

VAI TRÒ CỦA DỮ LIỆU HÀNH VI TRONG DỰ BÁO KHÁCH HÀNG RỜI BỎ DỊCH VỤ THẺ TÍN DỤNG: TIẾP CẬN TỪ GÓC ĐỘ PHÂN TÍCH DỮ LIỆU

NGUYỄN HẢI ANH^{1*} - NGUYỄN HOÀNG THAO¹

¹Ngân hàng TMCP Ngoại thương Việt Nam

Joint Stock Commercial Bank for Foreign Trade of Vietnam

* Tác giả liên hệ - Email: ahnguyen1511@gmail.com

Ngày nhận bài: 22/7/2026

Ngày sửa bài: 10/8/2026

Ngày duyệt đăng bài: 25/8/2026

Tóm tắt:

Bài nghiên cứu "Vai trò của dữ liệu hành vi trong dự báo khách hàng rời bỏ dịch vụ thẻ tín dụng: Tiếp cận từ góc độ phân tích dữ liệu" đánh giá khả năng dự báo của dữ liệu hành vi (behavioral data) đối với hiện tượng khách hàng rời bỏ (churn) trong lĩnh vực thẻ tín dụng. Nghiên cứu xem xét các yếu tố hành vi như thói quen chi tiêu và thói quen thanh toán, cùng mối tương quan của các yếu tố này với hiện tượng khách hàng rời bỏ. Trên cơ sở tiếp cận toàn diện về phân tích dữ liệu (data analytics) tham chiếu tới các khung lý thuyết về hành vi rời bỏ, tác giả xây dựng mô hình dự báo nhằm nhận diện chính xác nhóm khách hàng có nguy cơ rời bỏ trong tương lai. Kết quả nghiên cứu cung cấp những thông tin có giá trị giúp các ngân hàng bán lẻ hoàn thiện chiến lược giữ chân khách hàng. Nghiên cứu đồng thời đóng góp vào lĩnh vực phân tích dữ liệu nói chung thông qua việc làm rõ tiềm năng của dữ liệu hành vi trong mô hình hóa dự báo.

Từ khóa: dữ liệu hành vi, khách hàng rời bỏ, thẻ tín dụng, phân tích dữ liệu, ngân hàng bán lẻ

The role of behavioral data in predicting credit card customer churn: A data analytics approach

Abstract:

This study evaluates the predictive power of behavioral data for customer churn in the credit card sector. It examines behavioral factors such as spending and payment habits and their relationships with churn. Drawing on a comprehensive data analytics approach and theoretical frameworks of churn behavior, the authors develop a predictive model to identify customers at risk of leaving in the future. The findings provide useful information for retail banks seeking to improve customer retention strategies and also contribute to the broader field of data analytics by demonstrating the potential of behavioral data in predictive modeling.

Keywords: behavioral data, customer churn, credit card, data analytics, retail banking

1. Đặt vấn đề

Trong thị trường thẻ tín dụng có mức độ cạnh tranh cao, hiện tượng khách hàng rời bỏ dịch vụ tạo ra tổn thất kép cho ngân hàng: mất nguồn thu hiện hữu và phát sinh chi phí thu hút khách hàng mới thay thế. Theo Quinlan và cộng sự (2024), chi phí thu hút một khách hàng thẻ tín dụng mới tại thị trường Hồng Kông có thể lên tới 2.100 HKD (khoảng 250 USD), trong khi thời gian hòa vốn của một hợp đồng thẻ mới kéo dài tới 16 tháng. Reichheld và Sasser (1990) cùng Gupta và Lehmann (2003) đồng thời chỉ ra chi phí duy trì khách hàng hiện hữu thấp hơn đáng kể so với chi phí thu hút khách hàng mới, do đó việc dự báo và ngăn ngừa rời bỏ trở thành mục tiêu kinh doanh trọng yếu của các ngân hàng bán lẻ.

