khách hàng. Các biến còn lại có trọng số bằng 0, có thể được loại bỏ.



# Kết luận

Khách hàng có **xu hướng vay nhiều hơn** trong thời gian gần đây, bao gồm cả vay nợ và vay tín dụng. Đồng thời, khách hàng cũng thể hiện sự **quan tâm tới các khoản vay** của mình.



Tuy nhiên, phân tích cho thấy **rủi ro về khả năng trả nợ** của khách hàng trong tương lai đối với các khoản nợ của mình. Kết quả chạy mô hình cũng cho thấy rằng, khách hàng vay càng nhiều, số dư nợ càng tăng thì càng **có thể không trả nợ đúng hạn**.

# Đề xuất giải pháp

## Về tín dụng

**Giảm hạn mức tín dụng** đối với những khách hàng sở hữu nhiều thẻ tín dụng và có số dư nợ tín dụng cao.

**Tăng lãi suất tín dụng** đối với những khách hàng được dự báo có nguy cơ không trả nợ đúng hạn, đặc biệt là mức lãi suất khi không thanh toán dư nợ đầy đủ đúng hạn, buộc khách hàng phải thanh toán dư nợ đúng hạn.

## Về các khoản vay kỳ hạn

**Tăng tỷ lệ tài sản bảo đảm** đối với những khách hàng được dự báo khả năng trả nợ thấp.

**Tăng lãi suất cho vay**, đặc biệt là lãi suất của những khoản vay ngắn hạn (dưới 1 năm) nhằm giảm nhu cầu vay từ nhóm khách hàng có rủi ro cao.

Bên cạnh đó, ngân hàng cùng các tổ chức tín dụng cần tăng cường các hoạt động **tư vấn tài chính cá nhân** cho khách hàng nhằm giúp khách hàng vay và sử dụng các khoản vay hợp lý, đồng thời giảm thiểu rủi ro nợ xấu từ khách hàng.