

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

Vũ Thị Thúy Hằng

Trường Đại học Thương mại, Việt Nam

Ngày nhận: 04/06/2025

Ngày nhận bản sửa: 11/03/2026

Ngày duyệt đăng: 27/03/2026

Tóm tắt: Cho vay ngang hàng (P2P lending) mở rộng tiếp cận tín dụng nhưng cũng đặt ra yêu cầu đánh giá khả năng trả nợ của người vay cá nhân để quản trị rủi ro tín dụng. Nghiên cứu sử dụng 1.048.575 khoản vay cá nhân trên LendingClub (Hoa Kỳ, 2007- 2023) và ước lượng hồi quy logistic nhị phân nhằm nhận diện các yếu tố chi phối xác suất hoàn trả đúng hạn. Kết quả cho thấy lãi suất, tỷ lệ nợ/thu nhập, lịch sử chậm trả, hồ sơ công khai bất lợi, kỳ hạn vay và khoản trả góp hàng tháng có ý nghĩa thống kê; nhóm có thâm niên làm việc trên 10 năm có xu hướng hoàn trả tốt hơn. Mô hình hồi quy logistic cho thấy hiệu suất phân loại tốt, với diện tích dưới đường cong là 0,7814. Nghiên cứu chuẩn hóa biến kết quả trả nợ (loan outcome) từ trường loan_status theo

Determinants of borrower repayment ability on peer-to-peer lending platforms: Evidence from LendingClub data

Abstract: Peer-to-peer (P2P) lending expands access to credit but also necessitates assessing borrowers' repayment capacity to manage credit risk. This research employs 1,048,575 individual loans from LendingClub (USA, 2007- 2023) and utilizes binary logistic regression analysis to ascertain factors affecting the likelihood of prompt repayment. The results show that interest rate, debt-to-income ratio, history of late payments, unfavorable public records, loan term, and monthly installments are statistically significant; the group with more than 10 years of employment tends to have better repayment performance. The logistic regression model shows excellent classification performance, with an area under the curve of 0.7814. This study adjusts the loan outcome variable from the loan_status field using clear, consistent rules and demonstrates how financial and credit factors affect large, long-term datasets. The findings have ramifications for credit assessment and portfolio administration on P2P platforms. These findings offer portfolio management implications for Vietnamese peer-to-peer lending platforms in the context of the digital transformation of global financial services.

Keywords: Peer-to-peer lending, Debt repayment capacity, Fintech, Credit risk

Link Doi: <https://doi.org/10.59276/JELB.2026.05.3001>

Vu Thi Thuy Hang

Email: vuthuyhang.tmdt@tmu.edu.vn

Organization: Thuongmai University, Viet Nam

quy tắc minh bạch, có thể lặp lại và cung cấp bằng chứng thực nghiệm về vai trò của các yếu tố tài chính, tín dụng trên bộ dữ liệu lớn giai đoạn dài. Kết quả mang lại hàm ý ứng dụng cho hoạt động sàng lọc tín dụng và quản trị danh mục cho vay trên các nền tảng P2P. Các phát hiện này cung cấp những hàm ý ứng dụng cho hoạt động sàng lọc tín dụng và quản trị danh mục trên các nền tảng cho vay ngang hàng của Việt Nam trong bối cảnh chuyển đổi số các dịch vụ tài chính toàn cầu.

Từ khóa: Cho vay ngang hàng, Khả năng trả nợ, Công nghệ tài chính, Rủi ro tín dụng

Trích dẫn: Vũ Thị Thúy Hằng. (2026). Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub. *Tạp chí Kinh tế - Luật và Ngân hàng*, 28(5), 72-85. <https://doi.org/10.59276/JELB.2026.05.3001>

1. Đặt vấn đề

Từ giữa thập niên 2010, công nghệ tài chính (FinTech) thúc đẩy thay đổi mạnh mẽ trong cung ứng dịch vụ tài chính, đặc biệt thông qua nền tảng số và dữ liệu hóa quy trình. Trong bối cảnh đó, cho vay ngang hàng nổi lên như một kênh kết nối trực tiếp người vay và người cho vay, làm giảm chi phí trung gian và mở rộng tiếp cận tín dụng. Các công nghệ số (AI, dữ liệu lớn, học máy...) góp phần tăng tốc xử lý, cá nhân hóa dịch vụ và mở rộng bao trùm tài chính, đồng thời đặt ra yêu cầu cao hơn về quản trị rủi ro dựa trên dữ liệu. Cho vay ngang hàng cho phép người vay và người cho vay kết nối qua nền tảng trực tuyến, giảm vai trò trung gian truyền thống nhưng làm gia tăng vai trò của cơ chế sàng lọc và chấm điểm tín dụng của nền tảng (Ko và cộng sự, 2022; Vu, 2025).

