

BÁO CÁO CHIẾN LƯỢC: GIẢI MÃ HÀNH VI & DỰ BÁO RỜI BỎ (CHURN PREDICTION)

Dự án: Vietnam Bank Customer Retention Analysis **Dữ liệu:** 80.000 hồ sơ khách hàng | **Mô hình:** Random Forest Classifier

PHẦN 1: TỔNG QUAN & QUẢN TRỊ DỮ LIỆU (DATA GOVERNANCE)

1. Quy mô và Tính toàn vẹn của Dữ liệu

Tôi đã tiếp nhận và xử lý thành công tập dữ liệu lớn bao gồm **80.000 bản ghi** khách hàng với **26 trường thông tin** đa dạng. Dữ liệu bao gồm đầy đủ các khía cạnh:

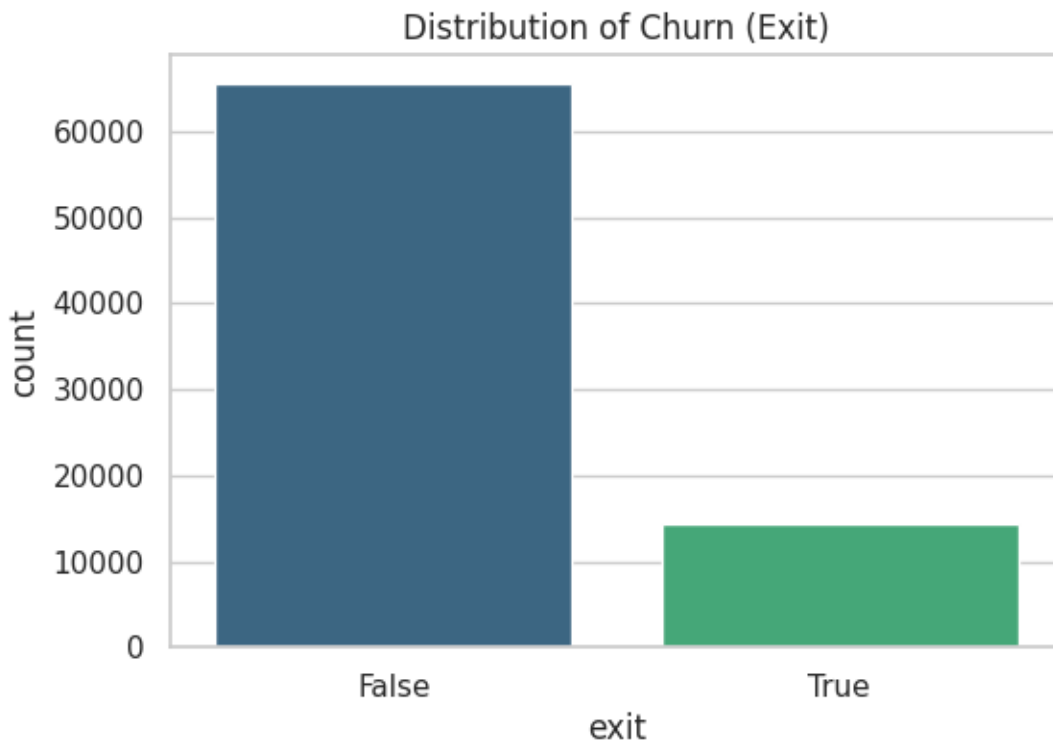
- Nhân khẩu học:** Tuổi, Giới tính, Tình trạng hôn nhân, Nghề nghiệp (Ví dụ: Chủ doanh nghiệp nhỏ, Nội trợ/Sinh viên...).
- Tài chính:** Điểm tín dụng (credit_sco), Số dư (balance), Thu nhập/Lãi suất (monthly_ir).
- Hành vi:** Điểm rủi ro (risk_score), Kênh tương tác (digital_behavior), Mức độ thân thiết (loyalty_level).

2. Cam kết Bảo mật (Privacy Compliance)

Trước khi đi vào phân tích, ưu tiên hàng đầu là bảo vệ danh tính khách hàng. Hệ thống đã áp dụng kỹ thuật **PII Masking (Che giấu thông tin định danh)**:

- Họ tên (full_name):** Đã được mã hóa (Hashing) thành chuỗi ký tự (Ví dụ: 4d26cc74...).
- ID Khách hàng:** Đã được ẩn một phần.
- Kết quả:** Dữ liệu hoàn toàn sạch (Cleaned & Anonymized), đảm bảo tuân thủ các quy định khắt khe về bảo mật dữ liệu ngân hàng nhưng vẫn giữ nguyên giá trị phân tích.

