

PHÂN TÍCH HÀNH VI VÀ PHÂN KHÚC THỊ TRƯỜNG: TIẾP CẬN TỪ BIG DATA TRONG NGÀNH NGÂN HÀNG VIỆT NAM

PHẠM VĂN HẬU

Email: phamvanhauhg@gmail.com

Ngày nhận bài: 3/6/2026

Ngày sửa bài: 19/6/2026

Ngày duyệt đăng bài: 4/7/2026

Tóm tắt:

Nghiên cứu này kết hợp mô hình Recency-Frequency-Monetary (RFM) với thuật toán phân cụm K-Means để phân tích hành vi khách hàng cá nhân tại một ngân hàng thương mại ở Việt Nam. Dữ liệu nghiên cứu gồm 4.642 bản ghi giao dịch của 845 khách hàng, được chiết xuất từ kho dữ liệu, hệ thống Core Banking và hệ thống quản lý quan hệ khách hàng (CRM). Kết quả xác định 4 phân khúc khách hàng có đặc điểm và giá trị giao dịch khác nhau, làm cơ sở xây dựng chính sách chăm sóc và khai thác khách hàng phù hợp. Nghiên cứu bổ sung bằng chứng thực nghiệm về việc ứng dụng RFM kết hợp K-Means trên dữ liệu giao dịch thực tế của ngân hàng Việt Nam, đồng thời cung cấp cơ sở khoa học cho việc nâng cao hiệu quả quản trị quan hệ khách hàng, giữ chân khách hàng và tăng cường năng lực cạnh tranh.

Từ khóa: dữ liệu lớn, RFM, phân khúc thị trường, CLV, K-Means, ngân hàng thương mại.

Customer behavior analysis and segmentation: A Big Data approach in Vietnam's banking sector

Abstract:

The rapid development of digital technologies and Big Data in the banking sector has facilitated the application of analytical models to identify customer behavior. This study applies the Recency-Frequency-Monetary (RFM) model to analyze the behavior of individual banking customers. The research dataset comprises 4,642 transaction records from 845 customers during the period from October 1, 2024, to October 30, 2025. The data were extracted from the data warehouse, the Core Banking system, and the Customer Relationship Management (CRM) system. The analytical results enable the identification of the characteristics and service usage behavior of different customer groups, thereby providing a basis for banks to develop business strategies and customer care policies tailored to the needs and expectations of each segment. These findings contribute to improving customer retention and strengthening banks' competitive capabilities.

Keywords: Big Data, RFM, market segmentation, customer lifetime value, K-Means, commercial banking.

1. Đặt vấn đề

Sự phát triển của công nghệ số và ngân hàng điện tử đã tạo ra khối lượng lớn dữ liệu giao dịch, phản ánh hành vi, nhu cầu và mức độ gắn kết của khách hàng. Trong bối cảnh cạnh tranh ngày càng gay gắt, dữ liệu không chỉ phục vụ hoạt động tác nghiệp mà còn trở thành nguồn lực chiến lược giúp các ngân hàng nâng cao hiệu quả quản trị quan hệ khách hàng (CRM). Thông qua Big Data, ngân hàng có thể phân tích hành vi giao dịch, dự báo nhu cầu và phân khúc khách hàng theo giá trị kinh tế, từ đó cá thể hóa sản phẩm, tối ưu hóa hoạt động tiếp thị và nâng cao hiệu quả kinh doanh (More & Moily, 2021).

Tuy nhiên, tại Việt Nam, việc ứng dụng Big Data trong phân khúc khách hàng ngân hàng còn hạn chế. Phần lớn các nghiên cứu chủ yếu dựa trên dữ liệu khảo sát hoặc quy mô mẫu nhỏ, trong khi các nghiên cứu khai thác dữ liệu giao dịch thực tế từ hệ thống ngân hàng còn ít. Đặc biệt, việc kết hợp đồng thời các chỉ số của Recency-Frequency-Monetary để hình thành mô hình RFM (Christy et al., 2018) chưa được các nghiên cứu liên quan lĩnh vực ngân hàng làm rõ. Trong đó, Recency là thời gian mua hàng gần nhất tính đến thời điểm nghiên cứu. Frequency là tần suất giao dịch của khách hàng. Monetary là tổng giá trị khách hàng thực hiện giao dịch trong thời gian nghiên cứu.