Các mô hình dự báo truyền thống chủ yếu dựa trên biến nhân khẩu học và các thông tin tĩnh, tần suất thay đổi thấp (tuổi, thu nhập, thâm niên quan hệ), do đó khó nắm bắt các tín hiệu suy giảm gắn kết xuất hiện trước thời điểm khách hàng rời bỏ. Dữ liệu hành vi - bao gồm tần suất và giá trị giao dịch, số dư tài khoản, thói quen thanh toán, mức độ sử dụng hạn mức - cung cấp nguồn thông tin động, cho phép giám sát liên tục và phát hiện sớm rủi ro rời bỏ (Hadden và cộng sự, 2007; Buckinx và Van den Poel, 2005).

Trên cơ sở đó, bài nghiên cứu trả lời câu hỏi trung tâm: dữ liệu hành vi có thể được sử dụng để dự báo khách hàng rời bỏ trong lĩnh vực thẻ tín dụng hay không, và nếu có, mức độ chính xác đạt được ở mức nào. Nghiên cứu sử dụng dữ liệu ẩn danh để xây dựng, kiểm định mô hình dự báo và đề xuất khung ứng dụng kết quả mô hình vào chiến lược giữ chân khách hàng. Phần còn lại của bài viết được sắp xếp như sau: Phần 2 trình bày cơ sở lý thuyết và tổng quan nghiên cứu; Phần 3 mô tả dữ liệu và phương pháp nghiên cứu; Phần 4 trình bày kết quả và ứng dụng phân tích đề xuất; Phần 5 kết luận toàn bài.

2. Cơ sở lý thuyết và tổng quan nghiên cứu

2.1. Khái niệm và phân loại khách hàng rời bỏ

Khách hàng rời bỏ (customer churn) được hiểu là quá trình khách hàng chấm dứt quan hệ với nhà cung cấp dịch vụ hoặc suy giảm mức độ gắn kết một cách rõ rệt (Hadden và cộng sự, 2007). Lazarov và Capota (2007) phân loại rời bỏ thành hai nhóm: rời bỏ tự nguyện (voluntary churn) - khách hàng chủ động chấm dứt quan hệ do không hài lòng hoặc do lựa chọn thay thế hấp dẫn hơn, trong đó tách biệt rời bỏ có chủ đích (deliberate) và rời bỏ tình cờ (incidental); và rời bỏ không tự nguyện (involuntary churn) - quan hệ bị chấm dứt bởi quyết định của ngân hàng, ví dụ do khách hàng vi phạm điều khoản hợp đồng. Việc phân biệt các loại hình rời bỏ có ý nghĩa trực tiếp đối với thiết kế biến phụ thuộc của mô hình dự báo và lựa chọn chiến lược can thiệp tương ứng.

2.2. Phân tích dữ liệu trong quản trị khách hàng rời bỏ

Tổng hợp nghiên cứu từ Davenport (2006), Shirazi (2019) và Lepenioti (2020), bốn phương pháp phân tích dữ liệu (data analytics) được sử dụng gồm: (i) phân tích mô tả (descriptive) tổng hợp hành vi lịch sử và phân khúc danh mục khách hàng; (ii) phân tích chẩn đoán (diagnostic) tìm nguyên nhân gốc của hiện tượng rời bỏ; (iii) phân tích dự báo (predictive) ước lượng xác suất rời bỏ của từng khách hàng bằng các kỹ thuật thống kê và học máy; (iv) phân tích đề xuất (prescriptive) chuyển hóa kết quả dự báo thành khuyến nghị hành động cụ thể về phân bổ nguồn lực giữ chân khách hàng. Nghiên cứu này vận dụng đồng thời cả 4 phương pháp, trong đó trọng tâm đặt vào phương pháp phân tích dự báo và phân tích đề xuất.

2.3. Tổng quan nghiên cứu thực nghiệm và khoảng trống nghiên cứu

Các nghiên cứu thực nghiệm về dự báo rời bỏ đã phát triển từ mô hình dựa trên biến nhân khẩu học sang mô hình tích hợp dữ liệu hành vi và kỹ thuật phân tích nâng cao. Buckinx và Van den Poel (2005) chứng minh biến hành vi giao dịch nâng cao đáng kể độ chính xác dự báo trong các ngành tiêu dùng nhanh Bán lẻ (Retail FMCG); Neslin và cộng sự (2006) hệ thống hóa các phương pháp đo lường độ chính xác của mô hình dự báo rời bỏ; Lemmens và Croux (2006) áp dụng kỹ thuật bagging và boosting

cây phân loại; Verbeke và cộng sự (2011) nhấn mạnh yêu cầu về tính diễn giải được (comprehensibility) của mô hình trong môi trường ngân hàng chịu giám sát chặt chẽ; Amin và cộng sự (2019) xử lý vấn đề độ tin cậy dữ liệu trong dự báo rời bỏ ngành viễn thông.

