

## CHẤM ĐIỂM TÍN DỤNG CÁ NHÂN TỪ DỮ LIỆU THAY THẾ: TRƯỜNG HỢP CÁC TỔ CHỨC TÍN DỤNG TẠI VIỆT NAM

Nguyễn Duy Vũ

*Trường Kinh tế, Đại học Bách Khoa Hà Nội, Hà Nội, Việt Nam*

Bùi Quốc Trung

*Trường Công nghệ thông tin và Truyền thông, Đại học Bách Khoa Hà Nội, Hà  
Nội, Việt Nam*

Nguyễn Thúc Hương Giang<sup>1</sup>

*Trường Kinh tế, Đại học Bách Khoa Hà Nội, Hà Nội, Việt Nam*

Ngày nhận: 07/03/2025; Ngày hoàn thành biên tập: 08/07/2025; Ngày duyệt đăng: 28/7/2025

DOI: <https://doi.org/10.38203/jiem.vi.032025.1261>

**Tóm tắt:** Bài viết phân tích tiềm năng ứng dụng dữ liệu thay thế trong việc đánh giá điểm tín dụng của khách hàng cá nhân tại các tổ chức tín dụng tại Việt Nam. Dữ liệu nghiên cứu từ cơ sở dữ liệu quốc gia về dân cư và lịch sử tín dụng từ các tổ chức tài chính tại Việt Nam được phân tích bằng phương pháp định lượng thông qua việc xây dựng mô hình hồi quy kết hợp với các thuật toán học máy. Kết quả thực nghiệm cho thấy việc tích hợp dữ liệu thay thế có thể cải thiện đáng kể khả năng dự báo điểm tín dụng, đồng thời mở rộng cơ hội tiếp cận dịch vụ tài chính cho người dân. Trên cơ sở đó, bài viết đưa ra những đóng góp cả về phương diện lý luận và thực tiễn, qua đó góp phần hoàn thiện khung đánh giá tín dụng cá nhân tại Việt Nam theo hướng hiện đại và toàn diện hơn.

**Từ khóa:** Dữ liệu truyền thống, Dữ liệu thay thế, Đánh giá tín dụng, Khách hàng cá nhân, Tổ chức tín dụng, Việt Nam

### INDIVIDUAL CREDIT SCORING USING ALTERNATIVE DATA: EVIDENCE FROM CREDIT INSTITUTIONS IN VIETNAM

**Abstract:** This paper analyzes the potential application of alternative data in assessing individual credit scores at financial institutions in Vietnam. The research utilizes data from the National Population Database and credit histories provided by financial organizations in Vietnam. A quantitative approach is employed, combining regression modeling with machine learning algorithms. The experimental results indicate that the integration of alternative data can significantly enhance the predictive accuracy of credit scoring models, while also expanding access to financial services for the population. Based on these findings, the study offers both theoretical and practical contri-

<sup>1</sup> Tác giả liên hệ, Email: [giang.nguyenthuchuong@hust.edu.vn](mailto:giang.nguyenthuchuong@hust.edu.vn)

butions, thereby improving the personal credit assessment framework in Vietnam toward a more modern and comprehensive approach.

**Keywords:** Traditional Data, Alternative Data, Credit Scoring, Individual Customers, Credit Institutions, Vietnam

---

## 1. Giới thiệu

Tín dụng ngày càng đóng vai trò quan trọng trong nền kinh tế hiện đại, không chỉ cung cấp nguồn vốn chủ yếu cho doanh nghiệp mà còn đáp ứng nhu cầu tiêu dùng cá nhân. Rủi ro tín dụng là khả năng khách hàng không thực hiện hoặc không sẵn sàng thực hiện các cam kết tài chính, đây luôn là mối quan tâm hàng đầu của các tổ chức tín dụng (Spuchl'áková & cộng sự, 2015). Trong đó, đánh giá tín dụng là một thành phần cốt lõi trong quản lý rủi ro tín dụng (Abdou & Pointon, 2011) và điểm tín dụng đóng vai trò then chốt trong việc ra quyết định cấp tín dụng. Theo Thomas & cộng sự (2002), đánh giá tín dụng là “một tập hợp các mô hình quyết định để hỗ trợ các tổ chức cho vay trong việc cấp tín dụng”. Bên cạnh đó, điểm tín dụng còn được sử dụng để xác định lãi suất, điều kiện vay và hỗ trợ quản lý danh mục tín dụng nhằm giảm thiểu tổn thất tài chính (Siddiqi, 2012; Bennouna & Tairi, 2019).

Tại Việt Nam, thị trường cho vay khách hàng cá nhân đã có sự phát triển nhanh chóng trong hai thập kỷ gần đây. Nhiều kiến nghị, đề xuất góp phần ngăn chặn đẩy lùi “tín dụng đen” đã được gửi tới Ngân hàng Nhà nước (Hiệp hội Ngân hàng Việt Nam, 2023). Tính đến ngày 31/12/2022, tổng dư nợ của 16 công ty tài chính được ngân hàng Nhà nước cấp phép đạt hơn 220 nghìn tỷ đồng, chiếm 1,87% tổng dư nợ toàn nền kinh tế và 8,5% dư nợ tiêu dùng của hệ thống tài chính, hỗ trợ khoảng 30 triệu người vay vốn với dư nợ trung bình 35-50 triệu đồng. Cùng với các ngân hàng thương mại, các công ty tài chính và công nghệ tài chính, thị trường tín dụng còn ghi nhận sự tham gia của nhiều tổ chức không chính thức như các cửa hàng cầm đồ, công ty P2P lending, công ty cho vay ngắn hạn và dịch vụ mua trước trả sau (Hoàng, 2023).

Trong bối cảnh đó, việc phát triển các mô hình đánh giá tín dụng sử dụng kết hợp giữa dữ liệu truyền thống và dữ liệu thay thế bao gồm hành vi trên mạng xã hội, lịch sử giao dịch trực tuyến, thông tin dân cư, mở ra tiềm năng to lớn cho cả khách hàng và tổ chức tín dụng. Đặc biệt, các mô hình này có thể hỗ trợ việc tiếp cận tín dụng cho nhóm khách hàng chưa có hoặc có rất ít lịch sử tín dụng (Phan, 2024). Trên thế giới, nhiều nghiên cứu đã khẳng định tiềm năng của dữ liệu thay thế trong cải thiện khả năng dự báo rủi ro tín dụng, đặc biệt đối với nhóm khách hàng chưa có lịch sử tín dụng chính thức. Chẳng hạn, Njuguna & Sowon (2021) đã tổng hợp các nghiên cứu thực nghiệm sử dụng dữ liệu viễn thông, hóa đơn tiện ích và hành vi kỹ thuật số để đánh giá tín dụng tại các nước đang phát triển. Trong khi đó, Wijaya & Nugroho (2023) chỉ ra Trung Quốc và Ấn Độ đã xây dựng hệ thống đánh giá tín dụng dựa trên mạng xã hội, thương mại điện tử và dữ liệu định danh công dân, từ đó hỗ trợ mục tiêu tài chính toàn diện.

