

DỰ ĐOÁN RỦI RO CỦA DOANH NGHIỆP TỪ CÁC KHOẢN VAY TRONG TƯƠNG LAI

GVHD: Đinh Thủy Tiên

NHÓM: aa



Danh sách thành viên



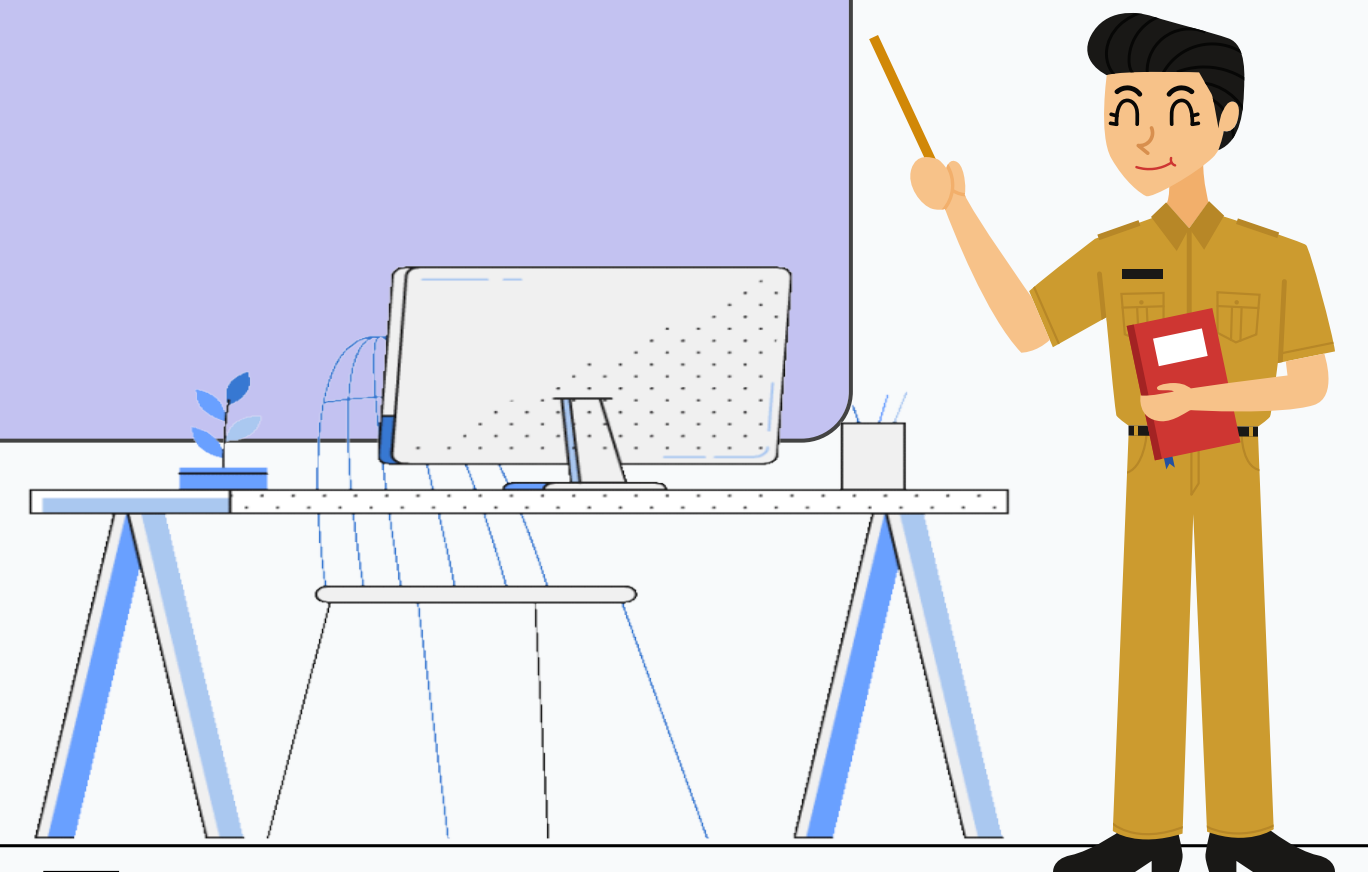
A47996_ Vũ Thành Đạt

A48443_ Đào_Trung_Hiếu



LENDING CLUB

Lending Club là một nền tảng cho vay ngang hàng (peer-to-peer lending) tại Mỹ, cho phép các cá nhân vay tiền trực tiếp từ các nhà đầu tư cá nhân, bỏ qua trung gian tài chính truyền thống như ngân hàng