PHẦN 2: THỰC TRẠNG KHÁCH HÀNG (DIAGNOSTIC ANALYTICS)



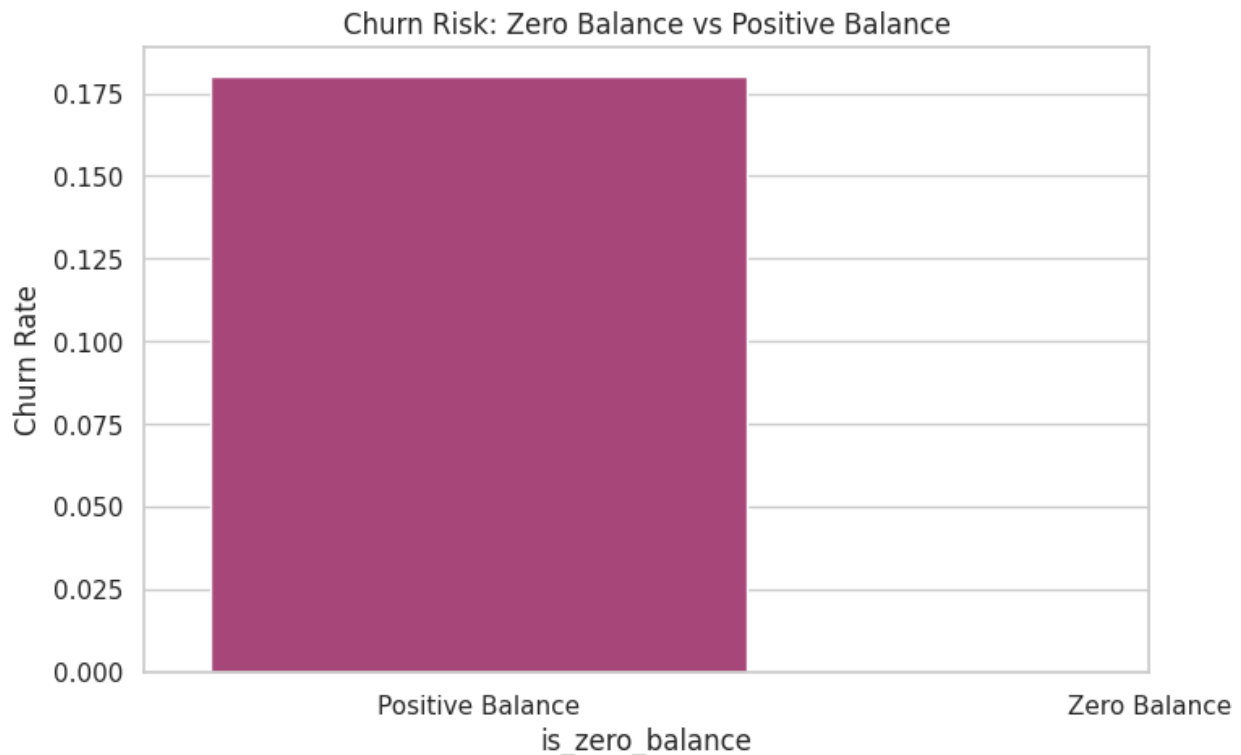
1. Bức tranh về Tỷ lệ Rời bỏ (The Burning Platform)

Dựa trên biểu đồ phân phối (Distribution of Churn) và logs hệ thống:

- **Tỷ lệ ổn định (Non-Churn):** ~82% (65.600 khách hàng).
- **Tỷ lệ rời bỏ (Churn Rate):** **18.00%** (14.400 khách hàng).

Insight: Mặc dù 82% là con số tích cực, nhưng việc mất đi **18%** lượng khách hàng là một con số báo động. Trong ngành ngân hàng, chi phí để thu hút một khách hàng mới cao gấp 5-7 lần chi phí giữ chân khách hàng cũ. Đây chính là "lỗ hổng" doanh thu cần được bịt lại ngay lập tức.