Mặc dù lĩnh vực này có ý nghĩa thực tiễn rõ rệt, nhưng nghiên cứu về đánh giá tín dụng cá nhân tại Việt Nam, đặc biệt với sự tham gia của dữ liệu thay thế, vẫn còn rất hạn chế (Huyen & Kleimeier, 2007; Tran & cộng sự, 2022; Gardner & Lainez, 2023). Do đó, nghiên cứu này được thực hiện nhằm khắc phục khoảng trống trong tài liệu hiện hành, thông qua việc làm rõ các câu hỏi nghiên cứu sau: Những loại dữ liệu thay thế nào hiện đang được sử dụng trong đánh giá tín dụng cá nhân tại Việt Nam và trên thế giới? Dữ liệu

từ cơ sở dữ liệu quốc gia về dân cư có thể được tích hợp vào mô hình đánh giá tín dụng cá nhân như thế nào? Mức độ hiệu quả của mô hình RBCredit, khi ứng dụng dữ liệu dân cư, trong việc phân loại khách hàng tín dụng tốt và xấu ra sao? Việc triển khai mô hình chấm điểm tín dụng dựa trên dữ liệu thay thế tại Việt Nam có phù hợp với bối cảnh pháp lý và điều kiện thực tiễn hiện nay hay không?

Thông qua việc làm rõ cơ sở lý luận, phân tích thực tiễn và thử nghiệm ứng dụng dữ liệu dân cư trong đánh giá điểm tín dụng khách hàng cá nhân, nghiên cứu kỳ vọng sẽ đóng góp vào việc hoàn thiện mô hình đánh giá tín dụng tại Việt Nam. Cụ thể, nghiên cứu đề xuất hướng tiếp cận dữ liệu mới, từ đó góp phần mở rộng khả năng tiếp cận tài chính cho người dân và tăng cường kiểm soát rủi ro tín dụng của các tổ chức cho vay.

Cấu trúc bài viết gồm 5 phần. Sau phần giới thiệu, phần 2 trình bày cơ sở lý thuyết làm nền tảng cho phân tích. Tiếp đó, phần 3 phương pháp nghiên cứu. Phần 4 trình bày kết quả nghiên cứu và thảo luận. Cuối cùng, phần 5 đưa ra kết luận.

## **2. Cơ sở lý thuyết**

### **2.1 Các mô hình đánh giá tín dụng truyền thống**

Các mô hình đánh giá tín dụng truyền thống chủ yếu dựa trên phương pháp thống kê, trong đó nổi bật là điểm tín dụng FICO do Công ty Fair Isaac phát triển từ năm 1989. Điểm FICO được tính toán dựa trên các dữ liệu về lịch sử thanh toán, dư nợ hiện tại, thời gian sử dụng tín dụng, cơ cấu tín dụng và số lượng đơn xin tín dụng mới (Hommonoff & cộng sự, 2021). Các tổ chức tài chính sử dụng điểm này để phân loại khách hàng và đưa ra quyết định tín dụng phù hợp.

Ưu điểm của các mô hình truyền thống là đơn giản, chi phí thấp và nhất quán, nhưng tất cả phụ thuộc vào lịch sử tín dụng. Điều này dẫn đến việc loại trừ nhóm cá nhân “không có hồ sơ tín dụng” như người trẻ tuổi, người nhập cư hoặc người chưa từng vay tiêu dùng, từ đó hạn chế khả năng tiếp cận tín dụng của các đối tượng này (Kiviat, 2019).

### **2.2 Dữ liệu thay thế trong đánh giá tín dụng**

Để khắc phục hạn chế của các mô hình truyền thống, khái niệm “đánh giá tín dụng thay thế” xuất hiện từ đầu những năm 2000. Dữ liệu thay thế cho phép mở rộng phạm vi đánh giá tín dụng bằng cách khai thác thông tin phi truyền thống, góp phần tạo ra bức tranh đầy đủ hơn về hành vi tài chính và mức độ rủi ro của khách hàng (FICO Blog, 2024).

Theo World Bank (2019), dữ liệu tín dụng có thể chia thành hai nhóm. Dữ liệu truyền thống bao gồm giao dịch ngân hàng, lịch sử tín dụng, dữ liệu thương mại. Dữ liệu thay thế bao gồm thanh toán tiện ích, hành vi trực tuyến, dữ liệu từ mạng xã hội, ứng dụng di động. Njuguna & Sowon (2021) cho thấy hóa đơn tiện ích là một trong những loại dữ liệu thay thế đầu tiên được sử dụng. Các nghiên cứu gần đây cũng đề cập đến việc sử dụng dữ liệu từ hồ sơ công cộng, lịch sử cư trú, điện thoại di động, hoạt động mạng xã hội, thói quen chi tiêu trực tuyến và thậm chí cả dữ liệu tâm lý học.

FICO định nghĩa dữ liệu thay thế là bất kỳ loại thông tin nào không trực tiếp phản ánh hành vi tín dụng truyền thống, đồng thời nhấn mạnh việc kết hợp cả hai loại dữ liệu để nâng cao độ chính xác dự báo. Mô hình VantageScore 4Plus ra mắt năm 2024 là một ví dụ điển hình, kết hợp dữ liệu truyền thống và dữ liệu mở từ hệ thống ngân hàng (VantageScore, 2024).

Tại Việt Nam, với tỷ lệ sử dụng điện thoại thông minh đạt 84% trong tháng 3/2024 (Edtech Agency, 2024) và mức độ phổ cập Internet di động cao cùng sự phát triển của các nền tảng mạng xã hội, việc ứng dụng dữ liệu thay thế trong chấm điểm tín dụng đang trở nên khả thi hơn bao giờ hết. Các quy định pháp lý hỗ trợ như Nghị định số 94/2025/NĐ-CP mới được ban hành quy định về cơ chế thử nghiệm có kiểm soát trong lĩnh vực ngân hàng, Nghị quyết số 175/NQ-CP phê duyệt dự án Trung tâm Dữ liệu quốc gia và Luật dữ liệu số 60/2024/QH15 đã mở ra cơ hội lớn cho việc khai thác và ứng dụng dữ liệu đa nguồn trong phát triển mô hình chấm điểm tín dụng tại Việt Nam. Tuy nhiên, các công ty hoạt động trong lĩnh vực này cũng đối diện với yêu cầu tuân thủ các quy định pháp lý nghiêm ngặt về bảo vệ dữ liệu cá nhân. Luật Bảo vệ dữ liệu cá nhân số 91/2025/QH15 và Nghị định số 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân đóng vai trò quan trọng trong việc định hình hành lang pháp lý cho lĩnh vực này.

### ***2.3 Thực trạng nghiên cứu trên thế giới***

Các quốc gia trên thế giới áp dụng đa dạng mô hình và phương pháp đánh giá tín dụng. Tại Hoa Kỳ và Canada, các tổ chức như Equifax và TransUnion sử dụng điểm tín dụng dựa trên lịch sử thanh toán, dư nợ và các chỉ số tín dụng truyền thống (Chase, 2024). Clean Credit (2017) cung cấp cái nhìn tổng quát về sự khác biệt giữa hệ thống báo cáo tín dụng ở Australia có 4 cơ quan tín dụng lớn (Experian, Equifax, Dun & Bradstreet, Tasmanian Collection Service) và đánh giá tín dụng dựa trên cả thông tin tích cực và tiêu cực. Tuy nhiên, Đức chỉ có một cơ quan tín dụng lớn tên là Schufa, sử dụng thông tin người tiêu dùng như lịch sử tài chính, thẻ tín dụng và thông tin cá nhân.