Quy mô thị trường cho vay ngang hàng P2P toàn cầu được tính toán là 176,5 tỷ đô la Mỹ vào năm 2025 và dự kiến sẽ đạt khoảng 1.380,80 tỷ đô la Mỹ vào năm 2034, tương ứng với tốc độ tăng trưởng kép hàng năm (CAGR) là 25,73% từ năm 2025 đến năm 2034 (Precedence Research, 2025). Tại Việt Nam, thị trường này dự kiến tăng lên 824,4 triệu USD vào năm 2028, với tốc

độ tăng trưởng hàng năm kép là 16,9% trong giai đoạn 2024- 2028 (Alpha AM, 2025). Đi cùng với tiềm năng tăng trưởng là những rủi ro tiềm ẩn, đặc biệt là rủi ro tín dụng, khi người vay không có khả năng hoàn trả khoản vay đúng hạn. Do không có tài sản đảm bảo và ít sự can thiệp của cơ quan kiểm soát tín dụng, việc đánh giá khả năng trả nợ của người vay trong các nền tảng cho vay ngang hàng P2P trở nên vô cùng quan trọng.

Mặc dù đã có nhiều nghiên cứu quốc tế về cho vay ngang hàng (trong đó có LendingClub), kết quả còn khác biệt về vai trò của một số biến và chưa thống nhất cách đo lường biến khả năng trả nợ của người vay khi dữ liệu có nhiều trạng thái (ví dụ trạng thái “late” theo thời điểm) (Nigmonov và cộng sự, 2024; Do, 2022; Nguyen & Ngo, 2025). Vì vậy, nghiên cứu này tập trung chuẩn hóa biến khả năng trả nợ theo “loan outcome” từ trường loan_status bằng một quy tắc mã hóa minh bạch, có thể lặp lại, đồng thời kiểm định vai trò của các yếu tố tài chính, tín dụng trên giai đoạn dài. Trên cơ sở đó, nghiên cứu sử dụng hồi quy logistic nhị phân với dữ liệu LendingClub giai đoạn 2007- 2023 để xác định các yếu tố ảnh hưởng đến khả năng trả nợ của người vay trên nền tảng P2P. Để đạt

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay

ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

được mục tiêu trên, bài báo được kết cấu thành 5 phần. Phần 1 giới thiệu và nêu vấn đề nghiên cứu. Phần 2 tổng quan lý thuyết và mô hình nghiên cứu. Phần 3 trình bày dữ liệu và phương pháp. Phần 4 phân tích kết quả và thảo luận. Phần 5 nêu kết luận và hàm ý chính sách.

2. Cơ sở lý thuyết và tổng quan nghiên cứu

2.1. Khả năng trả nợ của khách hàng cá nhân

Khả năng trả nợ của khách hàng cá nhân được định nghĩa là năng lực của người vay trong thực hiện đầy đủ nghĩa vụ vay vốn khi đến hạn (Balogun & Alimi, 1988). Một khoản vay được coi là quá hạn khi người vay chậm thanh toán và được coi là không thực hiện nghĩa vụ đúng hạn, khi khả năng thu hồi trở nên rất thấp theo tiêu chí của nền tảng hoặc điều khoản hợp đồng (CGAP, 1999).

Trong bối cảnh các nền tảng tài chính trực tuyến ít hoặc không yêu cầu tài sản bảo đảm, thông tin người vay phân tán và cơ chế giám sát khác biệt so với ngân hàng truyền thống, việc đánh giá chính xác khả năng trả nợ trở nên đặc biệt quan trọng. Điều này phải phát triển các công cụ định lượng phù hợp nhằm phân tích dựa trên dữ liệu có sẵn từ nền tảng, bao gồm đặc điểm cá nhân, khoản vay, lịch sử tín dụng- những yếu tố vốn khác biệt so với tín dụng ngân hàng, nơi thẩm định thường dựa nhiều vào quan hệ tín dụng, thẩm định tại chỗ.

2.2. Nền tảng cho vay ngang hàng (P2P lending) và bộ dữ liệu LendingClub

Cho vay ngang hàng là một mô hình tài chính trong đó các cá nhân có thể vay và cho vay tiền trực tiếp với nhau thông qua

các nền tảng trực tuyến, không cần sự tham gia của các tổ chức tài chính truyền thống như ngân hàng (Basha và cộng sự, 2021). Các nền tảng cho vay trực tuyến đóng vai trò trung gian giữa người vay và người cho vay, thu phí dịch vụ, mà không có sự hiện diện của ngân hàng truyền thống. Tuy không chịu ràng buộc vốn như ngân hàng, nhưng các nền tảng này vẫn phát sinh chi phí đánh giá khách hàng trước khi khoản vay được chấp thuận và sau đó là chi phí giám sát việc thanh toán cũng như thu hồi khoản vay khi xảy ra rủi ro.