CÁCH HOẠT ĐỘNG CỦA CÔNG TY

BÊN VAY

Khách hàng có thể đăng ký khoản vay từ 1.000 USD đến 40.000 USD. Hồ sơ sẽ được thẩm định nhanh chóng và giải ngân ngay sau khi được phê duyệt. Khoản vay được hoàn trả trong vòng ba năm với mức lãi suất cố định, mang đến sự an tâm và chủ động trong kế hoạch tài chính của Quý khách.

NHÀ ĐẦU TƯ

Nhà đầu tư có thể lựa chọn khoản vay phù hợp dựa trên thông tin người vay, đồng thời nhận lãi suất hấp dẫn từ khoản đầu tư của mình..

Lending Club

Chúng tôi thu phí khởi tạo từ người vay và phí dịch vụ từ nhà đầu tư, đồng thời chịu trách nhiệm quản lý các khoản vay và phân phối khoản thanh toán đến nhà đầu tư một cách minh bạch và hiệu quả.

BUSINESS CASE

MỤC ĐÍCH

Phân loại khoản vay dựa trên hành vi của khách hàng:

- Định giá rủi ro và lãi suất phù hợp
- Giảm thiểu rủi ro tín dụng
- Hỗ trợ ra quyết định chiến lược

GIẢ ĐỊNH

- Dữ liệu được sử dụng là đáng tin cậy
- Chi phí xử lý cho mỗi claim là bằng nhau
- Thời gian đánh giá cho mỗi claim là tương đương
- Khách hàng có đầy đủ các chỉ số cần thiết theo yêu cầu của mô hình

DỮ LIỆU

- Accepted_2007_to_2018Q4.csv

MVP

- Mô hình dự đoán khách hàng có khả năng trả được nợ hay không
- Xét duyệt mô hình

ĐỐI TƯỢNG

- Các tổ chức tín dụng

YẾU TỐ THÀNH CÔNG

- Sensitivity cao: 84%
- Precision cao: 91%
- Dễ giải thích và dễ thực thi

TÁC ĐỘNG

- Giúp tiết kiệm lên tới 50%- 70% thời gian đánh giá khoản vay
- Có thể tiết kiệm từ 30-60% chi phí xử lý khoản vay

MỤC TIÊU KHAI PHÁ DỮ LIỆU



PHÊ DUYỆT CÁC KHOẢN VAY

- Để tránh mất khách hàng tiềm năng .
- Giúp giảm thiểu rủi ro .
- Phân nhóm doanh nghiệp theo mức độ rủi ro .
- Dự đoán khả năng trả nợ hoặc vỡ nợ của từng doanh nghiệp .