Tổng quan từ các nghiên cứu thực nghiệm về dự báo khách hàng rời bỏ cho thấy ba khoảng trống chính: (i) mức độ tích hợp dữ liệu hành vi vào mô hình dự báo còn hạn chế; (ii) phần lớn nghiên cứu dừng ở khâu xây dựng mô hình, chưa gắn kết quả dự báo với khung hành động marketing cụ thể; (iii) yêu cầu về tính minh bạch của mô hình trong ngành tài chính chưa được đáp ứng đầy đủ khi nhiều nghiên cứu sử dụng thuật toán dạng "hộp đen". Nghiên cứu này xử lý cả ba khoảng trống thông qua việc xây dựng mô hình thể điểm logistic có tính diễn giải cao trên dữ liệu thực tế của ngân hàng Việt Nam và đề xuất khung phân khúc hành động dựa trên kết quả mô hình.

3. Dữ liệu và phương pháp nghiên cứu

3.1. Dữ liệu

Dữ liệu nghiên cứu được ẩn danh hóa, bao phủ giai đoạn từ tháng 01/2014 đến tháng 12/2018 trong phạm vi khách hàng cá nhân sử dụng thẻ tín dụng. Bộ dữ liệu gồm bốn nhóm thông tin: (i) nhân khẩu học (tuổi, giới tính, tình trạng hôn nhân, trình độ học vấn, thu nhập, nghề nghiệp); (ii) thông tin tài khoản (thâm niên quan hệ, hạng thẻ, hạn mức tín dụng, số dư); (iii) dữ liệu hành vi (tần suất và giá trị giao dịch, loại giao dịch, kênh giao dịch, mức độ sử dụng hạn mức); (iv) dữ liệu rủi ro tín dụng (lịch sử thanh toán, số lần quá hạn).

Quy trình tiền xử lý dữ liệu gồm 4 bước: loại bỏ bản ghi trùng lặp; chuẩn hóa định dạng dữ liệu không nhất quán; xử lý giá trị bị khuyến thông tin bằng phương pháp thay thế theo trung bình/trung vị đối với biến định lượng và phân nhóm riêng "không có thông tin" đối với biến định tính; nhận diện và xử lý giá trị ngoại lai bằng khoảng tứ phân vị (Interquartile Range - IQR) và điểm z (z-score), áp dụng biện pháp chặn trên (capping) hoặc loại bỏ tùy bản chất quan sát.

3.2. Định nghĩa biến phụ thuộc và thiết kế mẫu

Biến phụ thuộc nhận giá trị 1 (rời bỏ) khi khách hàng thỏa mãn ít nhất một trong ba tiêu chí sau trong vòng 12 tháng kể từ thời điểm quan sát: (i) đóng toàn bộ tài khoản thẻ tín dụng; (ii) không phát sinh bất kỳ giao dịch nào trong 06 tháng liên tục; (iii) không gia hạn thẻ đã hết hạn. Khách hàng không thỏa mãn tiêu chí nào nêu trên được gán giá trị 0 (không rời bỏ).

Mẫu nghiên cứu được thiết lập bằng việc lấy các quan sát tại tháng 6 và tháng 12 hằng năm trong giai đoạn 2014 - 2018; tại mỗi thời điểm quan sát, các biến độc lập được tính trên cửa sổ quan sát 06 tháng liên trước và trạng thái rời bỏ được xác định trong 12 tháng liên sau.