Nhật Bản không có hệ thống tín dụng quốc gia chính thức, các tổ chức tài chính tự quyết định mô hình đánh giá tín dụng. Tây Ban Nha không có điểm tín dụng, tuy nhiên trung tâm Quản lý rủi ro của quốc gia theo dõi tất cả các hoạt động tín dụng và vay mượn tập trung vào các mục tiêu tiêu cực trên báo cáo tín dụng của khách hàng và định nghĩa danh sách khách hàng có hành vi tín dụng xấu (Chase, 2024). Trung Quốc xây dựng hệ thống Tín dụng xã hội, sử dụng dữ liệu từ các ông lớn công nghệ như Baidu, Alibaba, Tencent bao gồm hồ sơ tín dụng tài chính, xu hướng hành vi trong giao dịch thương mại, tài sản và thông tin cá nhân, hành vi và sở thích và các mối quan hệ xã hội để đánh giá tín dụng (Wijaya & Nugroho, 2023).

Njuguna & Sowon (2021) cho thấy xu hướng gia tăng trong việc ứng dụng dữ liệu thay thế để mở rộng khả năng tiếp cận tín dụng, đặc biệt tại các quốc gia đang phát triển. Trong nghiên cứu này, các tác giả chỉ ra rằng hành vi sử dụng dữ liệu viễn thông, hóa đơn tiện ích, sử dụng thiết bị di động, lịch sử di chuyển cư trú, các hồ sơ về việc trễ hạn thanh toán trong mô hình đánh giá tín dụng đóng vai trò quan trọng trong việc đánh giá hành vi của khách hàng. Những đặc điểm hành vi này phản ánh mức độ ổn định và đáng tin cậy tài chính của người vay, đặc biệt hữu ích trong việc đánh giá tín dụng ở các thị trường mới nổi, nơi dữ liệu tín dụng truyền thống còn hạn chế.

### ***2.4 Thực trạng nghiên cứu tại Việt Nam***

Tại Việt Nam, nghiên cứu liên quan đến đánh giá tín dụng còn tương đối mới. Một số nghiên cứu đáng chú ý có thể kể đến như Huyen & Kleimeier (2007) đề xuất mô hình đánh giá tín dụng sử dụng dữ liệu truyền thống, Tran & cộng sự (2022) ứng dụng các kỹ thuật học máy để dự đoán điểm tín dụng và Gardner & Lainez (2023) nhấn mạnh vai trò tiên phong của Việt Nam trong ứng dụng tín dụng thay thế ở Đông Nam Á. Tuy nhiên,

các nghiên cứu hiện nay chủ yếu mang tính khái quát, chưa đi sâu phân tích các dạng dữ liệu thay thế cụ thể cũng như đánh giá hiệu quả của từng loại dữ liệu đối với khả năng dự báo rủi ro tín dụng, đặc biệt áp dụng cho đối tượng khách hàng cá nhân.

### **2.5 Đóng góp mới của nghiên cứu**

Nghiên cứu nhằm bổ sung và hoàn thiện những hạn chế còn tồn tại trong các nghiên cứu trước đó, cụ thể là tổng hợp cơ sở lý luận và thực tiễn về việc sử dụng dữ liệu thay thế trong chấm điểm tín dụng cho khách hàng cá nhân trên thế giới và tại Việt Nam. Thêm vào đó, nội dung trọng tâm là khai thác dữ liệu thay thế từ cơ sở dữ liệu quốc gia về dân cư - một nguồn dữ liệu mới, tiềm năng và chưa được nghiên cứu sâu. Trên cơ sở đó, nghiên cứu tiến hành xây dựng mô hình thử nghiệm sử dụng dữ liệu thay thế để đánh giá tín dụng cá nhân tại Việt Nam trên cơ sở áp dụng các phương pháp học máy hiện đại.

## **3. Phương pháp nghiên cứu**

### **3.1 Tổng quan phương pháp nghiên cứu**

Trước bối cảnh toàn cầu đang chuyển mình mạnh mẽ trong lĩnh vực đánh giá tín dụng nhờ vào sự phát triển của dữ liệu thay thế, kỹ thuật học máy và dữ liệu lớn, Việt Nam cũng không nằm ngoài xu thế này (Gardner & Lainez, 2023; Lainez, 2021; Lainez & cộng sự, 2021). Việc tiếp cận công nghệ mới nhằm cải thiện khả năng bao phủ tín dụng và thúc đẩy tài chính toàn diện đang được các tổ chức tín dụng trong nước tích cực triển khai. Nghiên cứu chọn Việt Nam làm trường hợp điển hình để phân tích tiềm năng ứng dụng dữ liệu thay thế trong chấm điểm tín dụng cá nhân, đặc biệt cho nhóm khách hàng chưa hoặc có ít lịch sử giao dịch tín dụng.

Sau khi sử dụng phương pháp nghiên cứu tại bàn nhằm tổng hợp cơ sở lý luận và thực tiễn về sử dụng dữ liệu thay thế trong chấm điểm tín dụng khách hàng cá nhân trên thế giới và tại Việt Nam, phương pháp nghiên cứu định lượng được sử dụng để xây dựng mô hình hồi quy ước tính mối quan hệ giữa các biến dữ liệu thay thế, điểm tín dụng cá nhân trên cơ sở kết hợp các thuật toán học máy. Tiếp đó, các kỹ thuật thống kê được sử dụng để phân tích dữ liệu, kiểm tra ý nghĩa của các mối quan hệ và đánh giá hiệu quả của mô hình và kiểm định khả năng áp dụng của mô hình vào thực tiễn. Để đánh giá độ phù hợp và khả năng dự đoán của mô hình, các chỉ số được áp dụng như chỉ số Gini đo lường khả năng phân biệt giữa các khách hàng có rủi ro tín dụng khác nhau và chỉ số KS (Kolmogorov-Smirnov) đo lường khả năng phân biệt giữa khách hàng tốt và khách hàng xấu tại các ngưỡng khác nhau.

### **3.2 Dữ liệu nghiên cứu**

Nghiên cứu sử dụng dữ liệu được trích xuất từ cơ sở dữ liệu quốc gia về dân cư và dữ liệu lịch sử tín dụng từ các tổ chức tài chính tại Việt Nam nhằm xây dựng và kiểm định mô hình đánh giá tín dụng cá nhân RBCredit. Đây là nguồn dữ liệu chính thống, có tính bao phủ toàn dân và tiềm năng cao trong việc phục vụ các giải pháp tài chính đổi mới. Nguồn dữ liệu này đóng vai trò then chốt trong việc xây dựng và kiểm định độ tin cậy của mô hình đánh giá tín dụng. Thêm vào đó, thông tin nhân khẩu học của khách hàng được thu thập, bao gồm độ tuổi, giới tính, địa chỉ cư trú và tình trạng hôn nhân, nhằm phục vụ quá trình phân tích và xây dựng mô hình.