Recall: 88%

Precision: 96%

LÝ DO CHỌN CÁC CHỈ SỐ

1. Precision

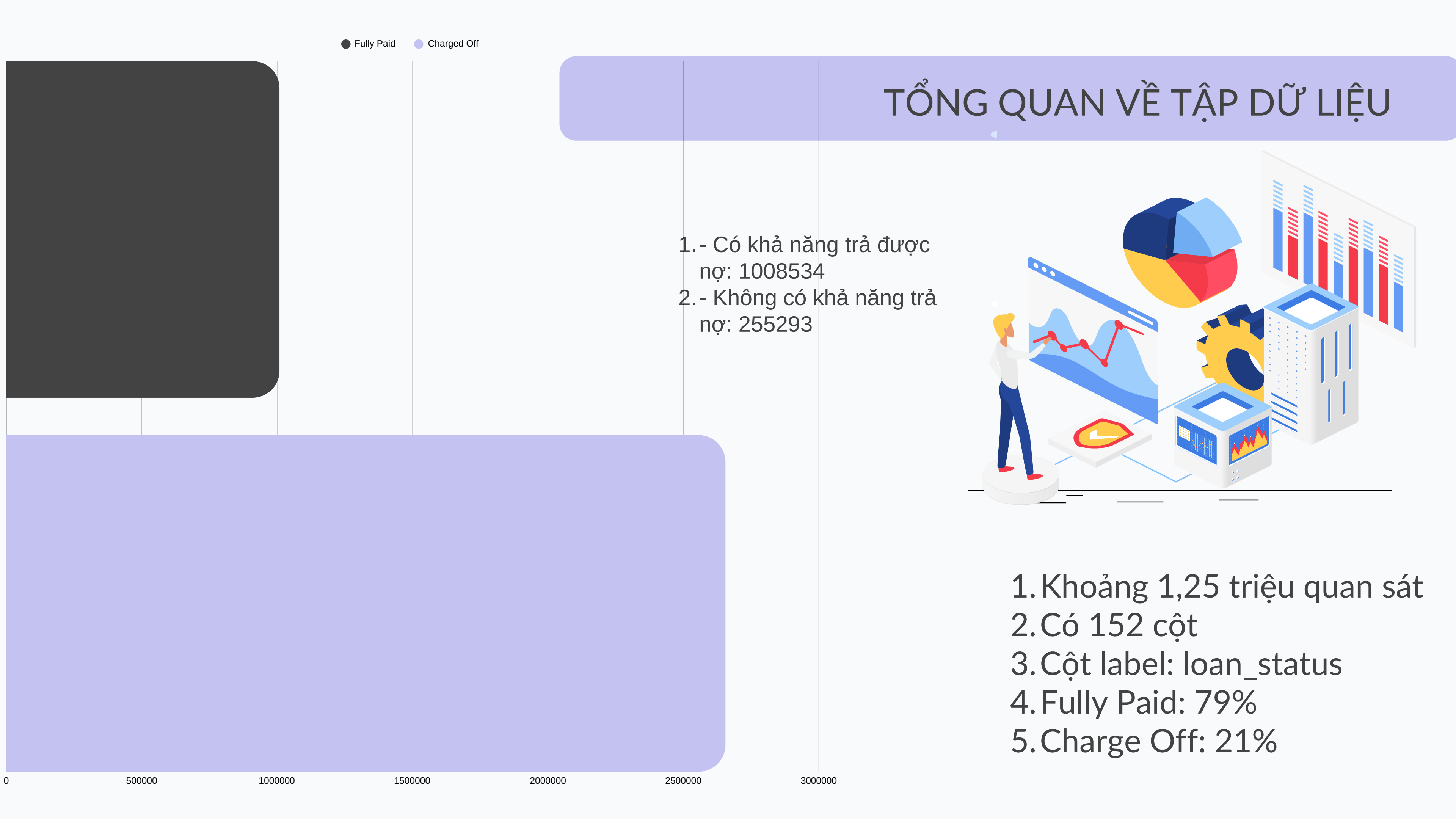
2. Khi bạn muốn giảm thiểu số lượng dự đoán sai dương (False Positives) (Bad Loan).
3. Phát hiện hiện khoản vay khó có khả năng thanh khoản, bạn muốn đảm bảo rằng các khoản vay được gán là 'Bad Loan' thực sự là 'Bad Loan', để tránh giảm thiểu rủi ro.
4. Precision cho "Bad Loan" là 0.96,

1. Recall

2. Recall cho "Good Loan": Để đảm bảo rằng các khoản vay tốt không bị từ chối một cách bất công.
3. Recall cho "Good Loan" là 0.88

TIỀN XỬ LÝ DỮ LIỆU





Fully Paid Charged Off

TỔNG QUAN VỀ TẬP DỮ LIỆU

- 1.- Có khả năng trả được nợ: 1008534
- 2.- Không có khả năng trả nợ: 255293



- 1. Khoảng 1,25 triệu quan sát
- 2. Có 152 cột
- 3. Cột label: loan_status
- 4. Fully Paid: 79%
- 5. Charge Off: 21%