Trong mô hình cho vay ngang hàng, người cho vay cá nhân trực tiếp cung cấp vốn cho người vay thông qua nền tảng trực tuyến, thay vì thông qua các tổ chức tài chính trung gian như ngân hàng (Basha và cộng sự, 2021). Chính vì loại bỏ vai trò của tổ chức tài chính truyền thống nên người cho vay cá nhân buộc phải tự chịu trách nhiệm với quyết định đầu tư của mình (Freedman & Jin, 2017). Điều này khiến việc đánh giá chính xác khả năng trả nợ của người vay trở thành yếu tố then chốt trong việc giảm thiểu rủi ro và bảo vệ lợi ích của người cho vay (Ma và cộng sự, 2021). Các nền tảng trực tuyến cho vay ngang hàng đối mặt với rủi ro tín dụng, nguy cơ người vay không hoàn trả khoản vay đúng hạn hoặc mất khả năng thanh toán hoàn toàn. Khi xảy ra tình huống không thực hiện nghĩa vụ đúng hạn, người cho vay có thể mất một phần hoặc toàn bộ khoản đầu tư mà không có sự bảo lãnh từ bất kỳ tổ chức tài chính nào.

Các nghiên cứu trên P2P lending (đặc biệt với LendingClub) cho rằng nhóm biến tài chính- tín dụng (lãi suất, DTI, lịch sử chậm trả, hồ sơ bất lợi) và đặc điểm khoản vay (kỳ hạn, khoản trả góp) có vai trò quan trọng trong dự báo rủi ro (Serrano-Cinca và cộng sự, 2015; Emekter và cộng sự, 2015; Xu và cộng sự, 2021).

Gần đây, một số nghiên cứu đã mở rộng

khung phân tích truyền thống bằng cách tích hợp các nguồn dữ liệu mới, đặc biệt là dữ liệu từ mạng xã hội nhằm cải thiện độ chính xác trong dự báo không thực hiện nghĩa vụ đúng hạn. Faturohman và cộng sự (2024) kết hợp dữ liệu truyền thống với dữ liệu hành vi từ mạng xã hội; Monje và cộng sự (2025) đề xuất hệ thống cảnh báo sớm dựa trên Explainable AI (XAI) trên dữ liệu LendingClub. Trên LendingClub, các nghiên cứu áp dụng đa mô hình cho thấy random forest có thể đạt hiệu suất cao (Núñez Mora và cộng sự, 2023), trong khi Sifrain (2023) so sánh logistic, random forest và neural network và xác định các biến dự báo chủ chốt như số tiền vay, DTI, thu nhập và đặc trưng tín dụng. Ngoài dự báo vỡ nợ/không đúng hạn, một số hướng nghiên cứu khác tập trung vào hành vi nhà đầu tư và ra quyết định trên nền tảng P2P (Yi, 2023) cũng như tác động của cú sốc vĩ mô như COVID-19 đến rủi ro tín dụng. Dù đã có nhiều nghiên cứu quốc tế về khả năng trả nợ trong P2P lending, nhưng phần lớn các nghiên cứu tập trung vào các thị trường đã phát triển như Mỹ, Châu Âu hoặc Trung Quốc- nơi mà dữ liệu được tổ chức đầy đủ, các nền tảng hoạt động ổn định và hệ thống đánh giá tín dụng đã tương đối hoàn thiện. Hơn nữa, số lượng nghiên cứu khai thác trực tiếp dữ liệu từ nền tảng LendingClub và sử dụng các mô hình định lượng để đánh giá xác suất không thực hiện nghĩa vụ đúng hạn dưới góc nhìn hệ thống vẫn còn hạn chế.

Trong khi đó, cho vay ngang hàng tại Việt Nam và nhiều thị trường mới nổi vẫn đang trong giai đoạn phát triển ban đầu, hệ thống dữ liệu phân tán, thiếu đồng bộ và chưa có cơ chế pháp lý rõ ràng. Một số nghiên cứu trong nước đã bước đầu đề cập đến rủi ro tín dụng và khả năng trả nợ trong bối cảnh P2P lending (Do, 2022; Bùi Quang Trường, 2025), tuy nhiên các công trình

