

1. Vấn đề bài toán



Giới thiệu

Tỷ lệ khách hàng rời bỏ ngân hàng (churn) là một trong những vấn đề ảnh hưởng trực tiếp đến doanh thu và chi phí duy trì khách hàng. Việc khách hàng rời đi thường dẫn đến mất nguồn thu lâu dài và tốn thêm chi phí để thu hút khách hàng mới.

Mục tiêu

Ngân hàng cần hiểu rõ lý do tại sao khách hàng rời đi để có chiến lược giữ chân hiệu quả và tối ưu hóa trải nghiệm khách hàng hiện tại.

Xây dựng mô hình dự đoán khả năng khách hàng sẽ rời bỏ ngân hàng dựa trên các đặc điểm như độ tuổi, thu nhập, thời gian gắn bó, số lượng sản phẩm sử dụng, v.v. Từ đó, vẽ được chân dung khách hàng rời bỏ và hỗ trợ doanh nghiệp có thể can thiệp sớm với các khách hàng có nguy cơ rời bỏ cao.

2. Tổng quan dữ liệu

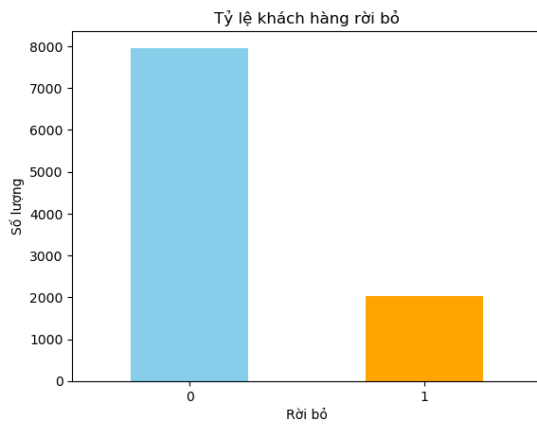
Nguồn dữ liệu: Tập dữ liệu công khai từ Kaggle – Churn Modelling Dataset.

Số lượng quan sát: 10,000 khách hàng từ một ngân hàng ở Châu Âu.

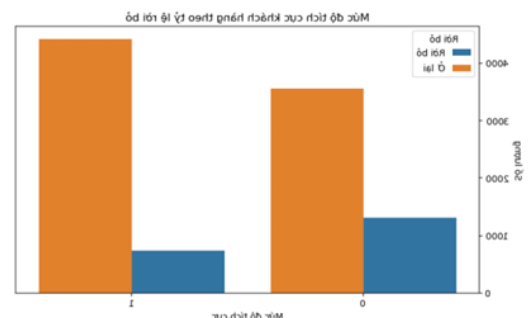
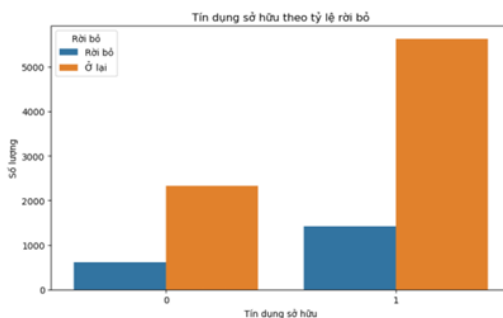
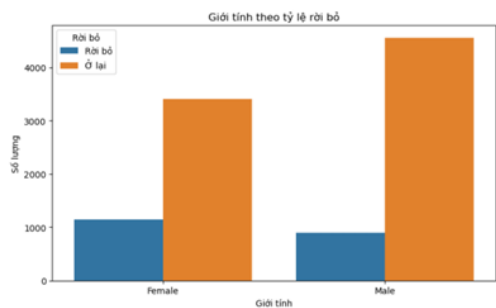
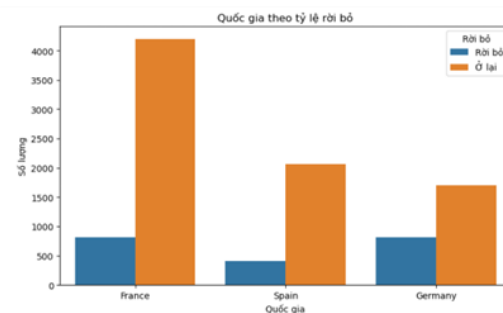
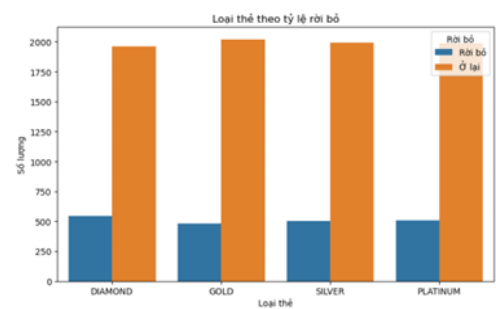
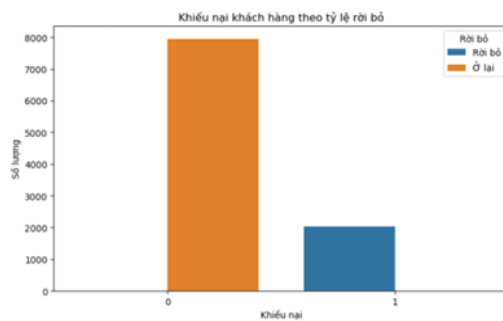
Biến mục tiêu: Exited (1: rời bỏ, 0: vẫn là khách hàng).