Mẫu cuối cùng gồm 7.889 quan sát, được phân tầng (stratified) theo tỷ lệ 70/30 thành mẫu phát triển và mẫu kiểm thử với tỷ lệ rời bỏ đồng nhất 34% ở cả 2 mẫu, bảo đảm 2 mẫu có tính chất tương đồng về tỷ lệ rời bỏ. Mô hình sẽ được xây dựng trên mẫu phát triển và được kiểm tra hiệu năng trên mẫu kiểm thử sau đó để đánh giá khả năng quá khớp của mô hình.

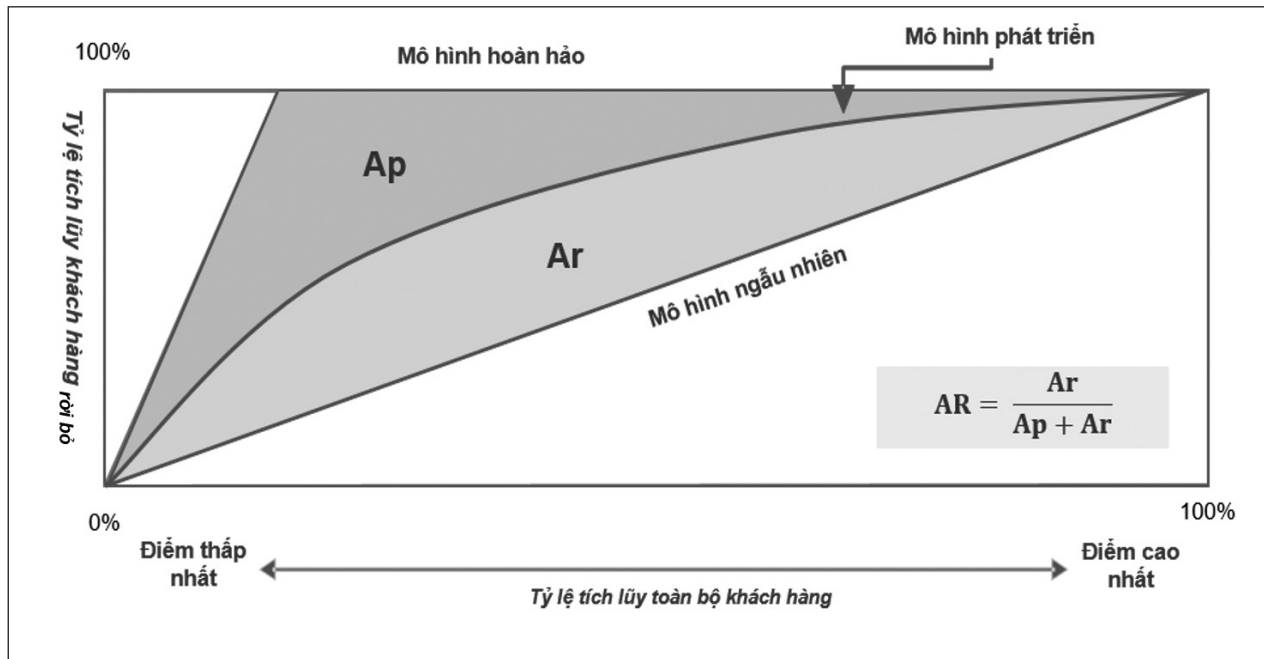
3.3. Phân tích đơn biến (Single Factor Analysis - SFA)

Từ danh sách dài (long-list) gồm 93 chỉ tiêu tiềm năng, nhóm nghiên cứu thực hiện phân tích đơn biến nhằm sàng lọc ra danh sách ngắn (short-list) theo 5 tiêu chí định tính: khả năng phân biệt khách hàng rời bỏ, tính khả thi trong thu thập dữ liệu, tính trực quan về nghiệp vụ, mức độ quen thuộc với đơn vị kinh doanh và tính phù hợp với điều kiện hiện tại của ngân hàng. Về mặt định lượng, sức mạnh dự báo của từng chỉ tiêu được đo bằng hệ số tỷ lệ chính xác (Accuracy Ratio - AR) đối với biến liên tục và giá trị thông tin (Information Value - IV) đối với biến phân nhóm.