0 500000 1000000 1500000 2000000 2500000 3000000

Tại sao lại là Fully
Paid và Charged-
Off





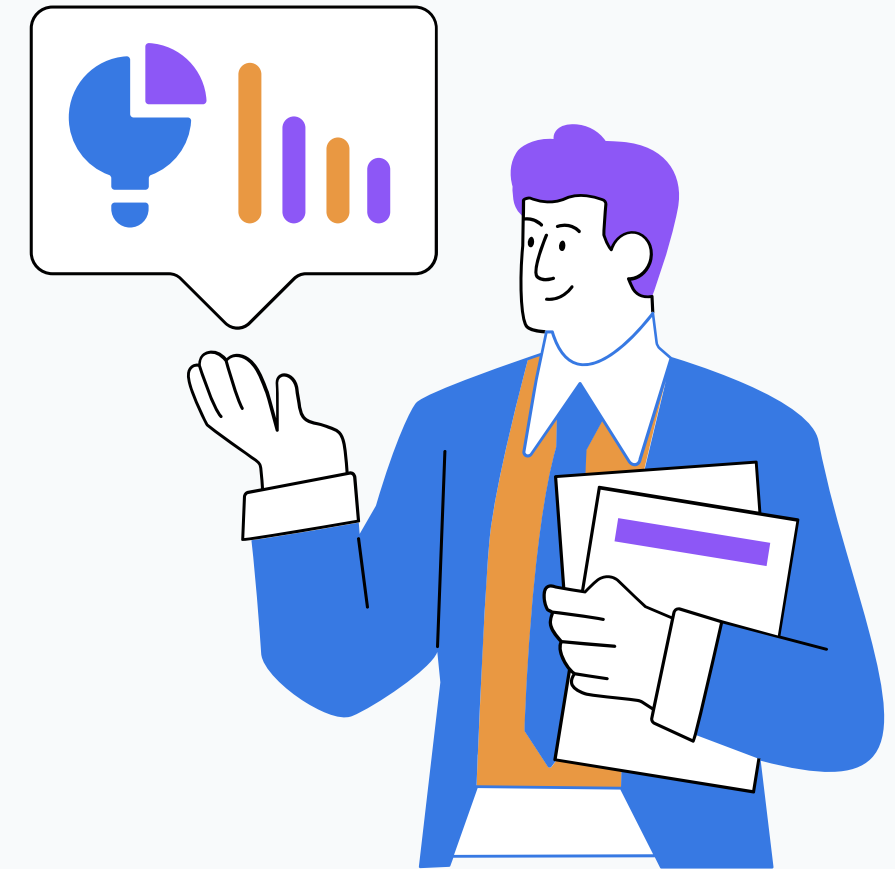
Sự rõ ràng

"Fully Paid: Khoản vay đã được thanh toán đầy đủ (cả gốc lẫn lãi).
Charge Off: Khoản vay được khoanh nợ, tức là khách hàng không trả được và tổ chức cho vay xác nhận khoản này là tổn thất.



Giảm sự mơ hồ

Cột Label "loan_status" có nhiều trạng thái khác nhau như "Current", "Late" hay "In Grace Period" không phản ánh kết quả cuối cùng của khoản vay. Chúng gây mơ hồ cho mô hình vì một khoản vay "Current" hôm nay có thể trở thành "Charged Off" trong tương lai.



Đơn giản hóa bài toán và cải thiện hiệu suất

Việc chỉ giữ lại các khoản vay có trạng thái "Fully Paid" và "Charged Off" giúp bài toán trở nên đơn giản hơn, loại bỏ sự mơ hồ trong dữ liệu.

Train Data 80%	Test Data 20%
Có khả năng trả nợ	Có khả năng trả nợ
Không có khả năng trả nợ	Không có khả năng trả nợ



TÁCH TẬP DỮ LIỆU:

Tập huấn luyện: 80%
Tập kiểm tra: 20%

Balanced Train Data	Test Data
Có khả năng trả được nợ	Có khả năng trả được nợ
Không có khả năng trả được nợ	Không có khả năng trả được nợ



CÂN BẰNG DỮ LIỆU

Sử dụng phương pháp Under Sampling.