Một số biến đầu vào chính: Age, Gender, Geography, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, CreditScore

EDA - Exploratory Data Analysis



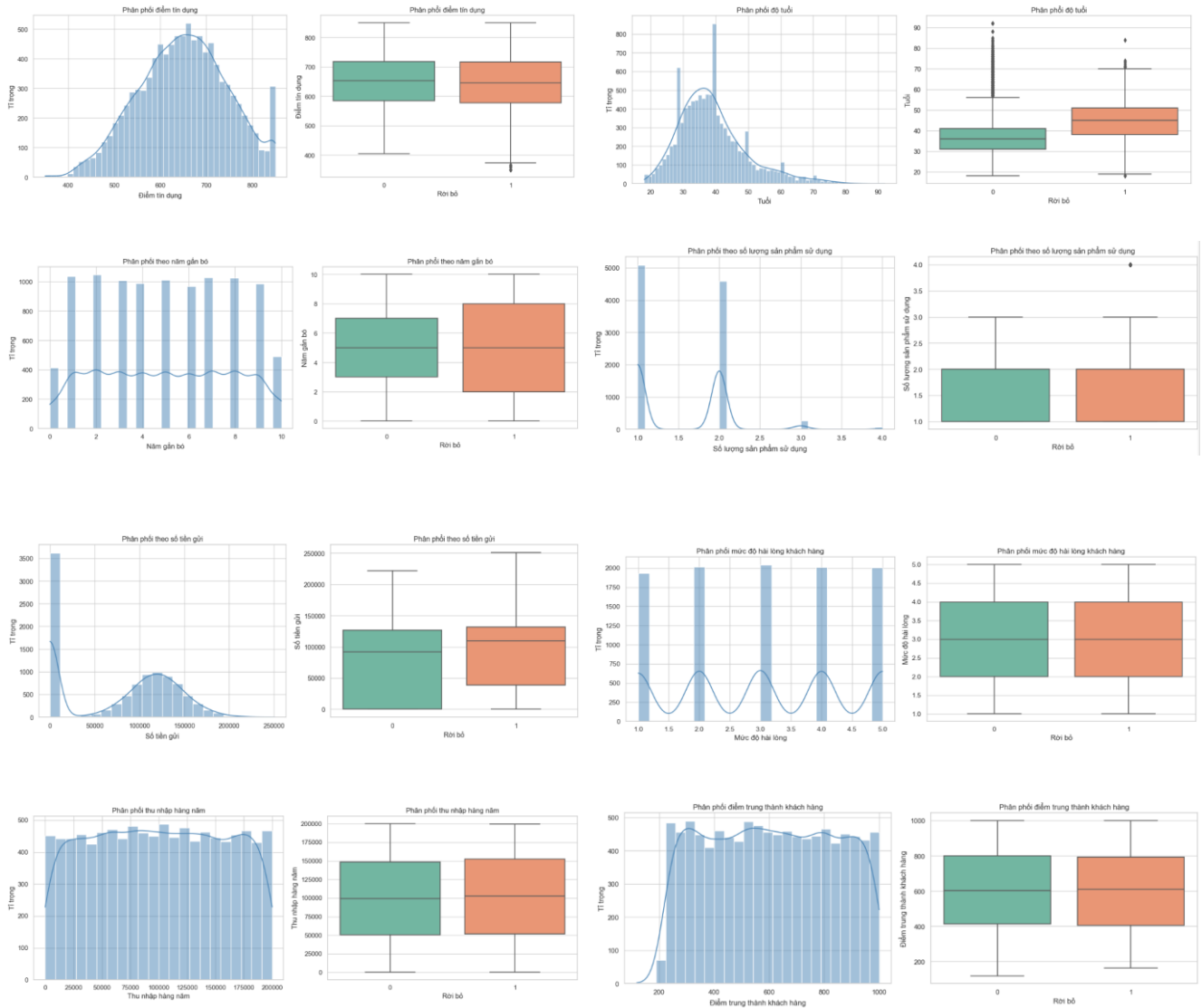
Có sự mất cân bằng dữ liệu rõ rệt: chỉ khoảng 20% khách hàng rời bỏ (label = 1), còn lại 80% vẫn tiếp tục sử dụng dịch vụ (label = 0). => sử dụng các kỹ thuật xử lý mất cân bằng dữ liệu (như SMOTE, class weight, undersampling...) khi xây dựng mô hình dự đoán