Hệ số AR được xác định theo diện tích giữa đường cong phân biệt của chỉ tiêu và mô hình ngẫu nhiên:

$$AR = A_r / (A_p + A_r)$$

Hình 1. Minh họa công thức tính hệ số AR



Nguồn: Tác giả tự tổng hợp

trong đó A_r là diện tích giữa đường cong của chỉ tiêu đang xét và mô hình ngẫu nhiên; A_p là diện tích giữa đường cong của mô hình hoàn hảo và chỉ tiêu đang xét. Giá trị thông tin IV được tính như sau:

$$IV = \sum_{i=1}^n (\% \text{ Không rời bỏ}_i - \% \text{ Rời bỏ}_i) \times WOE_i$$

với

$$WOE_i = \ln\left(\frac{nc_i / NC}{c_i / NC}\right)$$

trong đó nc_i , c_i là số khách hàng không rời bỏ và rời bỏ thuộc nhóm i ; NC , C là tổng số khách hàng không rời bỏ và rời bỏ của toàn mẫu. Ngưỡng tham chiếu IV: dưới 0,02 - không phù hợp đưa vào mô hình; [0,02; 0,1) - dự báo yếu; [0,1; 0,3) - dự báo trung bình; [0,3; 0,5] - dự báo mạnh; trên 0,5 - quá mạnh, cần thận trọng khi sử dụng. Bên cạnh đó, tương quan cặp giữa các chỉ tiêu được kiểm soát bằng hệ số Pearson và Spearman với ngưỡng tối đa 50% nhằm hạn chế đa cộng tuyến.

Sau khi lập danh sách ngắn, từng chỉ tiêu được phân nhóm (bucketing) để giảm ảnh hưởng của giá trị ngoại lai.

3.4. Phân tích đa nhân tố và lựa chọn mô hình

Phương pháp hồi quy logistic được lựa chọn cho phân tích đa nhân tố do ba ưu điểm: phù hợp với biến phụ thuộc nhị phân; cho phép ước lượng trực tiếp xác suất rời bỏ của từng khách hàng; và có tính diễn giải cao thông qua tỷ số odds - yêu cầu thiết yếu đối với mô hình vận hành trong môi trường ngân hàng (Verbeke và cộng sự, 2011). Kết quả mô hình được trình bày dưới dạng thẻ điểm (scorecard) với thang điểm 0 - 100, tạo thuận lợi cho việc triển khai tại các đơn vị kinh doanh. Với danh sách ngắn 12 chỉ tiêu tạo được các tổ hợp mô hình khác nhau. Mô hình cuối cùng được lựa chọn có AR cao nhất, toàn bộ chỉ tiêu trong mô hình có ý nghĩa thống kê (với p-value của từng chỉ tiêu không lớn hơn 0,1) và có xu hướng dữ liệu phù hợp với ý nghĩa nghiệp vụ. Điều này bảo đảm mô hình có hiệu năng dự báo tốt, có ý nghĩa thực tiễn cao, sát với tình hình kinh doanh.

4. Kết quả nghiên cứu

4.1. Kết quả phân tích đơn biến

Kết quả sàng lọc trên 93 chỉ tiêu của danh sách dài cho thấy toàn bộ các biến nhân khẩu học như giới tính, tình trạng hôn nhân, trình độ học vấn, tuổi của khách hàng, thu nhập tại thời điểm mở Thẻ, đều có hiệu năng dự báo đơn biến thấp hơn ngưỡng lựa chọn ($IV \geq 0,1$ hoặc $AR \geq 10\%$). Ngược lại, 12 chỉ tiêu được đưa vào danh sách ngắn đều thuộc nhóm dữ liệu tài khoản và dữ liệu hành vi (Bảng 1), trong đó chỉ tiêu có sức mạnh dự báo cao nhất - tổng số giao dịch thành công trong 03 tháng gần nhất - đạt AR 48,53%. Kết quả này cung cấp bằng chứng định lượng trực tiếp cho luận điểm trung tâm của nghiên cứu: dữ liệu hành vi có năng lực phân biệt khách hàng rời bỏ vượt trội so với dữ liệu nhân khẩu học tĩnh.