này chủ yếu dừng ở mức mô tả khái niệm, phân tích định tính hoặc sử dụng tập dữ liệu hạn chế. Một số nghiên cứu khai thác dữ liệu LendingClub nhưng chủ yếu như một nguồn dữ liệu mở phục vụ thử nghiệm mô hình chấm điểm tín dụng (Nguyễn Thị Thúy Quỳnh và cộng sự, 2018), trong khi số công trình đặt bộ dữ liệu này đúng trong bối cảnh P2P lending để mô hình hoá rủi ro không thực hiện nghĩa vụ đúng hạn và kiểm chứng khả năng trả nợ vẫn còn tương đối ít. Các nghiên cứu trước cho thấy các nhóm biến dự báo quan trọng thường thuộc: đặc điểm khoản vay (lãi suất, kỳ hạn, trả góp, mục đích), năng lực tài chính/đòn bẩy (DTI, thu nhập), lịch sử tín dụng (quá hạn, hồ sơ bất lợi) và ổn định việc làm, từ đó nghiên cứu này kế thừa để lựa chọn biến và xây dựng mô hình hồi quy logistic (Emekter và cộng sự, 2015; Sifrain, 2023). Việc lựa chọn LendingClub, dù là dữ liệu quốc tế vẫn phù hợp và có ý nghĩa đối với nghiên cứu tại Việt Nam vì (i) thị trường P2P trong nước còn thiếu bộ dữ liệu giao dịch toàn diện, chuẩn hoá và có thể truy vết theo vòng đời khoản vay (từ hồ sơ người vay đến trạng thái hoàn trả/vỡ nợ), khiến việc kiểm định mô hình định lượng gặp nhiều hạn chế; (ii) LendingClub cung cấp tập dữ liệu quy mô lớn, cấu trúc biến rõ ràng và nhãn kết quả tương đối chuẩn, qua đó cho phép kiểm định độ chính xác các mô hình dự báo không thực hiện nghĩa vụ đúng hạn một cách minh bạch; và (iii) trong bối cảnh mô hình quản trị rủi ro cho P2P lending ở Việt Nam còn mới, LendingClub giúp rút ra các hàm ý về nhóm biến rủi ro, đồng thời tạo nền tảng để điều chỉnh mô hình khi dữ liệu nội địa đầy đủ.

Do vậy, nghiên cứu này tập trung chuẩn hoá biến kết quả trả nợ (loan outcome) từ trường loan_status theo quy tắc mã hoá minh bạch, có thể lặp lại; đồng thời kiểm định vai trò của các yếu tố tài chính, tín

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay

ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

dụng đối với khả năng trả nợ đúng hạn bằng hồi quy logistic nhị phân trên bộ dữ liệu LendingClub giai đoạn 2007- 2023. Trên cơ sở đó, bài viết cung cấp bằng chứng thực nghiệm có tính hệ thống cho nghiên cứu rủi ro tín dụng trong P2P lending và hàm ý cho bối cảnh các thị trường còn thiếu dữ liệu chuẩn hoá như Việt Nam.

3. Phương pháp nghiên cứu

3.1. Mô hình nghiên cứu

Việc xác định các yếu tố ảnh hưởng đến khả năng trả nợ của người vay trên các nền tảng P2P là cần thiết nhằm hỗ trợ nền tảng xây dựng cơ chế đánh giá tín dụng và giảm thiểu rủi ro. Các nghiên cứu cho thấy rủi ro trả nợ trên P2P thường liên quan đến nhóm biến tài chính, tín dụng, đặc điểm khoản vay và một số đặc điểm nghề nghiệp, nhân khẩu học (Serrano-Cinca và cộng sự, 2015; Emekter và cộng sự, 2015; Weiss và cộng sự, 2010; Chong và cộng sự, 2022; Zeng và cộng sự, 2017; Ayal và cộng sự, 2019; Ho và cộng sự, 2024).

Trên cơ sở kế thừa các nghiên cứu trước, nghiên cứu này đề xuất đánh giá khả năng trả nợ của người vay trong mô hình cho vay ngang hàng P2P dựa trên các biến số như thông tin tài chính cá nhân, đặc điểm khoản vay, lịch sử vay và tín dụng. Nghiên cứu đề xuất xây dựng mô hình hồi quy logistic nhị phân, trong đó biến phụ thuộc là tình trạng trả nợ của người vay cá nhân. Mô hình hồi quy logistic được thể hiện như sau:

$$\log_e\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{11} X_{11} + \varepsilon$$

Trong đó:

$Y = \log_e\left(\frac{P(Y=1)}{1-P(Y=1)}\right)$ là xác suất người vay trả nợ đúng hạn (1), người vay không trả nợ đúng hạn (0)

$X_1 \rightarrow X_{11}$ là các biến độc lập có tác động để giải thích về khả năng trả nợ

$\beta_0 \rightarrow \beta_{11}$ là các hệ số ước lượng.

ε là sai số ngẫu nhiên.

Mô hình logistic phù hợp vì biến phụ thuộc dạng nhị phân và cho phép lượng hoá tác động của từng biến dự báo lên xác suất trả nợ, qua đó hỗ trợ ra quyết định trong thực tiễn.