Khách hàng có khiếu nại, ít hoạt động, không sở hữu thẻ tín dụng, và đến từ Germany là những nhóm có tỷ lệ rời bỏ cao.

Tình trạng hoạt động tài khoản là yếu tố phân biệt rõ ràng nhất giữa nhóm churn và không churn → đây có thể là feature quan trọng hàng đầu.

Sự khác biệt theo loại thẻ và giới tính không quá lớn, nhưng vẫn cần đưa vào kiểm tra vì có thể tương tác với các biến khác.

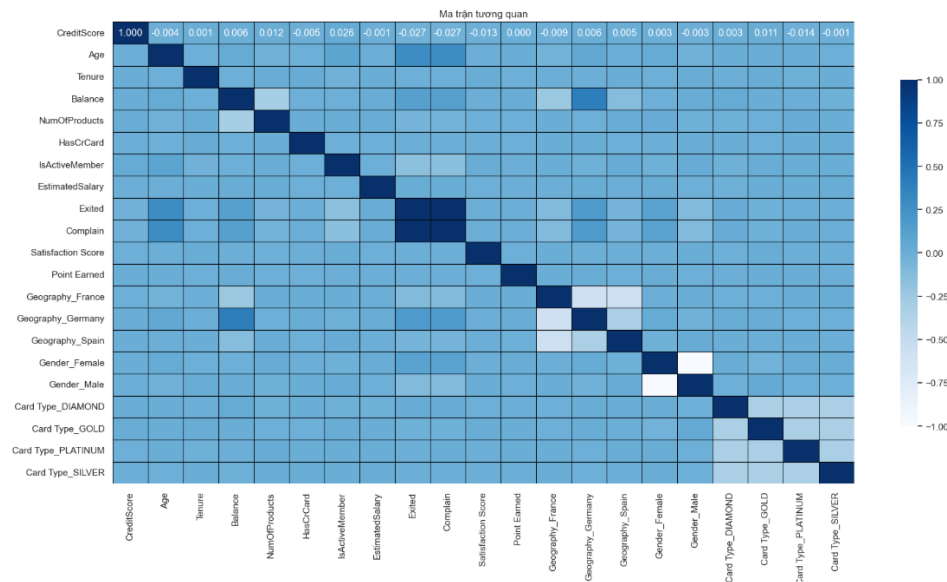


- Điểm tín dụng: Không khác biệt rõ giữa nhóm rời bỏ và ở lại.
- Tuổi: Khách rời bỏ thường lớn tuổi hơn.
- Số dư tài khoản: Nhóm rời bỏ thường có số dư cao hơn.
- Số sản phẩm sử dụng: Dùng 1 sản phẩm → tỷ lệ rời bỏ cao.
- Thời gian gần bỏ: Không ảnh hưởng rõ rệt đến churn.
- Hoạt động tài khoản: Rời bỏ nhiều ở nhóm hoạt động thấp.
- Thu nhập ước lượng: Phân bố đều, ít ảnh hưởng.