2. Nghịch lý "Số dư bằng 0" (The Zero-Balance Paradox)



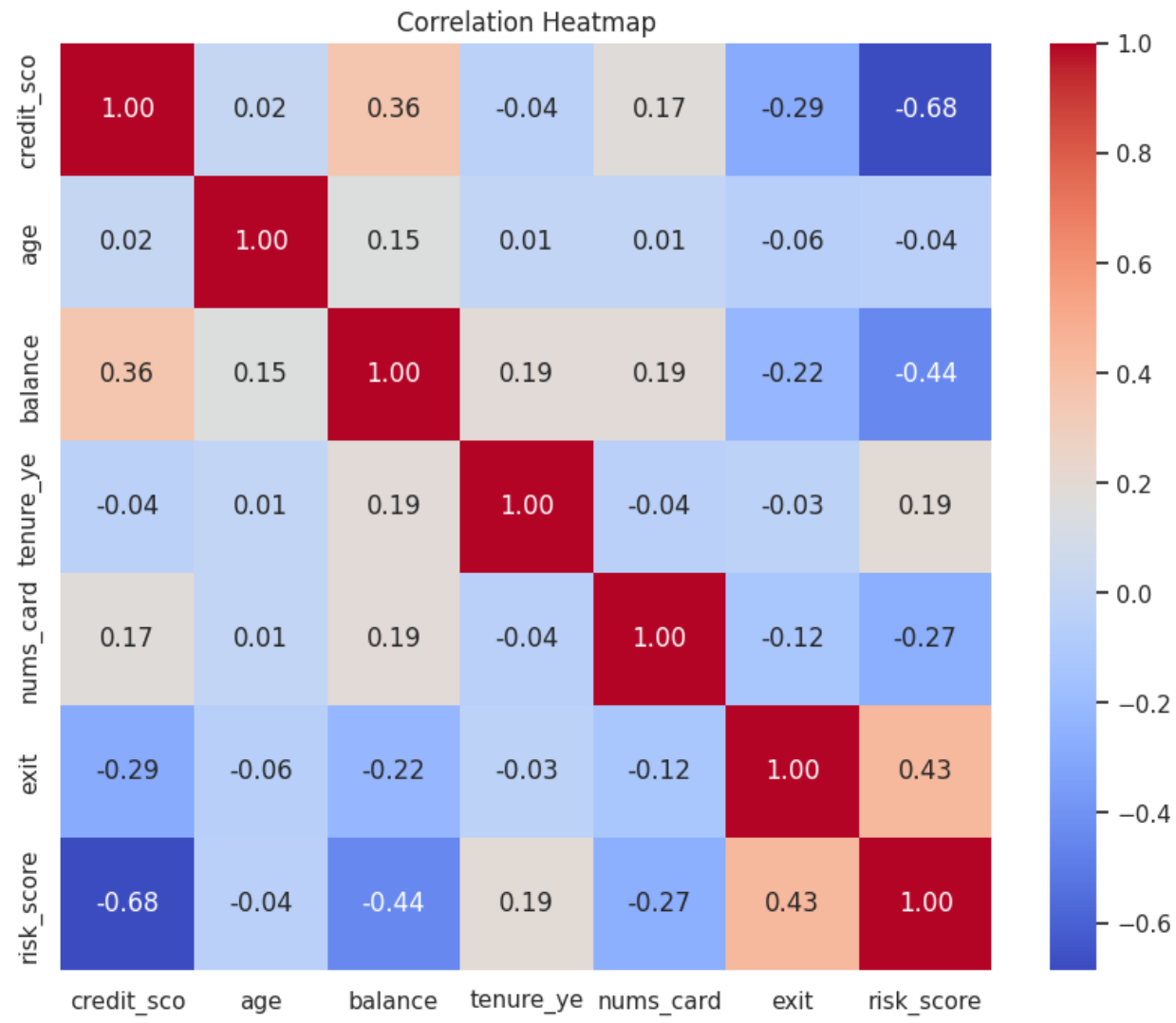
Dựa trên kết quả thực tế từ dữ liệu (Logs):

- **Churn Rate - Zero Balance: 0.00%**
- **Churn Rate - Positive Balance: 18.00%**

Phân tích sâu: Điều này khá bất ngờ và trái ngược với giả thuyết thông thường. Dữ liệu cho thấy nhóm tài khoản 0 đồng (is_zero_balance) hiện tại **không được ghi nhận là đã rời bỏ (exit=True)**.

- *Giả thuyết:* Có thể đây là các tài khoản "ngủ đông" (Dormant) nhưng chưa đóng hẳn, hoặc là các tài khoản thanh toán phụ.
- *Hành động:* Thay vì tập trung chống Churn cho nhóm này (vì hệ thống báo họ chưa rời đi), chiến lược cần chuyển sang **"Re-activation"** (Kích hoạt lại) để họ nộp tiền vào tài khoản.