Bảng 1. Danh sách ngắn 12 chỉ tiêu và hệ số AR đơn biến

#	Chỉ tiêu	AR
1	Hạng thẻ cao nhất khách hàng sở hữu (classic/gold/platinum)	12,09%
2	Hạn mức thẻ tín dụng tại thời điểm quan sát	14,47%
3	Số dư bình quân tài khoản thanh toán 06 tháng gần nhất (VND)	16,24%
4	Tổng số thẻ khách hàng sở hữu tại thời điểm quan sát	18,10%
5	Tổng số thẻ đã kích hoạt tại thời điểm quan sát	13,22%
6	Số thẻ mở trên 02 năm tính đến thời điểm quan sát	21,57%
7	Tuổi thẻ nhỏ nhất (tháng)	26,12%
8	Thời gian dài nhất đến ngày hết hạn thẻ (tháng)	34,82%
9	Dư nợ sao kê bình quân 03 tháng gần nhất	29,35%
10	Thời gian từ giao dịch gần nhất đến thời điểm quan sát (tháng)	36,21%
11	Tổng số giao dịch thành công (chi tiêu + rút tiền) 03 tháng gần nhất	48,53%
12	Tổng giá trị giao dịch thành công (chi tiêu + rút tiền) 03 tháng gần nhất	41,87%

Nguồn: Tác giả tự tổng hợp

4.2. Mô hình cuối cùng

Trong các mô hình ứng viên, lựa chọn mô hình đạt hệ số AR cao nhất đồng thời thỏa mãn các tiêu chí được đề cập tại Mục 3.4. Hai chỉ tiêu có trọng số lớn nhất - thời gian từ giao dịch gần nhất đến thời điểm quan sát (34,74%) và thời gian dài nhất đến ngày hết hạn thẻ (34,01%) - phản ánh trực tiếp mức độ gắn kết hiện hữu và triển vọng gia hạn quan hệ của khách hàng. Ba chỉ tiêu còn lại - dư nợ sao kê bình quân 03 tháng (13,80%), số dư bình quân tài khoản thanh toán 06 tháng (10,45%) và tổng số thẻ sở hữu (7,00%) - bổ sung thông tin về cường độ hoạt động tài chính và độ rộng quan hệ sản phẩm.

Bảng 2. Cấu trúc mô hình cuối cùng

#	Chỉ tiêu	Trọng số
1	Số dư bình quân tài khoản thanh toán 06 tháng gần nhất (VND)	10,45%
2	Tổng số thẻ khách hàng sở hữu tại thời điểm quan sát	7,00%

3	Thời gian dài nhất đến ngày hết hạn thẻ (tháng)	34,01%
4	Dư nợ sao kê bình quân 03 tháng gần nhất	13,80%
5	Thời gian từ giao dịch gần nhất đến thời điểm quan sát (tháng)	34,74%

Nguồn: Tác giả tự tổng hợp

4.3. Đánh giá hiệu năng mô hình

Hiệu năng mô hình được đánh giá theo hai phương diện: năng lực xếp hạng tổng thể (hệ số AR) và độ chính xác phân loại tại ngưỡng điểm cắt (cut-off score). Kết quả tại Bảng 3 cho thấy hệ số AR đạt 61,6% trên mẫu phát triển, 61,1% trên toàn mẫu và 59,9% trên mẫu kiểm tra - đều vượt ngưỡng 55% (ngưỡng tham chiếu: Rất tốt), với chênh lệch giữa mẫu phát triển và mẫu kiểm thử chỉ 1,7 điểm phần trăm, cho thấy mô hình ổn định và không có dấu hiệu quá khớp.

Bảng 3. Hệ số AR của mô hình

Chỉ tiêu	Giá trị
AR - Mẫu phát triển	61,6%
AR - Toàn mẫu	61,1%
AR - Mẫu kiểm thử	59,9%
Ngưỡng tham chiếu: Rất tốt	≥ 55%
Ngưỡng tham chiếu: Chấp nhận được	40% - 55%
Ngưỡng tham chiếu: Không chấp nhận	< 40%

Nguồn: Tác giả tự tổng hợp

Về độ chính xác phân loại, thang điểm rời bỏ (Churn Score) 0-100 được thiết lập theo nguyên tắc điểm càng thấp, xác suất rời bỏ càng cao. Tại ngưỡng phân loại 30 điểm (khách hàng có điểm từ 30 trở xuống được dự báo rời bỏ, từ 30 điểm trở lên được dự báo không rời bỏ), mô hình phân loại đúng 1.660/2.367 quan sát của mẫu kiểm thử, tương ứng độ chính xác 70,13%, và 5.540/7.889 quan sát của toàn mẫu, tương ứng độ chính xác 70,22% (Bảng 4). Trên thực tế vận hành, ngân hàng có thể điều chỉnh ngưỡng điểm theo quy mô chiến dịch marketing và ngân sách giữ chân khách hàng.