Tổng giá trị giao dịch/điểm hoạt động: Nhóm churn có điểm hoạt động thấp hơn

Tuổi, số dư, số lượng sản phẩm sử dụng, và mức độ hoạt động là những đặc điểm liên quan mật thiết đến churn.

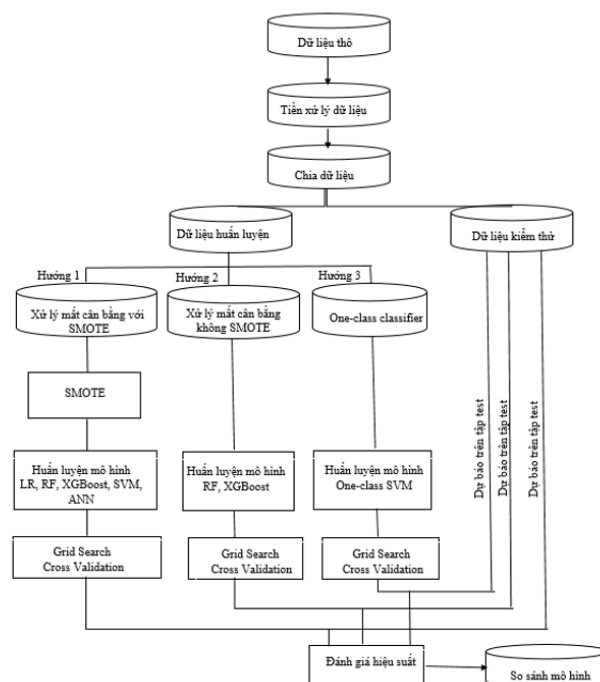
Một số biến như thu nhập, credit score, tenure không có sự khác biệt đáng kể → có thể ít quan trọng hơn trong mô hình hóa



Biến đầu vào xây dựng mô hình: Age, Balance, NumOfProducts, IsActiveMember, Complaint, Point Earned, Geography_Germany → có tương quan tốt với Exited.

CreditScore, EstimatedSalary, Tenure → tương quan rất thấp với Exited, nhưng vẫn có thể giữ để kiểm tra tương tác.

⇒ **Đề xuất hướng triển khai**



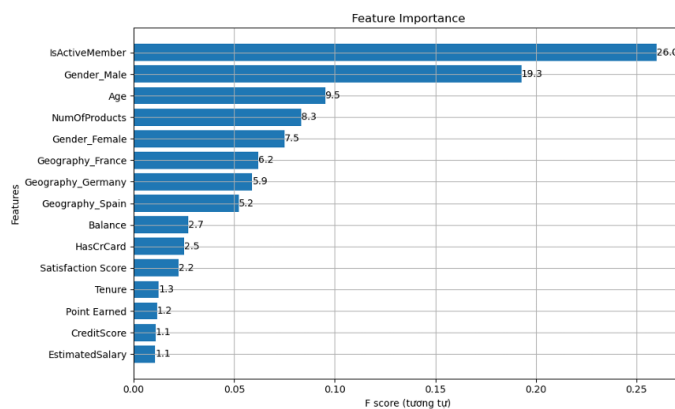
3. Tóm tắt kết quả

Hướng triển khai	Mô tả	Mô hình	ROC AUC	Accuracy	Precision	Recall	F1
Hướng 1	SMOTE	ANN	0.5	0.2	0.2	1	0.34
		RF	0.85	0.84	0.84	0.84	0.85
		XGBoost	0.86	0.86	0.86	0.86	0.86
		SVM	0.82	0.83	0.61	0.53	0.56
		LR	0.67	0.77	0.67	0.7	0.71
Hướng 2	Resample weight	XGBoost	0.83	0.83	0.58	0.59	0.58
		RF	0.85	0.86	0.83	0.39	0.53
Hướng 3	One - Class	SVM	0.51	0.74	0.22	0.11	0.14