Trên cơ sở tổng hợp từ các nghiên cứu trước, nghiên cứu lựa chọn 11 biến độc lập thuộc ba nhóm: (i) thông tin tài chính cá nhân, (ii) đặc điểm khoản vay, (iii) lịch sử vay và tín dụng. Chi tiết cách đo lường và mã hóa biến được trình bày tại Bảng 1 (mục 3.2).

3.2. Dữ liệu và phương pháp nghiên cứu

Nghiên cứu sử dụng mô hình hồi quy logistic nhị phân để ước lượng mối liên hệ giữa các biến giải thích và xác suất trả nợ đúng hạn. Biến phụ thuộc là khả năng trả nợ (đúng hạn/không đúng hạn). Các biến độc lập gồm cả biến định tính và định lượng, phù hợp với logistic. Đối với biến mục đích vay (PUR), nghiên cứu mã hoá theo biến giá, trong đó nhóm tham chiếu là mua xe (car) và được gán giá trị 0 xuyên suốt trong mô hình. Hiệu năng mô hình được đánh giá bằng ROC/AUC.

Nghiên cứu sử dụng dữ liệu từ LendingClub giai đoạn 2007- 2023. Dữ liệu được trích xuất từ Kaggle và LendingClub thông qua API (Kaggle, 2024). Sau khi làm sạch dữ liệu và loại bỏ quan sát thiếu biến chính, mẫu nghiên cứu còn lại gồm 1.048.575 hồ sơ vay. Để đảm bảo tính nhất quán và khả năng so sánh, nghiên cứu chuẩn hoá biến kết quả (loan outcome) từ trường loan_status theo quy tắc mã hoá minh bạch, có thể lặp lại. Cụ thể: Fully Paid được mã hoá 1; Charged Off, Default, Late (16–30 days) và Late (31–120 days) được mã hoá 0; Current và In Grace Period bị loại do chưa kết thúc vòng đời khoản vay. Ngưỡng từ 16

Bảng 1. Diễn giải các biến và cách đo lường

	Tên biến	Ký hiệu	Thang đo	Nghiên cứu có liên quan	Kỳ vọng
Biến phụ thuộc (Y)	Khả năng trả nợ	LOAN	1: trả đúng hạn/hoàn tất nghĩa vụ 0: Không thực hiện nghĩa vụ đúng hạn (bao gồm các trạng thái Charged Off, Default, quá hạn 16–30 ngày, quá hạn 31–120 ngày)	Xu và cộng sự (2021)	
Biến độc lập					
Thông tin tài chính cá nhân	X1-Thu nhập hàng năm	ANN	USD	Serrano-Cinca và cộng sự (2015)	+
	X2-Số năm làm việc	EMP	Từ < 1 năm đến +10 năm	Ayal và cộng sự (2019)	-
	X3-Số tiền vay	AMNT	USD	Zeng và cộng sự (2017)	+
	X4-Thời hạn vay	TERM	36 tháng hoặc 60 tháng	Zeng và cộng sự (2017)	-
Đặc điểm khoản vay	X5-Mục đích vay	PUR	Mua xe, trả nợ thẻ tín dụng, mua nhà, cải tạo nhà, học tập, y tế	Chong và cộng sự (2022)	-
	X6-Số tiền phải trả mỗi kỳ	INS	USD	Zeng và cộng sự (2017)	-
	X7-Lãi suất vay	INT	Phần trăm	Weiss và cộng sự (2010)	-
Lịch sử vay và tín dụng	X8- Tỷ lệ nợ / thu nhập	DTI	Phần trăm	Serrano-Cinca và cộng sự (2015)	-
	X9- Số lần chậm trả trong 2 năm gần nhất	DEL	Số lần (biến đếm: 0, 1, 2,...)	Chong và cộng sự (2022)	-
	X10- Số bản ghi nợ xấu công khai	PUB	Số bản ghi (biến đếm: 0, 1, 2,...)	Chong và cộng sự (2022)	-
	X11- Mức độ sử dụng tín dụng quay vòng	REV	Số	Chong và cộng sự (2022)	-

Chú thích: Charged Off: xóa nợ hoặc khoản vay bị xóa sổ (ngân hàng/tổ chức cho vay coi khoản vay không còn khả năng thu hồi và đưa ra khỏi sổ sách). Default: không thực hiện nghĩa vụ đúng hạn (người vay mất khả năng thanh toán đúng theo cam kết trong hợp đồng tín dụng).

Nguồn: Tác giả tổng hợp

ngày quá hạn được lựa chọn theo cách phân loại trạng thái trễ hạn của LendingClub. Thông tin về các trường dữ liệu đưa vào phân tích được trình bày tại Bảng 1.

Quy trình phân tích gồm làm sạch và mã hoá dữ liệu, thống kê mô tả, ước lượng mô hình bằng hợp lý tối đa và đánh giá hiệu năng bằng ROC/AUC. Mô hình được xử lý bằng RStudio (phiên bản 2025.05.0).