Đã có một vài nghiên cứu trước như Aggelis & Christodoulakis (2005) áp dụng thuật toán RFM để phân tích hành vi khách hàng tham gia giao dịch tại Ngân hàng. Tuy nhiên, nhóm tác giả này chưa khai thác RFM để thực hiện phân khúc thị trường. Tương tự, Mohammadi et al. (2014) áp dụng RFM để xếp hạng tín dụng cho khách hàng của ngân hàng cũng không đi phân tích kết hợp các thông số R, F, M để thực hiện phân khúc thị trường. Hạn chế này cũng được thấy trong nghiên cứu của Christy et al. (2021). Xuất phát từ những luận điểm của nghiên cứu trước cho thấy, tồn tại khoảng trống nghiên cứu liên quan đến vận dụng RFM để phân khúc khách hàng và xác định tỷ lệ khách hàng của từng phân khúc nhằm làm cơ sở xây dựng chiến lược marketing. Đây là cơ hội để nghiên cứu này lấp đầy khoảng trống nêu trên.

Xuất phát từ nhận định trên, mục tiêu của nghiên cứu này phân tích hành vi của khách hàng tham gia giao dịch tại ngân hàng. Trên cơ sở đó, để tiến đến thực hiện phân khúc thị trường dựa vào thuật toán RFM. Kết quả khám phá là cơ sở để nghiên cứu đề xuất hàm ý quản trị tối ưu hoạt động kinh doanh tại ngân hàng.

2. Lược khảo tài liệu

Với sự mở rộng nhanh chóng của ngành ngân hàng và sự gia tăng về mức độ phức tạp trong hành vi của khách hàng, ngân hàng phải đối mặt với những thách thức đáng kể trong việc quản lý rủi ro và tối ưu hóa phân bổ nguồn lực (More et al., 2021). Để giải quyết các vấn đề rủi ro, một số tác giả đã xây dựng mô hình thuật toán để phân tích hành vi khách hàng. Vasheghani et al. (2025) đã phát triển các thuật toán sử dụng mạng nơ-ron Perceptron đa lớp (Multi-Layer Perceptron - MLP) được tăng cường bằng Tối ưu hóa bầy đàn hạt tự thích nghi (Self-Adaptive Particle Swarm Optimization - PSO) và Thuật toán di truyền (Genetic Algorithms). Với sự phát triển của Big Data, phân tích hành vi khách hàng trong ngân hàng cực kỳ quan trọng (Dufouleur, 2026). Vasheghani et al. (2025) khám phá dữ liệu, trực quan hóa dữ liệu và triển khai các mô hình học máy trong lĩnh vực ngân hàng thông qua việc sử dụng các mô hình gồm SGD Classifier, K-Nearest Neighbor (KNN) Classifier và Random Forest Classifier. Kết quả cho thấy, mô hình Random Forest Classifier. Những phát hiện này cung cấp các hàm ý hữu ích cho các ngân hàng trong việc tối ưu hóa chiến lược tiếp thị trong bối cảnh ngành tài chính ngày càng năng động và cạnh tranh. Tuy nhiên nghiên cứu không đề cập đến phân tích hành vi và thực hiện phân khúc khách hàng, đây được xem như là khoảng trống mà nghiên cứu này sẽ hướng đến thông qua thuật toán RFM.

Để Mô hình RFM như đã nêu do Hughes (1994) đề xuất là công cụ phổ biến để đánh giá khách hàng dựa trên thời gian giao dịch gần nhất, tần suất và giá trị giao dịch. Phương pháp này là cơ sở giúp hướng đến phân khúc khách hàng dựa vào thuật toán phân tích cụm thông qua phương pháp K-Means Clustering (Aggelis & Christodoulakis, 2005). Điền hình Christy et al. (2018) đã sử dụng thuật toán RFM để phân khúc khách hàng thông qua phương pháp K-means truyền thống và Fuzzy C-Means (FCM) (Karahoca & Karahoca, 2011). Bên cạnh đó, nhóm tác giả đề xuất một phương pháp mới để lựa chọn các tâm cụm khởi tạo (initial centroids) cho thuật toán K-means. Kết quả thu được từ các phương pháp được so sánh dựa trên các tiêu chí gồm số lần lặp (iterations), độ chặt của cụm (cluster compactness) và thời gian thực thi (execution time).