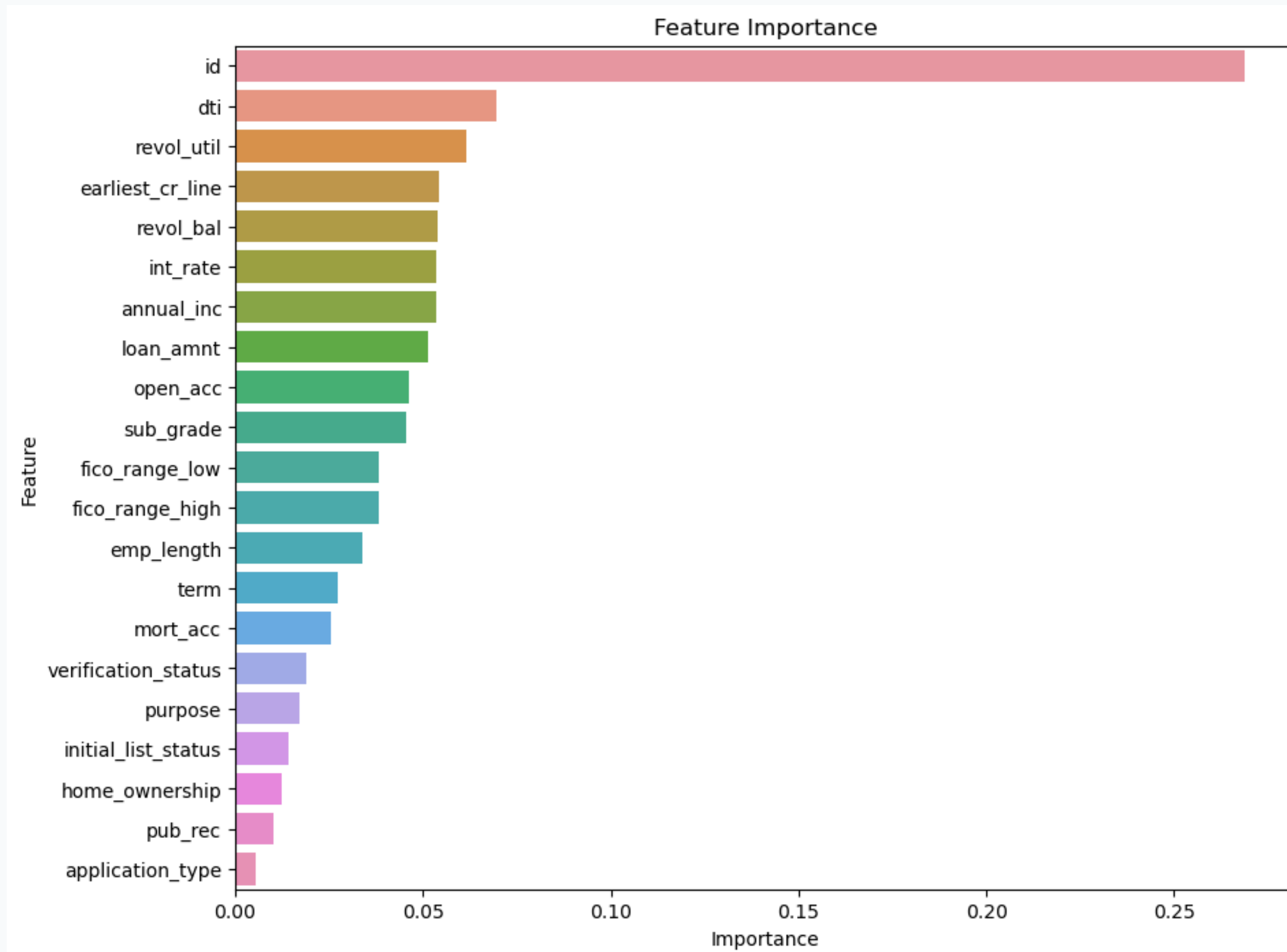
Khoản vay trả đầy đủ: 50%

Khoản nợ: 50%

TIỀN XỬ LÝ DỮ LIỆU

- Lọc dữ liệu theo mục tiêu ,chỉ giữ lại các khoản vay có trạng thái rõ ràng, loại bỏ các khoản vay còn đang hoạt động hoặc chưa xác định kết quả.
- Xem xét các đặc điểm có ý nghĩa thống kê (sử dụng Rừng ngẫu nhiên) để đánh giá tầm quan trọng của các đặc trưng.
- EDA các cột để xem xét các đặc điểm quan trọng để xác định sự khác biệt trong phân phối theo biến mục tiêu và đề xuất các giả thuyết về ảnh hưởng của từng đặc trưng đối với khả năng rủi ro tín dụng.