4. Kết quả và thảo luận nghiên cứu

4.1. Mô tả đặc điểm mẫu

Trước khi tiến hành phân tích hồi quy, nghiên cứu thực hiện thống kê mô tả để làm rõ đặc điểm của mẫu dữ liệu. Bảng 2 dưới đây trình bày các thống kê cơ bản (trung bình, độ lệch chuẩn, giá trị nhỏ nhất và lớn nhất) của một số biến định lượng chính trong mô hình.

Các số liệu cho thấy quy mô vay phổ biến ở mức trung bình- thấp, với mức vay trung bình 14.603 USD, thấp hơn nhiều so với ngưỡng tối đa 35.000 USD của LendingClub. Thu nhập năm trung bình của mẫu đạt 79.630 USD, cao hơn mức thu nhập hộ gia đình trung vị tại Mỹ cùng kỳ

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

Bảng 2. Đặc điểm mẫu

Biến	Ký hiệu	Trung bình	Độ lệch chuẩn	Giá trị tối thiểu	Giá trị tối đa	Loại biến
Khả năng trả nợ	LOAN	0,82	0,38	0	1	Nhị phân
Thu nhập năm (USD)	ANN	79.630	46.720	4.800	9.000.000	Định lượng
Số năm làm việc	EMP	6,4	3,2	< 1	10+	Thứ bậc
Số tiền vay (USD)	AMNT	14.603	5.385	1.000	35.000	Định lượng
Thời hạn vay (tháng)	TERM	47,8	8,9	36	60	Nhị phân
Mục đích vay	PUR	-	-	Nhiều nhóm	Nhiều nhóm	Phân loại
Khoản trả hàng tháng (USD)	INS	363	145	5,59	1,820	Định lượng
Lãi suất vay (%)	INT	13,19	4,27	5,32	26,06	Định lượng
Tỷ lệ nợ/ thu nhập (%)	DTI	18,86	9,70	0	380,53	Định lượng
Số lần chậm trả 2 năm gần nhất	DEL	0,19	0,61	0	20	Định lượng
Số bản ghi nợ xấu công khai	PUB	0,07	0,28	0	9	Định lượng
Mức độ sử dụng tín dụng quay vòng	REV	0,53	0,21	0	1	Định lượng

Nguồn: Kết quả phân tích từ RStudio

(khoảng 70.000 USD (Statista, 2023), cho thấy nền tảng phục vụ đối tượng người vay có thu nhập trung bình- khá. Đối với biến DTI, độ lệch chuẩn lớn (9,70) và khoảng giá trị rất rộng (0–380,53%) phản ánh sự khác biệt đáng kể giữa các nhóm vay. Một số trường hợp có DTI rất cao cho thấy tồn tại nhóm khoản vay rủi ro. Phần lớn khoản vay thuộc nhóm nhỏ- vừa, phù hợp với đặc trưng vay tiêu dùng trên P2P.

Xét về mục đích vay, hai nhóm chiếm tỷ trọng lớn nhất là hợp nhất nợ với 46.698 khoản vay và thanh toán thẻ tín dụng với 20.531 khoản vay, chiếm hơn 70% tổng số mẫu, tiếp theo là vay phục vụ cho hoạt động kinh doanh nhỏ, vay để mua sắm, chi phí y tế. Điều này cho thấy người vay chủ yếu tìm đến nền tảng cho vay ngang hàng P2P để giảm áp lực nợ hiện tại hoặc tái cấu trúc tài chính cá nhân. Các mục đích tiêu dùng không thiết yếu xuất hiện với tần suất thấp, phản ánh mức độ ưu tiên thấp của các khoản vay mang tính dài hạn.

Ngoài ra, lãi suất trung bình của các khoản vay là 13,19%, dao động trong khoảng từ 5,32% đến 26,06%, cho thấy sự phân hóa mạnh mẽ theo mức độ rủi ro của người vay. Khoản thanh toán trung bình hàng tháng là 363 USD. Về kỳ hạn, đa số người vay chọn 36 tháng (hơn 70%), thay vì 60 tháng, phản ánh xu hướng giảm rủi ro chi phí dài hạn. Về kinh nghiệm làm việc, nhóm thâm niên 10 năm trở lên chiếm tỷ lệ cao nhất, cho thấy mẫu nghiêng về người vay có mức độ ổn định nghề nghiệp nhất định và nhu cầu vay chủ yếu phục vụ quản lý nợ.

4.2. Kết quả hồi quy

Kết quả hồi quy cho thấy một số biến độc lập có ý nghĩa thống kê cao và đóng vai trò quan trọng trong việc dự đoán hành vi trả nợ của người vay. Cụ thể như Bảng 3.