3. Phương pháp nghiên cứu

Nguồn dữ liệu của nghiên cứu này được chiết xuất từ nguồn Big Data của ngân hàng từ hệ thống core T24 (hệ thống lõi của ngân hàng), CRM (hệ thống thu thập và quản lý thông tin khách hàng), mis (data warehouse), portal, của một ngân hàng thương mại cổ phần. Do tính bảo mật của ngân hàng, nên nghiên cứu này xin phép được giấu tên. Đây là những nguồn dữ liệu quan trọng bởi độ tin cậy lớn và ghi lại chính xác hành vi giao dịch của từng khách hàng. Mặc dù số liệu không thuộc dạng khảo sát, tuy nhiên việc sử dụng nguồn dữ liệu Big Data này được xem như là một xu hướng mới trong việc nghiên cứu hành vi trong kỷ nguyên công nghệ thông tin phát triển.

Số liệu được tập hợp từ các nguồn khác nhau theo loại khách hàng, trong đó có khách hàng sử dụng thẻ tín dụng, khách hàng tiền gửi và lịch sử giao dịch tiền gửi và khách hàng tiền vay. Để hạn chế những thông tin nhiễu, các khách hàng được chọn đưa vào nghiên cứu có thời gian tham gia dịch vụ trên 12 tháng. Thời gian nghiên cứu của số liệu đối với khách hàng là 12 tháng, tính từ ngày 1/10/2024 đến ngày 30/10/2025. Các đối tượng khách hàng trong nghiên cứu là khách hàng có tham gia giao dịch trong thời gian nghiên cứu. Công cụ được sử dụng hỗ trợ phục vụ phân tích là SPSS Modeler 18.

Quy trình nghiên cứu gồm 4 bước: (i) làm sạch và chuẩn hóa dữ liệu; (ii) tính toán các thông số của R, F, M (RFM); (iii) áp dụng thuật toán K-Means Clustering để phân nhóm khách hàng có đặc điểm hành vi tương đồng. Phương pháp RFM dựa vào một vài nghiên cứu của Christy et al. (2018); Monalisa et al. (2019) được tổng hợp ở Hình 1.

Sau khi chuẩn hóa dữ liệu, thuật toán K-Means (Chang & Tsay, 2004) được sử dụng để xác định các phân khúc khách hàng, với số lượng cụm được lựa chọn thông qua quá trình thử nghiệm nhằm bảo đảm tính đồng nhất trong từng cụm và sự khác biệt giữa các cụm. Toàn bộ quá trình xử lý và phân tích dữ liệu được thực hiện bằng IBM SPSS Modeler 18.

4. Kết quả phân tích dữ liệu

Như đã trình bày trước đó, có 845 khách hàng với 4.642 giao dịch được khách hàng thực hiện trong thời gian nghiên cứu trong giai đoạn từ ngày 01/10/2024 đến ngày 30/10/2025. Kết quả thống kê mô tả của R, F, M là giá trị bình quân của một khách hàng được tính từ ngày 01/10/2024 đến ngày 30/10/2025. Kết quả tính được $R = 12$ ngày, tức là ngày giao dịch gần nhất một khách hàng tính đến thời điểm nghiên cứu là 12 ngày. Tổng số lần giao dịch bình quân $F = 95$ ngày và tổng số tiền giao dịch bình quân $M = 480$ triệu đồng. Dựa trên giá trị thống kê mô tả, R, F, M được chuyển thành giá trị 5 thứ bậc (từ 1 đến 5). Trong đó 1 được đánh giá là rất thấp và 5 là rất cao. Việc xếp hạng và đặt tên nhóm phân khúc là dựa vào Dash & Mishra (2010).

Hình 1. Thông tin thuật toán của RFM



Nguồn: Tự tổng hợp.

Theo nhóm tác giả, nếu thành phần R cao, F cao, M cao, được gọi là nhóm khách hàng đóng góp tốt nhất (best group). Trường hợp R thấp, F thấp, M thấp được gọi là nhóm có nguy cơ rời dịch vụ. R cao, F thấp, M thấp được gọi là nhóm khách hàng mới. R thấp, F cao, M cao được gọi là nhóm khách hàng thích chi tiêu. Việc định tên này được áp dụng trong kết quả dữ liệu của nghiên cứu. Kết quả ở Bảng 1, R (thời gian giao dịch sau cùng tính đến thời điểm hiện tại), tức là thời gian mua hàng lần gần nhất tính đến thời gian nghiên cứu có số ngày nhỏ được xếp vào giá trị 5 và có số ngày lớn được xếp vào giá trị 1. Riêng với F và M thì ngược lại, khi được chuyển đổi vào giá trị thứ bậc từ 1 đến 5 thì giá trị càng lớn của F và M càng tiến dần về 5. Như vậy, các chỉ số R, F, M được chuyển đổi về giá trị thứ bậc từ 1 đến 5. Nên giá trị trong dấu ngoặc của bảng là giá trị bình quân được tính từ 1 đến 5.