Bảng 4. Độ chính xác của mô hình tại ngưỡng điểm 30

Mẫu	Tổng số quan sát	Số quan sát dự báo đúng	Độ chính xác
Mẫu kiểm thử	2.367	1.660	70,13%
Toàn mẫu	7.889	5.540	70,22%

Nguồn: Tác giả tự tổng hợp

4.4. Ứng dụng phân tích đề xuất: phân khúc khách hàng theo Churn Score và B-Score

Ở cấp độ phân tích đề xuất, điểm rời bỏ được kết hợp với điểm hành vi tín dụng (Behavioral Score hay còn gọi là B-Score) - thước đo rủi ro tín dụng của khách hàng trên thang 0-100, điểm càng thấp rủi ro không trả được nợ càng cao - nhằm hình thành ma trận phân khúc bốn nhóm phục vụ phân bổ nguồn

lực giữ chân khách hàng (Bảng 5). Cách tiếp cận này bảo đảm chi phí giữ chân được tập trung vào nhóm khách hàng vừa có nguy cơ rời bỏ cao vừa có chất lượng tín dụng tốt, đồng thời kiểm soát rủi ro tín dụng đối với các nhóm còn lại.

Bảng 5. Ma trận phân khúc khách hàng và định hướng chiến lược

Phân khúc	Đặc điểm	Định hướng chiến lược
Không ưu tiên (No Priority)	Churn Score cao, B-Score thấp	Nguy cơ rời bỏ thấp nhưng rủi ro tín dụng cao: tập trung quản trị rủi ro tín dụng (giám sát định kỳ, điều chỉnh giảm hạn mức, sản phẩm có bảo đảm); duy trì tương tác ở mức cơ bản, không đầu tư ưu đãi chi phí cao.
Duy trì (Maintenance)	Churn Score cao, B-Score cao	Khách hàng ổn định và tín nhiệm tốt: chương trình khách hàng thân thiết theo bậc, bán chéo sản phẩm cao cấp, truyền thông cá nhân hóa nhằm tối đa hóa giá trị vòng đời.
Không giá trị (No Values)	Churn Score thấp, B-Score thấp	Nguy cơ rời bỏ và rủi ro vỡ nợ đồng thời cao: ưu tiên giảm thiểu rủi ro, cân nhắc chiến lược rút lui có kiểm soát; nếu giữ chân, tập trung tái cấu trúc nợ và chuyển đổi sang sản phẩm có bảo đảm.
Giữ chân (Retention)	Churn Score thấp, B-Score cao	Nguy cơ rời bỏ cao nhưng hồ sơ tín dụng tốt — phân khúc ưu tiên cao nhất: ưu đãi giữ chân cá nhân hóa (miễn phí thường niên, nâng hạng ưu đãi), chăm sóc chủ động, kênh dịch vụ ưu tiên và khuyến mại gia hạn thẻ.

Nguồn: Tác giả tự tổng hợp

Hiệu quả các chiến dịch theo phân khúc cần được giám sát bằng hệ thống chỉ tiêu đo lường hiệu quả trọng yếu (Key Performance Indicators - KPIs) như tỷ lệ giữ chân, mức độ tương tác và lợi nhuận, kết hợp cơ chế kiểm thử so sánh (A/B testing) và rà soát định kỳ để tối ưu hóa liên tục nội dung ưu đãi, kênh truyền thông và thời điểm tiếp cận, nhằm tăng hiệu quả giữ chân khách hàng.