Các yếu tố tài chính như lãi suất vay ($\beta = -0,13162$, $p < 0,001$), tỷ lệ nợ trên thu nhập (DTI) ($\beta = -0,01830$, $p < 0,001$), và

Bảng 3. Kết quả mô hình logistic

Các yếu tố tác động	Hệ số hồi quy	Sai số chuẩn	Z-value	P-value
(Intercept)	10,96137	43,95414	0,249	0,80307
ANN	0,00000	0,00000	3,792	0,00015
EMP_length1 year	-0,02403	0,04558	-0,527	0,59800
EMP_length10+ years	0,07254	0,03439	2,109	0,03494
EMP_length2 years	0,02967	0,04265	0,696	0,48661
EMP_length3 years	0,00504	0,04335	0,116	0,90751
EMP_length4 years	0,00604	0,04793	0,126	0,89968
EMP_length5 years	-0,03167	0,04718	-0,671	0,50197
EMP_length6 years	0,08237	0,05535	1,488	0,13667
EMP_length7 years	0,06457	0,05605	1,152	0,24936
EMP_length8 years	0,01034	0,05016	0,206	0,83669
EMP_length9 years	-0,03195	0,05444	-0,587	0,55720
AMNT	0,00004	0,00001	4,729	0,00000
TERM	-0,90208	0,05415	-16,658	0,00000
PUR_credit_card	-0,09259	0,10911	-0,849	0,39613
PUR_debt_consolidation	-0,12929	0,10776	-1,200	0,23022
PUR_home_improvement	-0,15080	0,11412	-1,321	0,18634
PUR_house	-0,01261	0,18091	-0,070	0,94444
PUR_major_purchase	-0,15941	0,12610	-1,264	0,20619
PUR_medical	-0,14253	0,13622	-1,046	0,29542
PUR_moving	0,03791	0,15617	0,243	0,80822
PUR_renewable_energy	0,14385	0,37436	0,384	0,70079
PUR_small_business	-0,29637	0,13721	-2,160	0,03077
PUR_vacation	-0,05194	0,16779	-0,310	0,75692
PUR_other	0,01179	0,11418	0,103	0,91775
INS	-0,00137	0,00026	-5,321	0,00000
INT	-0,13162	0,00331	-39,743	0,00000
DTI	-0,01830	0,00112	-16,299	0,00000
DEL	-0,05765	0,00943	-6,114	0,00000
PUB	-0,06700	0,01343	-4,988	0,00000
REV	-0,00028	0,00041	-0,671	0,50195

Nguồn: Kết quả phân tích từ RStudio

khoản thanh toán hàng kỳ (installment) ($\beta = -0,00137$, $p < 0,001$) đều có tác động tiêu cực đáng kể đến xác suất trả nợ đúng hạn.

Bên cạnh đó, số tiền vay (loan_amnt) có ảnh hưởng dương nhỏ nhưng có ý nghĩa thống kê cao ($\beta = 0,00004$, $p < 0,001$), cho

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay

ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

thấy những người vay số tiền lớn hơn- có thể do có hồ sơ tín dụng tốt hơn có xác suất trả nợ cao hơn.

Một số yếu tố hồ sơ tín dụng cũng có ảnh hưởng rõ rệt: số lần chậm trả trong 2 năm gần nhất (*delinq_2yrs*) và số bản ghi nợ xấu công khai (*pub_rec*) có hệ số âm ($p < 0,001$), phản ánh lịch sử tín dụng kém làm giảm khả năng hoàn trả. Thời hạn vay (*term*) có tác động tiêu cực rất mạnh ($\beta = -0,90208$, $p < 0,001$): nhóm vay 60 tháng có rủi ro không trả đúng hạn cao hơn đáng kể so với nhóm 36 tháng.

Về đặc điểm cá nhân, thu nhập hàng năm (*annual_inc*) có ảnh hưởng tích cực đến khả năng trả nợ ($p = 0,00015$), dù hệ số gần bằng 0 do biến được chuẩn hóa theo đơn vị rất nhỏ. Ngoài ra, thâm niên làm việc trên 10 năm là nhóm duy nhất trong biến số năm làm việc (*emp_length*) cho thấy ảnh hưởng dương có ý nghĩa thống kê ($\beta = 0,07254$, $p = 0,03494$), cho thấy sự ổn định nghề nghiệp góp phần làm tăng xác suất hoàn trả đúng hạn. Phần lớn các biến liên quan đến mục đích vay (*purpose*) không có ý nghĩa thống kê đáng kể, ngoại trừ nhóm vay cho kinh doanh nhỏ (*small_business*) với ảnh hưởng âm rõ rệt ($\beta = -0,29637$, $p = 0,03077$). Điều này gợi ý rằng các khoản vay mang tính rủi ro kinh doanh có nhiều khả năng dẫn đến tình trạng không hoàn trả đúng hạn.