MỨC ĐỘ ƯU TIÊN



CHỌN CỘT DỮ LIỆU

- **loan_amnt**: Số tiền vay ban đầu mà khách hàng yêu cầu.
- **term**: Thời hạn của khoản vay, thường là 36 tháng hoặc 60 tháng.
- **int_rate**: Lãi suất hàng năm (%) áp dụng cho khoản vay.
- **num_op_rev_tl**: Số lượng tài khoản tín dụng quay vòng đang mở (thường là thẻ tín dụng).
- **sub_grade**: Hạng tín dụng chi tiết của khách hàng (ví dụ: A1 đến G5), đánh giá rủi ro tín dụng.
- **home_ownership**: Tình trạng sở hữu nhà của khách hàng (OWN, RENT, MORTGAGE, etc.).
- **annual_inc**: Thu nhập hàng năm của khách hàng.
- **purpose**: Mục đích sử dụng khoản vay (ví dụ: debt_consolidation, credit_card, home_improvement, etc.).
- **bc_open_to_buy**: Số tiền khả dụng để chi tiêu qua thẻ tín dụng (hạn mức còn lại).
- **Region**: Khu vực cư trú của khách hàng.
- **dti**: Tỷ lệ nợ hàng tháng trên tổng thu nhập hàng tháng (Debt-to-Income ratio).
- **delinq_2yrs**: Số lần trễ hạn thanh toán trong vòng 2 năm gần nhất.
- **earliest_cr_line**: Ngày mở tài khoản tín dụng đầu tiên của khách hàng.
- **fico_score**: Điểm tín dụng FICO, phản ánh độ uy tín tài chính (300–850).
- **pub_rec**: Số lần xuất hiện trong hồ sơ công khai (ví dụ: phá sản, kiện tụng).
- **revol_bal**: Số dư nợ quay vòng hiện tại (ví dụ: nợ thẻ tín dụng chưa trả).
- **mort_acc**: Số lượng tài khoản liên quan đến bất động sản (thế chấp).

LỰA CHỌN MÔ HÌNH

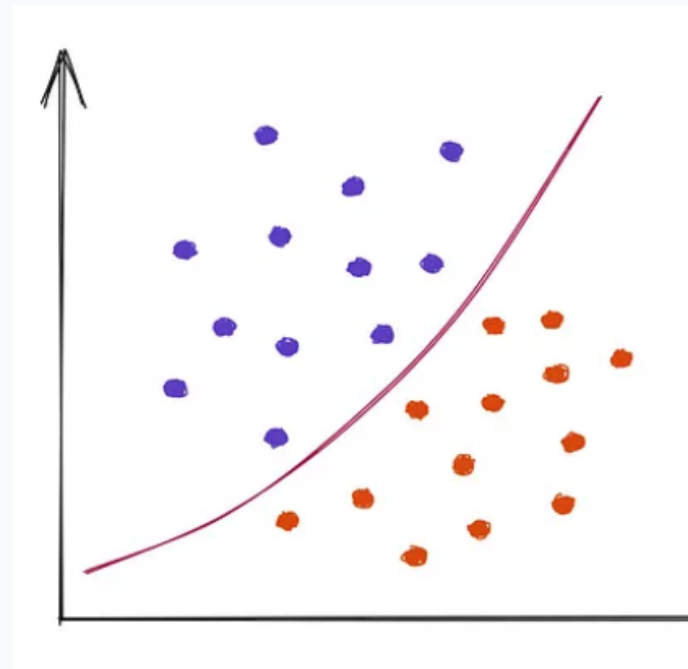
	KNN	LG	RF
Đơn giản – Dễ hiểu	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Hợp với bài toán	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Dễ triển khai	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Tuân thủ quy định	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

CÁC BIẾN CỦA MÔ HÌNH VÀ KẾT QUẢ MÔ HÌNH

BIẾN ĐẦU VÀO

'loan_amnt', 'term',
'int_rate',
'num_op_rev_tl',
'sub_grade',
'home_ownership',
'annual_inc', 'purpose',
'bc_open_to_buy',
'region', 'dti',
'delinq_2yrs',
'earliest_cr_line',
'fico_score', 'pub_rec',
'revol_bal', 'mort_acc'

Logistic MODEL



ĐẦU RA MÔ HÌNH

- Khoản vay tốt (Fully Paid)
- Khoản vay xấu (Charge Off)

BUSINESS VALUE

Không dùng mô hình Số quan sát

Tổng số claims	10,000
Số claims không được cho vay	2000
Sai số nếu làm thủ công	1000

Tổng sai số 1000

Khi áp dụng mô hình Số quan sát

Số dự đoán đúng	9668
Số claims mà mô hình phân loại không được cho vay	7737
Mô hình phân loại sai	260

Tổng sai số 260

- Giúp tiết kiệm lên tới 50%-70% thời gian đánh giá
- Có thể tiết kiệm từ 30-60% chi phí xử lý khoản vay

Cảm ơn cô và các
bạn đã lắng nghe



THE END !!!