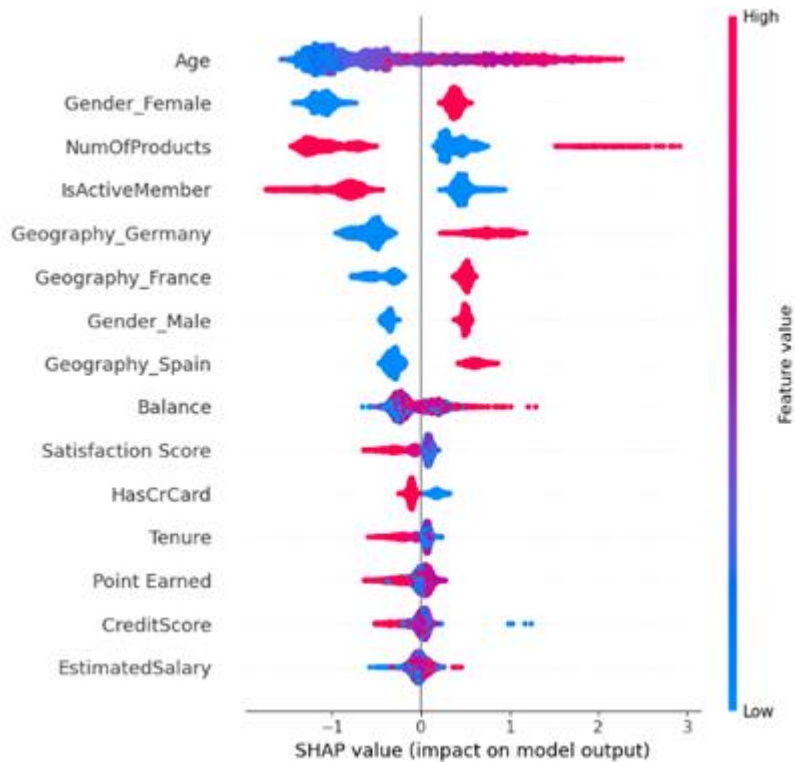
Hướng 1 - SMOTE => XGBoost và Random Forest (RF) cho hiệu suất cao nhất: ROC AUC ~0.85–0.86, F1 ~0.85–0.86 — là lựa chọn tốt nhất.

Hướng 2 - Resample weight => XGBoost và RF giữ được ROC AUC khá (~0.83–0.85) nhưng Recall giảm mạnh, đặc biệt với RF giảm khả năng nhận diện khách hàng churn.

Hướng 3 - One-Class SVM => kết quả kém, không phù hợp với bài toán này.



Mô hình ưu tiên các yếu tố hành vi và nhân khẩu học, đặc biệt là mức độ tương tác với ngân hàng (IsActiveMember), hơn là thông tin tài chính



Tuổi cao, sử dụng ít sản phẩm, và không còn hoạt động ($IsActiveMember = 0$) là những yếu tố tăng mạnh khả năng churn.

Khách hàng đến từ Germany, có satisfaction thấp và thu nhập thấp cũng dễ rời bỏ hơn.

Một số biến như EstimatedSalary, CreditScore, Point Earned có ít ảnh hưởng đến quyết định churn.

4. Đề xuất

Chân dung khách hàng rời bỏ

- Tuổi cao (trên 50)
- Không hoạt động trong thời gian gần đây
- Chỉ sử dụng 1 sản phẩm của ngân hàng
- Đến từ Germany
- Điểm hài lòng thấp
- Không sở hữu thẻ tín dụng
- Thu nhập và điểm tín dụng ở mức thấp

Đề xuất các giải pháp hỗ trợ

Tăng mức độ tương tác với khách hàng: Khách hàng rời bỏ thường không còn hoạt động, vì vậy cần duy trì liên lạc và thúc đẩy họ sử dụng dịch vụ thường xuyên hơn.

Đa dạng hóa và cá nhân hóa sản phẩm: Khách hàng chỉ dùng ít sản phẩm để rời đi, do đó nên khuyến khích họ trải nghiệm thêm các dịch vụ phù hợp với nhu cầu

Cải thiện trải nghiệm khách hàng: Khách hàng có mức độ hài lòng thấp hoặc từng khiếu nại dễ rời bỏ, ngân hàng cần rút ngắn thời gian phản hồi và nâng chất lượng phục vụ

Thiết kế chính sách phù hợp theo từng nhóm khách hàng: Khách lớn tuổi hoặc thuộc khu vực có churn cao cần được chăm sóc bằng các chính sách riêng biệt và thân thiện hơn.

Theo dõi và dự báo hành vi rời bỏ: Ứng dụng mô hình AI để phát hiện sớm khách hàng có rủi ro churn và đưa ra biện pháp can thiệp kịp thời.