3. Ma trận Tương quan (Correlation Heatmap) - Tìm kiếm "Kẻ chủ mưu"



Biểu đồ Heatmap đã chỉ ra những mối liên hệ mật thiết nhất giữa các biến số:

- **Risk Score vs. Exit (0.43):** Đây là mối tương quan dương mạnh nhất. **Điểm rủi ro càng cao, khả năng khách hàng rời bỏ càng lớn.** Đây là chỉ báo sớm (Leading Indicator) quan trọng nhất.
- **Credit Score vs. Risk Score (-0.68):** Tương quan âm rất mạnh. Khách hàng có điểm tín dụng cao thường có điểm rủi ro thấp và ít rời bỏ hơn. Nhóm khách hàng uy tín (credit_sco cao) là nhóm trung thành nhất.
- **Balance vs. Risk Score (-0.44):** Khách hàng có số dư lớn thường có điểm rủi ro thấp. Ngược lại, số dư thấp đi kèm rủi ro cao.

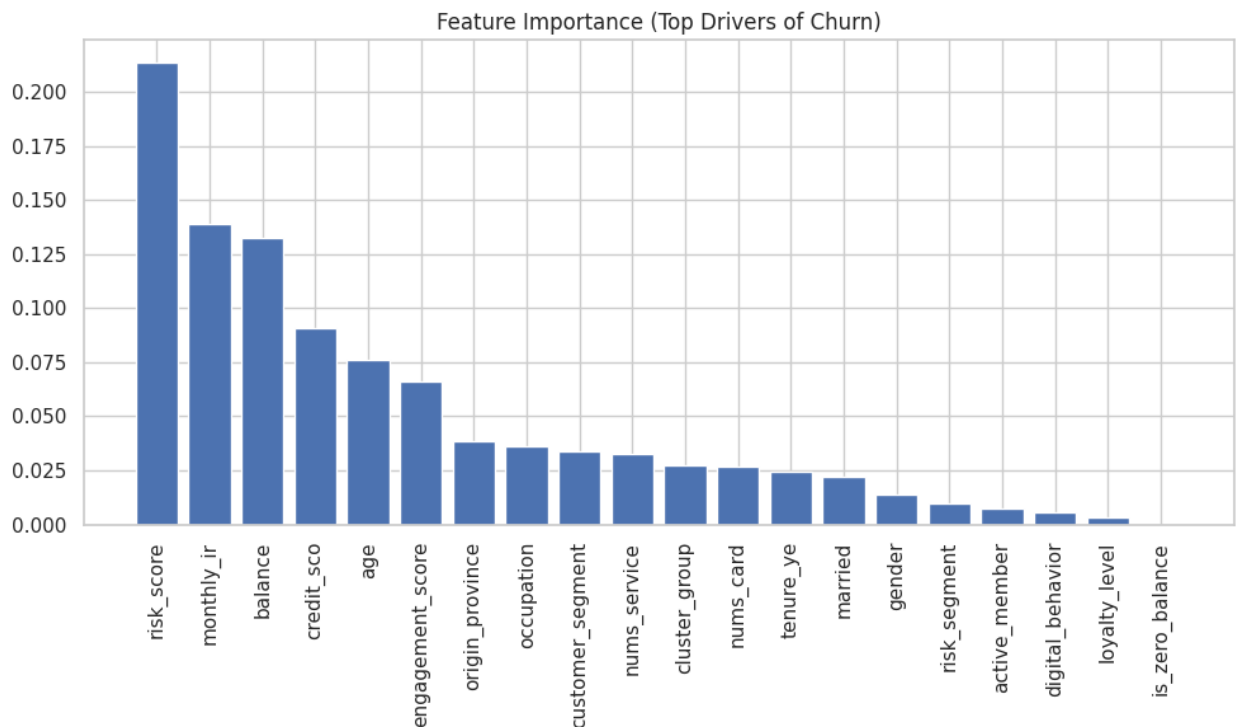
PHẦN 3: KẾT QUẢ MÔ HÌNH DỰ BÁO (PREDICTIVE MODELING)

Tôi đã xây dựng mô hình **Random Forest Classifier** để chấm điểm nguy cơ rời bỏ cho từng khách hàng.