Tính đến năm 2024, cơ sở dữ liệu quốc gia về dân cư đã thu thập được hơn 104 triệu

hồ sơ công dân (mẫu DC01) với 17 trường thông tin, bao gồm các yếu tố định danh và nhân khẩu học như họ tên, ngày sinh, giới tính, tình trạng hôn nhân, địa chỉ cư trú, thông tin về cha mẹ, vợ/chồng, người đứng đầu hộ gia đình (Công an nhân dân, 2024). Nguồn dữ liệu này đóng vai trò trung tâm trong đề án 06 của Chính phủ về chuyển đổi số quốc gia (Công thông tin điện tử Chính phủ, 2022) và đang được khai thác để nâng cao hiệu quả quản lý nhà nước cũng như phát triển các ứng dụng kinh tế - xã hội.

Dữ liệu sử dụng trong mô hình RBCredit bao gồm hơn 120.000 khoản vay cá nhân, cụ thể vay tiêu dùng, vay tiền mặt, thẻ tín dụng, được thu thập từ các tổ chức tín dụng trong giai đoạn 2020-2024. Các đặc trưng đầu vào của mô hình được trích xuất từ dữ liệu dân cư, chủ yếu là các đặc điểm nhân khẩu học của người vay và người thân, bao gồm độ tuổi, giới tính, tình trạng hôn nhân, địa chỉ và khu vực cư trú như nông thôn hoặc đô thị. Qua phân tích yếu tố và đánh giá mức độ ảnh hưởng, các biến có sức dự báo mạnh nhất được xác định là giới tính, nhóm tuổi, tình trạng hôn nhân và khu vực sinh sống.

Việc khai thác và xử lý dữ liệu được thực hiện dưới sự phối hợp giữa Trung tâm nghiên cứu, ứng dụng dữ liệu dân cư và căn cước công dân (RAR center) – thuộc Trung tâm dữ liệu quốc gia về dân cư, Cục C06 - Bộ Công an và trường Công nghệ Thông tin và Truyền thông, Đại học Bách Khoa Hà Nội. Mục tiêu là phát triển mô hình RBCredit như một giải pháp bổ sung đáng tin cậy cho hệ thống chấm điểm tín dụng hiện có, đặc biệt đối với nhóm khách hàng chưa có lịch sử tín dụng chính thức (Phạm, 2023).

Bên cạnh việc khai thác dữ liệu từ cơ sở dữ liệu quốc gia về dân cư, nghiên cứu cũng tiến hành tổng hợp và phân tích các nguồn cung cấp dữ liệu thay thế đang hoạt động tại thị trường Việt Nam. Đây là các đơn vị đóng vai trò quan trọng trong phục vụ chấm điểm tín dụng cá nhân. Cụ thể, Trung tâm Thông tin tín dụng quốc gia Việt Nam (CIC), cơ quan thuộc Ngân hàng Nhà nước Việt Nam, hiện là các tổ chức chính thống lưu trữ và cung cấp nền tảng thông tin tín dụng truyền thống của quốc gia. Thêm vào đó, Công ty cổ phần Thông tin tín dụng Việt Nam (PCB) hoạt động như một tổ chức tư nhân chuyên cung cấp dịch vụ thông tin tín dụng phục vụ nhu cầu đánh giá rủi ro tín dụng của các ngân hàng và tổ chức tài chính.

Ngoài ra, dữ liệu thay thế còn được cung cấp bởi các nhà mạng viễn thông lớn như Viettel, Mobifone và Vinaphone, thông qua thông tin về hành vi sử dụng điện thoại, vị trí địa lý và lịch sử giao dịch kỹ thuật số của người dùng. Các công ty công nghệ tài chính (Fintech) tiên phong như Trusting Social, Data Nest, Kalapa là những nhà cung cấp giải pháp dữ liệu và phân tích chuyên sâu, tận dụng dữ liệu phi truyền thống từ mạng xã hội, hành vi tiêu dùng và các nền tảng số khác để xây dựng mô hình đánh giá tín dụng mới, hỗ trợ nhóm khách hàng chưa hoặc có rất ít lịch sử tín dụng chính thức.

### **3.3 Mô hình nghiên cứu**

Để xây dựng mô hình chấm điểm tín dụng cá nhân sử dụng dữ liệu thay thế, nghiên cứu sử dụng mô hình truyền thống hồi quy logistic và các kỹ thuật học máy hiện đại phổ biến như Random Forest, XGBoost và CatBoost để so sánh hiệu quả cao trong lĩnh vực phân tích dữ liệu tài chính.

Mô hình hồi quy tuyến tính Logistic để ước tính xác suất khách hàng không vỡ nợ dựa trên các đặc điểm nhân khẩu học được trích xuất từ cơ sở dữ liệu quốc gia về dân cư. Mô hình nghiên cứu được đề xuất như sau:

$$\text{Logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

trong đó,  $P$  là xác suất khách hàng không vỡ nợ;  $X_1, X_2, \dots, X_n$  là các biến độc lập, đại diện cho các đặc điểm của dữ liệu thay thế;  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  là các hệ số hồi quy, thể hiện mức độ của mối quan hệ giữa từng biến độc lập và điểm tín dụng.

Đề đo lường biến  $P$ , khách hàng được coi là «vỡ nợ» nếu khoản vay của họ có số ngày quá hạn thanh toán ít nhất 60 ngày trong vòng sáu kỳ thanh toán hàng tháng đầu tiên. Điểm tín dụng được gán giá trị 1 cho khách hàng không vỡ nợ và 0 cho khách hàng vỡ nợ. Dữ liệu thay thế bao gồm tuổi của khách hàng, được đo bằng số năm; giới tính của khách hàng, được mã hóa là 1 cho nam và 0 cho nữ; địa chỉ cư trú của khách hàng, được mã hóa thành các khu vực địa lý khác nhau; tình trạng hôn nhân của khách hàng, được mã hóa là 1 cho đã kết hôn và 0 cho chưa kết hôn.

Random Forest là một thuật toán học tập tổng hợp mạnh mẽ, hoạt động bằng cách xây dựng và kết hợp dự đoán từ một số lượng lớn các cây quyết định. Trong quá trình huấn luyện, mỗi cây quyết định được tạo ra dựa trên một mẫu ngẫu nhiên của dữ liệu và đưa ra dự đoán riêng lẻ. Để đưa ra được dự đoán cuối cùng, Random Forest tổng hợp kết quả từ tất cả các cây, thường bằng cách lấy trung bình đối với các bài toán hồi quy hoặc lấy phiếu đa số đối với các bài toán phân loại. Một trong những ưu điểm nổi bật của Random Forest là khả năng chống chịu tốt với nhiễu và các giá trị ngoại lệ trong dữ liệu. Kết quả cuối cùng được tổng hợp từ nhiều cây quyết định, nên tác động của bất kỳ cây nào bị ảnh hưởng bởi nhiễu sẽ được giảm thiểu. Thêm vào đó, Random Forest cũng cung cấp một khả năng đánh giá tầm quan trọng của các biến đầu vào, giúp hiểu rõ hơn về mức độ ảnh hưởng của từng yếu tố đến điểm tín dụng (Breiman, 2001).