5. Kết luận

Nghiên cứu cung cấp bằng chứng thực nghiệm về vai trò của dữ liệu hành vi trong dự báo khách hàng rời bỏ dịch vụ thẻ tín dụng tại thị trường Việt Nam. Trên bộ dữ liệu được thu thập, mô hình thẻ điểm hồi quy logistic gồm 05 chỉ tiêu đạt hệ số AR dao động từ 59,9% đến 61,6% và độ chính xác phân loại 70,13% trên mẫu kiểm thử tại ngưỡng điểm 30. Kết quả trả lời khẳng định cho câu hỏi nghiên cứu trung tâm: dữ liệu hành vi có thể được sử dụng để dự báo rời bỏ với độ chính xác cao, trong khi các biến nhân khẩu học tĩnh (giới tính, tình trạng hôn nhân, trình độ học vấn, tuổi) đều không đạt ngưỡng dự báo tối thiểu $IV \geq 0,1$ hoặc $AR \geq 10\%$.

Về hàm ý quản trị, kết quả nghiên cứu định hướng ba nhóm hành động cho các ngân hàng bán lẻ: (i) ưu tiên đầu tư hạ tầng thu thập và khai thác dữ liệu hành vi giao dịch thay vì chỉ dựa vào hồ sơ nhân khẩu học; (ii) vận hành mô hình dự báo dưới dạng thẻ điểm có tính diễn giải cao nhằm đáp ứng yêu cầu minh bạch trong môi trường ngân hàng chịu giám sát; (iii) kết hợp điểm rời bỏ với điểm rủi ro tín dụng để phân bổ ngân sách giữ chân khách hàng theo nguyên tắc tối đa hóa giá trị điều chỉnh theo rủi ro, thay vì giữ chân đồng loạt.

Nghiên cứu có 3 hạn chế chính. Một là, dữ liệu được thu thập từ một tổ chức tài chính duy nhất tại Việt Nam, do đó kết quả có thể phản ánh đặc thù hành vi khách hàng và điều kiện thị trường trong nước, cần thận trọng khi khái quát hóa sang thị trường khác. Hai là, yêu cầu ẩn danh hóa dữ liệu giới hạn khả năng khai thác một số thông tin chi tiết về nhân khẩu học và giao dịch. Ba là, mô hình dựa trên giả định hành vi quá khứ phản ánh hành vi tương lai - giả định có thể suy giảm hiệu lực khi điều kiện thị trường hoặc chính sách sản phẩm thay đổi đáng kể. Hướng nghiên cứu tiếp theo có thể mở rộng mẫu sang nhiều tổ chức tín dụng, tích hợp dữ liệu tương tác đa kênh (ứng dụng di động, tổng đài chăm sóc khách hàng) và so sánh hiệu năng giữa mô hình thẻ điểm với các thuật toán học máy có kiểm soát tính diễn giải, đồng thời đo lường tác động dài hạn của các chiến dịch giữ chân được triển khai theo khung phân khúc đề xuất ■

TÀI LIỆU THAM KHẢO:

- Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94, 290-301. <https://doi.org/10.1016/j.jbusres.2018.03.003>
- Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), 252-268. <https://doi.org/10.1016/j.ejor.2003.12.010>
- Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98-107.
- Gupta, S., & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing*, 17(1), 9-24. <https://doi.org/10.1002/dir.10045>
- Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10), 2902-2917. <https://doi.org/10.1016/j.cor.2005.11.007>
- Lazarov, V., & Capota, M. (2007). Churn prediction. Business Analytics Course, TUM Computer Science.
- Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2), 276-286. <https://doi.org/10.1509/jmkr.43.2.276>
- Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50, 57-70. <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>
- Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, 43(2), 204-211.
- Quinlan, B., Chung, J., Liu, G., Wang, D. B., & Fan, M. (2024). From acquisition to activation: Revisiting credit card issuer strategies in Hong Kong. Quinlan & Associates.
- Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: Quality comes to services. *Harvard Business Review*, 68(5), 105-111.
- Shirazi, F., & Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, 48, 238-253. <https://doi.org/10.1016/j.ijinfomgt.2018.10.005>
- Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3), 2354-2364. <https://doi.org/10.1016/j.eswa.2010.08.023>