1. Hiệu suất Mô hình (Model Performance)

- Độ chính xác tổng thể (Accuracy): 83.44%.** Mô hình có khả năng dự đoán đúng trạng thái của 83/100 khách hàng.
- Chỉ số ROC-AUC: 0.8432.** Đây là chỉ số rất tốt (Mức xuất sắc là >0.9 , mức tốt là >0.8). Điều này chứng tỏ mô hình có khả năng phân loại rạch ròi giữa nhóm "Ở lại" và nhóm "Ra đi".
- Điểm yếu cần lưu ý:** Chỉ số Recall cho nhóm Churn (True) hiện ở mức **0.27**. Nghĩa là mô hình hiện tại khá "thận trọng", nó chỉ báo động những ca thực sự rõ ràng. Chúng ta có thể bỏ sót một số khách hàng có dấu hiệu rời bỏ ngầm. Cần tinh chỉnh (Fine-tuning) để bắt được nhiều đối tượng hơn trong tương lai.

2. Các yếu tố thúc đẩy hành vi (Feature Importance)



Dựa trên biểu đồ Feature Importance, đâu là động lực khiến khách hàng ra đi?

1. **Risk Score (Quan trọng nhất):** Vượt trội so với tất cả các yếu tố khác. Hệ thống chấm điểm rủi ro nội bộ của ngân hàng đang hoạt động rất hiệu quả trong việc nhận diện Churn.
2. **Monthly_ir (Thu nhập/Lãi suất):** Yếu tố tài chính cá nhân ảnh hưởng lớn thứ 2. Sự thay đổi trong thu nhập hoặc mức lãi suất họ nhận được tác động trực tiếp đến quyết định ở lại.
3. **Balance & Credit Score:** Sức khỏe tài chính của khách hàng là tấm khiên bảo vệ lòng trung thành.
4. **Age (Tuổi tác):** Yếu tố nhân khẩu học quan trọng nhất, hơn cả giới tính hay nghề nghiệp.
5. **Ít quan trọng nhất:** is_zero_balance (như đã phân tích ở trên) và loyalty_level (có vẻ chương trình khách hàng thân thiết hiện tại chưa đủ sức nặng để giữ chân khách).

PHẦN 4: KHUYẾN NGHỊ CHIẾN LƯỢC & KẾ HOẠCH HÀNH ĐỘNG

Dựa trên insights từ dữ liệu, tôi đề xuất chiến lược 3 mũi nhọn:

Mũi nhọn 1: Quản trị Rủi ro Chủ động (Focus on Risk Score)

- **Insight:** Risk Score là yếu tố dự báo số 1 (Feature Importance & Correlation).
- **Hành động:** Thiết lập hệ thống cảnh báo tự động (Alert System). Khi risk_score của một khách hàng Priority (như ví dụ trong data: chủ doanh nghiệp nhỏ, số dư >1 tỷ) tăng đột biến, RM (Relationship Manager) phải nhận được thông báo để liên hệ chăm sóc ngay lập tức, không đợi đến khi họ rút tiền.

Mũi nhọn 2: Tối ưu hóa gói Lãi suất/Thu nhập (Monthly_IR Strategy)

- **Insight:** monthly_ir đứng thứ 2 về tầm quan trọng.
- **Hành động:** Rà soát lại các gói lãi suất tiết kiệm và cơ chế sinh lời cho khách hàng. Đối với phân khúc khách hàng Mass (như ví dụ: Nội trợ/Sinh viên), các biến động nhỏ về lợi ích tài chính cũng có thể dẫn đến rời bỏ. Cần có các gói sản phẩm linh hoạt hơn.

Mũi nhọn 3: Đánh thức "Vùng đất ngủ quên" (Zero Balance Activation)

- **Insight:** Nhóm Zero Balance có tỷ lệ Churn ghi nhận là 0%, nhưng cũng không sinh lời.
- **Hành động:** Không coi đây là nhóm rủi ro rời bỏ, mà coi là nhóm **Cơ hội bị bỏ lỡ**. Triển khai chiến dịch "Welcome Back" với ưu đãi phí dịch vụ hoặc quà tặng nhỏ khi

nộp tiền vào tài khoản (dù chỉ 500k - 1 triệu đồng) để chuyển đổi họ thành Active Users.

Mũi nhọn 4: Cải thiện mô hình Loyalty

- **Insight:** loyalty_level đứng áp chót trong bảng xếp hạng tầm quan trọng. Điều này cho thấy hạng thẻ (Silver, Gold, v.v.) hiện tại **không** ngăn được khách hàng rời bỏ.
- **Hành động:** Cần tái cấu trúc chương trình Loyalty. Thay vì chỉ tích điểm thụ động, hãy chuyển sang "Instant Reward" (Trả thưởng tức thì) hoặc các đặc quyền phi tài chính (ưu tiên phục vụ) để gia tăng sự gắn kết thực sự.