XGBoost (Extreme Gradient Boosting) là một thuật toán sử dụng cây quyết định. Tuy nhiên, khác với Random Forest xây dựng các cây độc lập, XGBoost xây dựng chúng theo thứ tự, trong đó cây sau ‘học hỏi’ từ những sai lầm của cây trước. Cách thức hoạt động này tương tự như việc học một kỹ năng mới. Mỗi cây mới trong XGBoost tập trung vào việc sửa các lỗi mà các cây trước đó đã mắc phải, giúp mô hình ngày càng chính xác hơn. XGBoost cũng tích hợp các ‘công cụ’ đặc biệt để tránh tình trạng ‘học tử’ (quá khớp), từ đó nâng cao khả năng hoạt động tốt trên dữ liệu mới (IBM, 2024).

CatBoost (Category Boosting) là một thuật toán khác cũng sử dụng cây quyết định và học hỏi từ những sai lầm, tương tự như XGBoost. Tuy nhiên, điểm đặc biệt của CatBoost là được thiết kế tối ưu cho việc xử lý dữ liệu ‘danh mục’. Dữ liệu ‘danh mục’ bao gồm các biến được chia thành các nhóm hoặc loại như màu sắc (đỏ, xanh, vàng), loại hình nhà ở (chung cư, nhà riêng, biệt thự) hoặc tình trạng hôn nhân (độc thân, đã kết hôn, ly hôn). CatBoost có các phương pháp xử lý thông minh để làm việc trực tiếp với dữ liệu ‘danh mục’ mà không cần chuyển đổi chúng thành dạng số, giúp tăng hiệu quả của mô hình và giảm thời gian cần thiết cho việc chuẩn bị dữ liệu (Medium, 2024).

Bốn thuật toán này đã có trong thư viện Scikit-Learn. Dữ liệu sau khi làm sạch sẽ được dùng đưa vào làm đầu vào của thuật toán để xây dựng mô hình. Tham số mặc định của thuật toán sẽ được dùng để huấn luyện.

### **3.4 Quy trình xây dựng mô hình chấm điểm tín dụng**

Quy trình xây dựng mô hình được triển khai theo 5 bước. Bước 1 là lựa chọn biến đầu

vào. Các biến đầu vào từ cơ sở dữ liệu quốc gia về dân cư có liên quan đến khả năng vỡ nợ của khách hàng dựa trên phân tích tương quan và ý kiến chuyên gia. Bước 2 là tiền xử lý dữ liệu. Xử lý các giá trị thiếu, loại bỏ các giá trị ngoại lệ và mã hóa các biến phân loại. Bước 3 là huấn luyện mô hình, các thuật toán được huấn luyện trên tập dữ liệu huấn luyện. Quá trình huấn luyện sử dụng kỹ thuật điều chỉnh siêu tham số nhằm tối ưu hóa hiệu quả dự đoán. Bước 4 là đánh giá mô hình. Mô hình được kiểm định trên tập dữ liệu kiểm tra và đánh giá bằng chỉ số Gini và Kolmogorov-Smirnov (KS). Cuối cùng, bước 5 sẽ biểu diễn mô hình tổng quát.

Nghiên cứu xây dựng mô hình tổng quát nhằm đánh giá khả năng tín dụng cá nhân dựa trên dữ liệu nhân khẩu học từ cơ sở dữ liệu quốc gia về dân cư. Để đảm bảo độ tin cậy và so sánh hiệu quả dự báo, bốn mô hình học máy và thống kê được triển khai gồm hồi quy Logistic, Random Forest, XGBoost và CatBoost. Mô hình tổng quát được biểu diễn như sau:

$$Score = f(X_1, X_2, \dots, X_n)$$

trong đó, *Score* là điểm tín dụng dự đoán (hoặc xác suất không vỡ nợ); *f* là hàm dự đoán được xây dựng bởi từng mô hình (mô hình hồi quy tuyến tính Logistic, Random Forest, XGBoost, CatBoost);  $X_1, X_2, \dots, X_n$  là các biến đầu vào từ cơ sở dữ liệu quốc gia về dân cư (đặc điểm nhân khẩu học như tuổi, giới tính, tình trạng hôn nhân).

Việc so sánh kết quả giữa các mô hình dựa trên các chỉ số đánh giá như Gini và Kolmogorov-Smirnov, giúp lựa chọn mô hình có khả năng phân biệt rủi ro tốt nhất và ổn định theo thời gian.

### 3.5 Đánh giá hiệu suất mô hình

Mô hình RBCredit được đánh giá qua hai bước. Bước một là đánh giá nội bộ (internal validation). Dữ liệu giai đoạn 2020-2023 được chia ngẫu nhiên thành 5 phần bằng nhau để thực hiện quy trình kiểm định chéo 5 lần, phục vụ tính hiệu suất trung bình, đánh giá khả năng tổng quát hóa. Sau đó, mô hình được kiểm định với dữ liệu năm 2024 để đánh giá độ chính xác và ổn định. Bước hai là đánh giá độc lập (external validation). Dữ liệu năm 2024 từ cơ sở dữ liệu quốc gia về dân cư và dữ liệu lịch sử tín dụng thực tế thu thập từ các tổ chức tài chính như FE Credit, Home Credit, Mcredit; BIDV, VIB, NCB, PVCombank đã hoàn thành quá trình đánh giá. Các ngân hàng lớn đang chờ đánh giá như Vietcombank, Vietinbank, Agribank, MBBank, Techcombank, TPBank, HDBank, Wooribank và LPBank, được sử dụng để kiểm tra độc lập chất lượng của mô hình đã xây dựng.

Theo Siddiqi (2012), để đánh giá chất lượng của mô hình đánh giá tín dụng, chỉ số Gini cho độ chính xác và PSI cho sự ổn định của mô hình được sử dụng. Gini là một chỉ số được sử dụng để đánh giá độ phân biệt của mô hình phân loại, thường dùng trong các mô hình dự đoán khả năng vỡ nợ của khách hàng. Chỉ số này càng cao thì mô hình phân biệt càng tốt giữa các nhóm khách hàng có khả năng vỡ nợ và không vỡ nợ. PSI là một chỉ số đo lường sự ổn định của mô hình theo thời gian, bằng cách so sánh sự thay đổi trong phân phối của kết quả dự đoán giữa các giai đoạn khác nhau.

## 4. Kết quả và thảo luận

### 4.1 Kết quả đánh giá nội bộ

Kết quả kiểm định nội bộ cho thấy chỉ số Gini ổn định giữa 5 lần huấn luyện (nhỏ

hơn 2%). Sai khác chỉ số Gini giữa các tập dữ liệu huấn luyện từ năm 2020 đến 2023 và tập kiểm tra năm 2024 là khá nhỏ (dưới 2%) trong tất cả 5 lần huấn luyện mô hình. Chỉ số PSI trong suốt quá trình huấn luyện đều duy trì mức thấp (dưới 10%), khẳng định mô hình có tính ổn định cao theo thời gian. Ngoài ra, kết quả thực nghiệm cũng chỉ ra rằng chỉ số Gini dao động từ 30% - 55% tùy thuộc vào từng sản phẩm tài chính và phân khúc khách hàng. Cụ thể, chỉ số Gini cao nhất được ghi nhận ở các sản phẩm vay tiêu dùng và vay xe máy, trong khi chỉ số Gini thấp nhất xuất hiện đối với thẻ tín dụng và các khoản vay tiền mặt có giá trị trên 100 triệu đồng.