Bảng 1. Kết quả phân khúc thị trường dựa vào RFM

Label	Description	Size	Inputs		
			Recency	Frequency	Monetary
R= LOW; F= LOW; M= LOW	Churn group	26.5%	1.35	1.43	1.54
R= HIGH; F= HIGH; M= HIGH	Best group	34.5%	3.73	4.54	4.48
R= HIGH; F= LOW; M= LOW	New group	19.5%	3.39	1.96	2.05
R= LOW; F= HIGH; M= HIGH	Frequent group	19.8%	2.20	3.22	3.34

Ghi chú: Churn group: Nhóm khách hàng có nguy cơ rời dịch vụ; Best group: Nhóm khách hàng đóng góp tốt nhất; New group = Nhóm khách hàng mới; Frequeny group: Nhóm khách hàng thích chi tiêu.

Nguồn: Dữ liệu big data của ngân hàng

Nhóm khách hàng “Churn group”: được gọi là nhóm có nguy cơ rời dịch vụ. Nhóm này chiếm tỷ lệ 26,5%, tương đương với 225 khách hàng. Đây là nhóm có các giá trị $R = \text{THẤP}$, $F = \text{THẤP}$ và $(M = \text{THẤP})$.

Nhóm khách hàng “Best group”: Đây là nhóm khách hàng đóng góp lợi ích lớn nhất cho ngân hàng. Nhóm này chiếm 34,5% tương đương với 288 khách hàng. Các chỉ số của R , F , M đều có giá trị ở mức cao ($R = \text{Cao}$; $F = \text{Cao}$; $M = \text{Cao}$). Tức là hành vi giao dịch của nhóm khách hàng thường xuyên thực hiện giao dịch tại ngân hàng, có giá trị giao dịch cao và các giao dịch thực hiện cách thời gian nghiên cứu rất gần.

Nhóm khách hàng “New group” là nhóm khách hàng được xác định là nhóm mới chiếm tỷ lệ 19,5% tương đương với 164 khách hàng. Nhóm khách hàng này được xem là nhóm mới vì có chỉ số $R = \text{CAO}$, tần suất giao dịch $F = \text{THẤP}$ và giá trị giao dịch $M = \text{THẤP}$.

Nhóm khách hàng “Frequent group” là nhóm khách hàng thường xuyên chiếm tỷ lệ 19,8% tương đương với 168 khách hàng. Nhóm được xem là nhóm khách hàng thường xuyên vì có chỉ số $R = \text{THẤP}$, tần suất giao dịch $F = \text{CAO}$, giá trị giao dịch $M = \text{CAO}$). Đây là nhóm khách hàng thương đến ngân hàng để thực hiện giao dịch.

5. Thảo luận

Dựa vào tiếp cận thuật toán của RFM đối với nguồn dữ liệu Big Data của ngân hàng, kết quả nghiên cứu đã nhận diện được hành vi khách hàng và được phân thành 4 nhóm khách hàng (Nhóm có nguy cơ rời dịch vụ- churn group; Nhóm tốt nhất - best group; Nhóm khách hàng mới - new group; Nhóm khách hàng thường xuyên- frequent group). Kết quả cho một giá trị đóng góp quan trọng trong chiến lược phân khúc thị trường. Kết quả này hợp với (Hughes, 1994) và khẳng định mức độ hiệu quả của mô hình RFM trong đánh giá giá trị khách hàng. Điểm mới của nghiên cứu là mô hình được kiểm định trên dữ liệu giao dịch thực tế của khách hàng tại một ngân hàng thương mại tại Việt Nam thay vì dữ liệu khảo sát hoặc mô phỏng.

Sự khác biệt giữa các phân khúc cho thấy giá trị khách hàng cần được đánh giá đồng thời theo thời gian giao dịch, tần suất sử dụng và giá trị giao dịch. Nhóm “best group” có mức độ gắn kết và giá trị kinh tế cao, trong khi nhóm “new group” tuy có giá trị hiện tại thấp nhưng sở hữu tiềm năng phát triển nếu được chăm sóc phù hợp. Đối với nhóm “frequent group”, giá trị giao dịch lớn chưa đi kèm mức độ trung thành tương xứng, hàm ý ngân hàng cần tăng cường bán chéo, nâng cao trải nghiệm và gia tăng tương tác để cải thiện khả năng duy trì khách hàng. Đối với nhóm “churn group” ngân hàng cần ưu tiên theo dõi, tái kích hoạt và tăng cường tương tác với nhóm khách hàng này nhằm khai thác tiềm năng và giảm nguy cơ rời bỏ.

Nghiên cứu đồng thời khẳng định ưu thế của mô hình phân khúc dựa trên dữ liệu giao dịch so với các phương pháp truyền thống dựa trên đặc điểm nhân khẩu học hoặc kinh nghiệm quản trị. Việc tích hợp kết quả phân khúc vào hệ thống CRM giúp ngân hàng tối ưu hóa phân bổ nguồn lực, nâng cao hiệu quả chăm sóc khách hàng và hỗ trợ ra quyết định trong bối cảnh chuyển đổi số.

6. Kết luận và hàm ý quản trị

Nghiên cứu xây dựng mô hình phân tích hành vi khách hàng dựa trên dữ liệu giao dịch tại ngân hàng thương mại thông qua việc kết hợp các thành phần Recency - Frequency - Monetary để tạo thành thuật toán của RFM, trên cơ sở đó vận dụng phương pháp phân tích cụm của K-Means. Kết quả cho thấy, mô hình giúp nhận diện hiệu quả các nhóm khách hàng theo đặc điểm hành vi và giá trị kinh tế, khẳng định

vai trò của dữ liệu giao dịch trong quản trị quan hệ khách hàng dựa trên dữ liệu. Khác với các nghiên cứu sử dụng dữ liệu khảo sát, nghiên cứu khai thác trực tiếp dữ liệu giao dịch thực tế, góp phần bổ sung bằng chứng thực nghiệm về khả năng ứng dụng Big Data trong quản trị quan hệ khách hàng tại các ngân hàng thương mại Việt Nam.

Từ kết quả nghiên cứu, bài báo đề xuất một số hàm ý quản trị cần được quan tâm tại một ngân hàng thương mại. Thứ nhất, chuyển từ phân khúc khách hàng truyền thống sang phân khúc dựa trên dữ liệu giao dịch và cập nhật kết quả phân khúc định kỳ để nâng cao hiệu quả quản trị khách hàng. Thứ hai, xây dựng chiến lược CRM theo từng phân khúc nhằm triển khai các chương trình chăm sóc, bán chéo sản phẩm và giữ chân khách hàng phù hợp. Thứ ba, tăng cường ứng dụng Big Data và tích hợp các mô hình phân tích vào hệ thống CRM để hỗ trợ dự báo hành vi, nhận diện nguy cơ rời bỏ và nâng cao chất lượng ra quyết định. Thứ tư, đầu tư phát triển hạ tầng và năng lực phân tích dữ liệu, đồng thời tăng cường kết nối giữa các hệ thống Core Banking, CRM và Data Warehouse nhằm hình thành nền tảng quản trị khách hàng theo định hướng dữ liệu trong bối cảnh chuyển đổi số ■

TÀI LIỆU THAM KHẢO:

- Aggelis, V., & Christodoulakis, D. (2005). RFM analysis for decision support in e-banking area. *WSEAS Transactions on Computers*, 4(8), 943-950.
- Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking - An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1251-1257. <https://doi.org/10.1016/j.jksuci.2018.09.004>
- Dash, P., & Mishra, S. (2010). Developing Rfm Model for Customer. *International Journal of Marketing & Human*, 1(1), 58-69.
- Dufouleur, M. (2026). Bitcoin market segmentation and regulatory effect. *Journal of International Money and Finance*, 165, 103570. <https://doi.org/10.1016/j.jimonfin.2026.103570>
- Mohammadi, R., Bidabad, B., Nourasteh, T., & Sherafati, M. (2014). Credit Ranking of Bank Customers (An Integrated Model of. *European Online Journal of Natural and Social Science*, 3(3), 564-571.
- More, R., Moily, Y., & Student, B. E. (2021). Big Data Analysis in Banking Sector. *International Journal of Engineering Research and Applications Wwww.Ijera.Com*, 11, 1-05. <https://doi.org/10.9790/9622-1104020105>
- Vasheghani, M., Farokhi, E. N., & Dolatshah, B. (2025). Forecasting loan, deferred rate and customer segmentation in banking industry: A computational intelligence approach. *Array*, 27, 100460. <https://doi.org/10.1016/j.array.2025.100460>