#### 4.2 Kết quả đánh giá độc lập

Các tổ chức tài chính đã cung cấp tập dữ liệu từ 10.000 đến 50.000 khoản vay để nhóm phát triển mô hình chấm điểm, không được biết trước nhãn vỡ nợ/không vỡ nợ của các khoản vay này. Sau đó, các tổ chức tài chính thực hiện đánh giá tính chính xác và độ ổn định của mô hình dựa trên dữ liệu thực tế.

Các chỉ số trong Bảng 1, Bảng 2 và Bảng 3 được tính toán dựa trên điểm tín dụng do mô hình RBCredit. Mô hình này được phát triển bằng thuật toán CatBoost, với dữ liệu đầu vào từ cơ sở dữ liệu quốc gia về dân cư và dữ liệu lịch sử tín dụng thực tế thu thập từ các tổ chức tài chính. Để tính toán tỷ lệ vỡ nợ và không vỡ nợ, một khoản vay được coi là vỡ nợ là khoản vay có số ngày quá hạn (DPD) đạt ít nhất 60 trong sáu kỳ thanh toán đầu tiên.

Mỗi quan sát trong tập dữ liệu đánh giá tương ứng với một khoản vay, được mô hình RBCredit chấm một điểm tín dụng nằm trong thang điểm từ 300 đến 850. Việc chia khoảng điểm tín dụng được thực hiện theo hai phương pháp là chia theo phân tích phân vị nhằm đảm bảo số lượng quan sát tương đối đồng đều giữa các khoảng, hoặc chia theo các mức điểm dễ phân biệt để thuận tiện cho việc phân tích. Việc xác định cách chia này nhằm đảm bảo rằng mỗi khoảng điểm có đủ số lượng quan sát cần thiết để tính toán tỷ lệ vỡ nợ và không vỡ nợ một cách ổn định và có ý nghĩa thống kê.

Việc tính toán các tỷ lệ trong các bảng kết quả được thực hiện như sau:

$$\text{Tỷ lệ vỡ nợ trong khoảng điểm} = (\text{Số quan sát vỡ nợ trong khoảng điểm}) / (\text{Tổng số quan sát trong khoảng điểm}) \times 100\%.$$

$$\text{Tỷ lệ không vỡ nợ trong khoảng điểm} = (\text{Số quan sát không vỡ nợ trong khoảng điểm}) / (\text{Tổng số quan sát trong khoảng điểm}) \times 100\%.$$

Tổng tích lũy tỷ lệ được tính bằng cách cộng dồn tỷ lệ từ khoảng đầu tiên cho đến khoảng đang xét, nhằm phản ánh xu hướng tích lũy rủi ro hoặc không rủi ro theo điểm tín dụng.

**Bảng 1. Phân bố tỷ lệ vỡ nợ và không vỡ nợ đối các khoản vay không có tài sản đảm bảo và khách hàng có hồ sơ tín dụng mỏng theo các khoảng điểm liên tiếp**

TT	Khoảng điểm	Tỷ lệ không vỡ nợ (%)	Tổng tích lũy tỷ lệ không vỡ nợ (%)	Tỷ lệ vỡ nợ (%)	Tổng tích lũy tỷ lệ vỡ nợ (%)
1	350 - 568	5,3	5,3	24,7	24,7
2	569 - 670	7,7	13,0	16,8	41,5

**Bảng 1. Phân bố tỷ lệ vỡ nợ và không vỡ nợ đối các khoản vay không có tài sản đảm bảo và khách hàng có hồ sơ tín dụng mỏng theo các khoảng điểm liên tiếp (tiếp theo)**

TT	Khoảng điểm	Tỷ lệ không vỡ nợ (%)	Tổng tích lũy tỷ lệ không vỡ nợ (%)	Tỷ lệ vỡ nợ (%)	Tổng tích lũy tỷ lệ vỡ nợ (%)
3	671 - 724	8,7	21,7	14,8	56,3
4	725 - 757	9,1	30,7	12,6	68,9
5	758 - 780	9,9	40,6	10,4	79,3
6	781 - 799	10,6	51,2	7,9	87,1
7	800 - 815	12,1	63,3	4,1	91,2
8	816 - 825	11,6	74,9	3,8	95,0
9	825 - 836	12,2	87,0	2,8	97,8
10	836 - 850	13,0	100,0	2,2	100,0

*Nguồn: Tính toán của nhóm tác giả*

**Bảng 2. Phân bố tỷ lệ các trường hợp vỡ nợ và không vỡ nợ cho sản phẩm thẻ tín dụng và khách hàng có hồ sơ mỏng theo các khoảng điểm liên tiếp**

TT	Khoảng điểm	Tỷ lệ không vỡ nợ (%)	Tổng tích lũy tỷ lệ không vỡ nợ (%)	Tỷ lệ vỡ nợ (%)	Tổng tích lũy tỷ lệ vỡ nợ (%)
1	300 - 660	8,2	8,2	29,6	29,6
2	661 - 724	9,4	17,6	14,6	44,2
3	724 - 755	9,7	27,3	12,3	56,5
4	756 - 773	9,8	37,0	10,8	67,3
5	774 - 789	9,9	46,9	8,3	75,6
6	790 - 803	10,3	57,1	6,7	82,3
7	804 - 815	10,4	67,5	5,3	87,6
8	816 - 825	10,5	78,0	6,2	93,8
9	826 - 834	10,5	88,4	4,4	98,2
10	835 - 849	11,6	100,0	1,8	100,0

*Nguồn: Tính toán của nhóm tác giả*

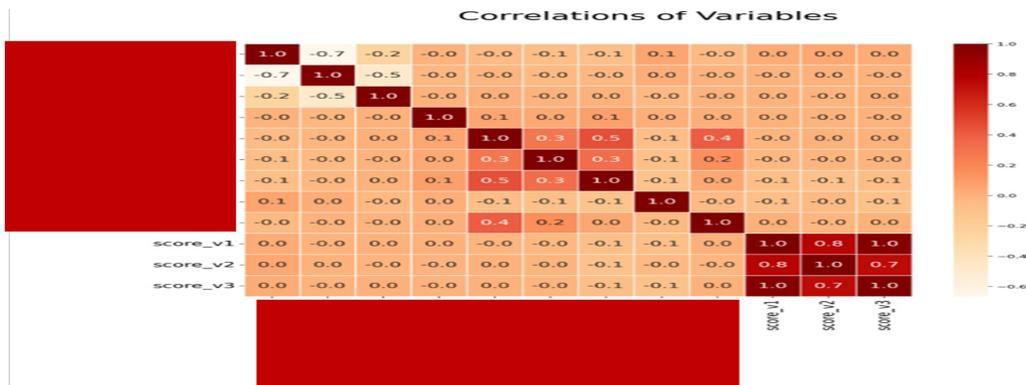
**Bảng 3. Phân bố tỷ lệ các trường hợp vỡ nợ và không vỡ nợ cho sản phẩm thẻ tín dụng theo các khoảng điểm liên tiếp**

TT	Khoảng điểm	Tỷ lệ không vỡ nợ (%)	Tổng tích lũy tỷ lệ không vỡ nợ (%)	Tỷ lệ vỡ nợ (%)	Tổng tích lũy tỷ lệ vỡ nợ (%)
1	300 - 638	8,7	8,7	26,4	26,4
2	661 - 718	9,7	18,4	14,2	40,6
3	719 - 752	9,8	28,2	11,9	52,5
4	753 - 777	10,1	38,3	9,5	62,0
5	778 - 795	10,2	48,5	8,2	70,3
6	796 - 809	10,0	58,5	7,1	77,4
7	810 - 820	10,0	68,4	6,5	83,9
8	821 - 830	10,7	79,1	5,9	89,7
9	831 - 839	10,0	89,2	5,7	95,4
10	840 - 850	10,9	100,0	4,6	100,0

*Nguồn: Tính toán của nhóm tác giả*

Kết quả nghiên cứu đánh giá độc lập chỉ ra khả năng phân biệt khách hàng tốt/xấu. Mô hình RBCredit có khả năng phân biệt giữa khách hàng tốt (không vỡ nợ) và khách hàng xấu (vỡ nợ), với chỉ số Gini dao động từ 30% đến 52% (mô hình tín dụng có chỉ số Gini trên 20% đã có tính ứng dụng). Ngoài ra, Bảng 1, Bảng 2 và Bảng 3 trình bày kết quả đánh giá độc lập trên các phân khúc khách hàng và sản phẩm khác nhau. Trong đó, cột “tỷ lệ không vỡ nợ (%)” và “tỷ lệ vỡ nợ (%)” lần lượt thể hiện tỷ lệ phần trăm các khoản vay không vỡ nợ và vỡ nợ, tương ứng với từng khoảng điểm tín dụng. Các kết quả cho thấy xu hướng nhất quán, khi điểm tín dụng RBCredit tăng, tỷ lệ vỡ nợ giảm rõ rệt, trong khi tỷ lệ không vỡ nợ tăng lên, khẳng định khả năng phân loại khách hàng tốt/xấu của mô hình.

Ngoài ra, điểm tín dụng RBCredit khi được tích hợp vào mô hình nội bộ cũng làm tăng hiệu quả đáng kể. Việc sử dụng điểm tín dụng RBCredit như một đặc trưng trong các mô hình chấm điểm tín dụng của các tổ chức tài chính giúp tăng chỉ số chính xác Gini thêm từ 3% đến 8%. Hình 1 minh họa kết quả phân tích tương quan giữa điểm RBCredit (ba phiên bản) với điểm tín dụng nội bộ của một công ty tài chính lớn, cho thấy mức độ tương quan thấp. Điều này hàm ý rằng điểm RBCredit mang lại thông tin bổ sung có giá trị, từ đó giúp cải thiện đáng kể hiệu quả của mô hình tín dụng nội bộ khi được kết hợp.



**Hình 1. Giá trị tương quan giữa ba phiên bản điểm RBCredit với điểm nội bộ**

*Nguồn: Tính toán của nhóm tác giả*

Về sự khác biệt giữa các nhóm tổ chức tài chính và ngân hàng, kết quả đánh giá độ lập cho thấy mô hình RBCredit đạt chỉ số Gini cao hơn khi áp dụng cho các công ty tài chính so với các ngân hàng thương mại, phản ánh mức độ phù hợp cao hơn với đặc thù danh mục khách hàng của các công ty tài chính.

Về hiệu quả theo phân khúc sản phẩm và khách hàng, mô hình RBCredit cho kết quả đánh giá cao hơn đối với các sản phẩm vay tiêu dùng, khoản vay có giá trị nhỏ (dưới 50 triệu đồng) và các nhóm khách hàng có hồ sơ tín dụng mỏng (những khách hàng chưa có hoặc có rất ít lịch sử tín dụng trong hệ thống tài chính).

Các kết quả trên cho thấy dữ liệu từ cơ sở dữ liệu quốc gia về dân cư đóng vai trò quan trọng trong hoạt động đánh giá tín dụng trong lĩnh vực tài chính. Bên cạnh việc nâng cao khả năng phân loại rủi ro tín dụng, việc ứng dụng dữ liệu này góp phần mở rộng tỷ lệ tiếp cận tài chính chính thống cho người dân, qua đó gián tiếp giảm thiểu tình trạng tín dụng đen và góp phần đảm bảo an sinh xã hội.

## 5. Hàm ý chính sách và kết luận

Nghiên cứu tập trung phân tích vai trò của dữ liệu thay thế trong việc nâng cao hiệu quả mô hình chấm điểm tín dụng cá nhân tại Việt Nam. Kết quả cho thấy so với dữ liệu truyền thống, dữ liệu thay thế như dữ liệu viễn thông, hành vi tiêu dùng, mạng xã hội có thể cải thiện đáng kể khả năng phân loại khách hàng, đặc biệt đối với nhóm chưa có hoặc có rất ít lịch sử tín dụng.

Việc ứng dụng dữ liệu thay thế mang lại nhiều triển vọng cho các tổ chức tài chính, giúp mở rộng khả năng tiếp cận tín dụng, thúc đẩy tài chính toàn diện và quản trị rủi ro hiệu quả hơn. Tuy nhiên, khai thác dữ liệu này đòi hỏi năng lực xử lý dữ liệu lớn, kỹ thuật tiền xử lý phù hợp và các thuật toán phức tạp vốn có hạn chế là thiếu tính minh bạch và khó giải thích.

Bên cạnh đó, để dữ liệu thay thế thực sự phát huy giá trị, cần có sự đồng bộ về khung pháp lý, cơ chế bảo vệ quyền riêng tư và cơ sở hạ tầng dữ liệu. Trong bối cảnh Việt Nam

chuyên mình mạnh mẽ sang nền tài chính số, các kết quả nghiên cứu có thể đóng góp thiết thực cho tổ chức tín dụng và nhà hoạch định chính sách trong việc thiết kế các mô hình chấm điểm tín dụng hiện đại, toàn diện và công bằng hơn.

Từ kết quả thực nghiệm, một số hàm ý quản trị được đưa ra nhằm thúc đẩy ứng dụng hiệu quả dữ liệu thay thế trong chấm điểm tín dụng cá nhân tại Việt Nam. Trước hết, các tổ chức tín dụng cần chủ động tích hợp dữ liệu thay thế vào mô hình chấm điểm tín dụng, đặc biệt với khách hàng chưa có lịch sử tín dụng. Việc đầu tư vào hạ tầng công nghệ, quản trị dữ liệu và năng lực phân tích là cần thiết. Các mô hình nên kết hợp học máy hiện đại với phương pháp truyền thống để cân bằng giữa độ chính xác và khả năng giải thích. Về phía cơ quan quản lý, cần hoàn thiện khung pháp lý cho việc thu thập, chia sẻ và sử dụng dữ liệu thay thế, đảm bảo tuân thủ quy định về bảo mật và quyền riêng tư. Đồng thời, khuyến khích cơ chế chia sẻ dữ liệu an toàn giữa các bên và xây dựng hệ thống giám sát chất lượng mô hình. Các doanh nghiệp cung cấp dữ liệu như công ty viễn thông, nền tảng số hay công nghệ tài chính số cần chuẩn hóa dữ liệu và phối hợp với tổ chức tín dụng để cung cấp chỉ số hành vi giá trị. Cuối cùng, các nhà nghiên cứu và chuyên gia trong lĩnh vực tài chính cần tiếp tục thực hiện các nghiên cứu thực nghiệm tại Việt Nam nhằm kiểm chứng hiệu quả và phát triển các mô hình tín dụng vừa đảm bảo độ chính xác, vừa đáp ứng yêu cầu về tính giải thích.

### **Tài liệu tham khảo**

- Abdou, H.A. & Pointon, J. (2011), “Credit scoring, statistical techniques and evaluation criteria: a review of the literature”, *Intelligent Systems in Accounting, Finance and Management*, Vol. 18, pp. 59-88.
- Bennouna, G. & Tairi, M. (2019), “Scoring in microfinance: credit risk management tool -Case of Morocco”, *Procedia Computer Science*, Vol. 148, pp. 522-531.
- Breiman, L. (2001), “Random Forests”, *Machine Learning*, Vol. 45, pp. 5-32.
- Chase (2024), “Do other countries have credit scores?”, <https://www.chase.com>, truy cập ngày 20/04/2025.
- Clean Credit (2017), “How different is AUS credit reporting to US credit reporting?”, <https://www.cleancredit.com.au/different-aus-credit-reporting-us-credit-reporting>, truy cập ngày 20/04/2025.
- Công an nhân dân (2024), “Cơ sở dữ liệu quốc gia về dân cư là “trái tim” của chuyển đổi số”, <https://cand.com.vn/su-kien-binh-luan-thoi-su/co-so-du-lieu-quoc-gia-ve-dan-cu-la-trai-tim-cua-chuyen-doi-so-i730549>, truy cập ngày 20/4/2025.
- Công Thông tin điện tử Chính phủ (2022), “Quyết định số 06/QĐ-TTg: Phê duyệt Đề án phát triển ứng dụng dữ liệu về dân cư, định danh và xác thực điện tử”, <https://vanban.chinhphu.vn/?pageid=27160&docid=205022>, truy cập ngày 20/4/2025.
- Edtech Agency (2024), “Việt Nam hướng đến 100% người dùng có điện thoại thông minh vào cuối năm 2024”, <https://edtechagency.net/vi/viet-nam-huong-den-100-nguoi-dung-co-dien-thoai-thong-minh-vao-cuoi-nam-2024>, truy cập ngày 18/04/2025.
- FICO Blog (2024), “How to use alternative data in credit risk analytics”, <https://www.fico.com/blogs/how-use-alternative-data-credit-risk-analytics>, truy cập ngày

18/04/2025.

- Gardner, J. & Lainez, N. (2023), “Algorithmic credit scoring in Vietnam: a legal proposal for maximizing benefits and minimizing risks”, *Asian Journal of Law and Society*, Vol. 10, No. 3, pp. 401-432.
- Hiệp hội Ngân hàng Việt Nam (2023), “Hiệp hội Ngân hàng Việt Nam đề xuất nhiều giải pháp hạn chế tín dụng đen”, <https://vnba.org.vn/vi/hiep-hoi-ngan-hang-viet-nam-de-xuat-nhieu-giai-phap-han-che--tin-dung-den-10775.htm>, truy cập ngày 10/4/2025.
- Homonoff, T., O’Brien, R. & Sussman, A.B. (2021), “Does knowing your FICO score change financial behavior? Evidence from a field experiment with student Loan borrowers”, *The Review of Economics and Statistics*, Vol. 103, No. 2, pp. 236-250.
- Hoàng, L. (2023), “Ba vấn đề then chốt trên thị trường cho vay tiêu dùng”, *VnEconomy*, <https://vneconomy.vn/ba-van-de-then-chot-tren-thi-truong-cho-vay-tieu-dung.htm>, truy cập ngày 10/4/2025.
- Huyen, D.T.T. & Kleimeier, S. (2007), “A credit scoring model for Vietnam’s retail banking market”, *International Review of Financial Analysis*, Vol. 26, No. 5, pp. 471-495.
- IBM (2024), “What is XGBoost?”, <https://www.ibm.com/think/topics/xgboost>, truy cập ngày 21/04/2025.
- Kiviat, B. (2019), “Credit Scoring in the United States”, *Economic Sociology: The European Electronic Newsletter*, Vol. 21, No. 1, pp. 33-42.
- Lainez, N. (2021), “The prospects and dangers of algorithmic credit scoring in Vietnam: regulating a legal blindspot”, Economics Working Paper.
- Lainez, N., Bui, T.T.D., Phan, K.T., Le, T.L. & Pannier, E. (2021), “Financial inclusion and consumer finance in Vietnam: challenges and approaches to credit scoring perspective”, *ISEAS Perspective*, 2021, 112, 9 p. ISSN 2335-6677.
- Medium (2024), “Understanding catboost: the gradient boosting algorithm for categorical data”, <https://aravindkolli.medium.com/understanding-catboost-the-gradient-boosting-algorithm-for-categorical-data-73ddb200895d>, truy cập ngày 21/04/2025.
- Njuguna, R. & Sowon, K. (2021), “A scoping review of alternative credit scoring literature”, *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 437-444.
- Phạm, V. (2023), “Ứng dụng dữ liệu dân cư đánh giá khả tín, giúp giảm rủi ro cho vay đến 20%”, *VnEconomy*, <https://vneconomy.vn/ung-dung-du-lieu-dan-cu-danh-gia-kha-tin-giup-giam-rui-ro-cho-vay-den-20.htm>, truy cập ngày 20/04/2025.
- Phan, L. (2024), “Thị trường tài chính tiêu dùng chờ cú hích từ chấm điểm tín dụng”, *VnEconomy*, <https://vneconomy.vn/thi-truong-tai-chinh-tieu-dung-cho-cu-hich-tu-cham-diem-tin-dung.htm>, truy cập ngày 10/4/2025.

- Siddiqi, N. (2012), *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wiley & Sons, Hoboken.
- Spuchl'áková, E., Valášková, K. & Adamko, P. (2015), "The credit risk and its measurement, hedging and monitoring", *Procedia Economics and Finance*, Vol. 24, pp. 675-681.
- Thomas, L.C., Edelman, D.B. & Crook, J.N. (2002), "Readings in Credit Scoring: Foundations, Developments, and Aims", *Oxford University Press*, Oxford.
- Tran, Q., Duong, B.V., Lam, G.H., Vuong, D. & Do, T.H. (2022), "An investigation on vietnamese credit scoring based on big data platform and ensemble learning", *Intelligence of Things: Technologies and Applications*, Vol. 148, pp. 289-298.
- VantageScore (2024), "Credit bureaus + bank data = VantageScore 4 Plus", *Technical Report*, <https://vs4plus.vantagescore.com>, truy cập ngày 18/04/2025.
- Wijaya, T. & Nugroho, M. (2023), "International experiences with innovative credit scoring: lessons for Indonesia", *Centre for Indonesian Policy Studies*, No. 14, 45.
- World Bank (2019), "Credit scoring approach guidelines", <https://www.worldbank.org>, truy cập ngày 18/04/2025.