Để làm rõ ý nghĩa kinh tế, các hệ số hồi quy logit có thể diễn giải thông qua tỷ số odds (odds ratio – OR), với $OR = \exp(\beta)$. Theo đó, các biến phản ánh “gánh nặng trả nợ”

như INT, DTI và INS đều có β âm, hàm ý khi các chỉ báo gánh nặng tăng lên thì odds trả đúng hạn giảm xuống, qua đó rủi ro không thực hiện nghĩa vụ đúng hạn tăng lên. Đặc biệt, kỳ hạn vay (TERM) có $\beta = -0,90208$, tương ứng $OR \approx \exp(-0,902) \approx 0,406$, cho thấy việc lựa chọn kỳ hạn 60 tháng làm odds trả đúng hạn chỉ còn khoảng 40,6% so với kỳ hạn 36 tháng (giảm khoảng 59,4%), phản ánh rủi ro tích lũy theo thời gian và áp lực trả nợ dài hạn. Ở nhóm “lịch sử tín dụng”, các biến DEL và PUB có β âm, cho thấy quá khứ chậm trả và hồ sơ bất lợi làm suy giảm đáng kể odds trả đúng hạn, phù hợp với lập luận rằng lịch sử tín dụng là tín hiệu rủi ro quan trọng. Ngược lại, thâm niên làm việc trên 10 năm có $\beta = 0,07254$ ($OR \approx 1,075$), hàm ý ổn định nghề nghiệp giúp odds trả đúng hạn tăng nhẹ khoảng 7,5%. Đối với số tiền vay (AMNT), hệ số dương rất nhỏ ($\beta = 0,00004$) cho thấy tác động cận biên theo đơn vị USD là nhỏ. Tuy nhiên, dấu dương gợi ý các khoản vay lớn có thể đi kèm sàng lọc tín dụng chặt hơn và hồ sơ vay tốt hơn khi đã kiểm soát các biến tín dụng khác. Nhìn chung, kết quả nhấn mạnh vai trò nổi trội của ba nhóm chỉ báo: (i) gánh nặng trả nợ (INT, DTI, INS, TERM), (ii) lịch sử tín dụng (DEL, PUB) và (iii) ổn định việc làm (EMP_length 10+ years) trong việc giải thích xác suất trả nợ đúng hạn trên nền tảng P2P.

4.3. Kết quả dự báo của mô hình

Sau khi xây dựng mô hình hồi quy logistic

Bảng 4. Kết quả dự báo của mô hình

Khả năng dự báo	Thực tế = 0 (Không trả được nợ)	Thực tế = 1 (Trả được nợ)
Không trả được nợ (Dự báo = 0)	2.269 (TN)	1.888 (FN)
Trả được nợ (Dự báo = 1)	13.901 (FP)	64.169 (TP)
Khả năng dự báo chính xác	80,8%	

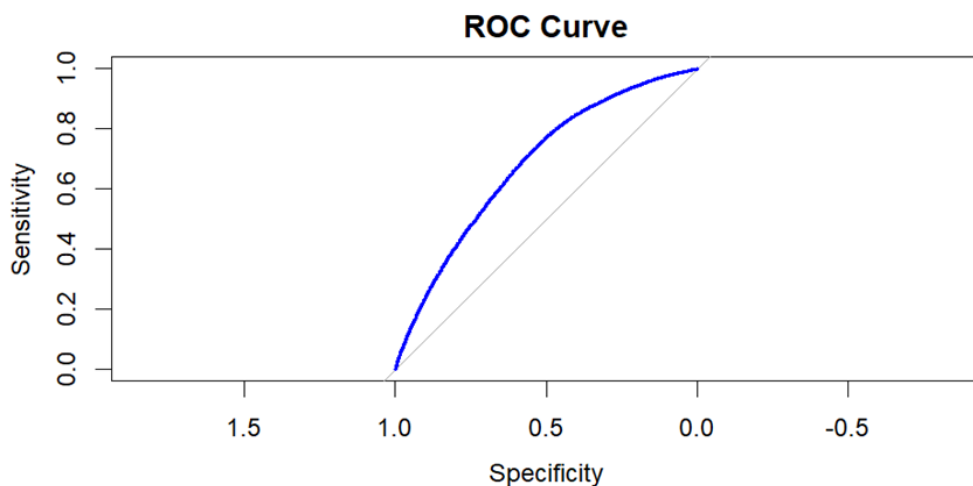
Nguồn: Kết quả phân tích từ RStudio

đề dự báo khả năng hoàn trả nợ đúng hạn của người vay, tác giả tiến hành kiểm định độ chính xác và hiệu suất phân loại của mô hình thông qua ma trận nhầm lẫn và các chỉ số thống kê liên quan. Kết quả ma trận dự báo được trình bày ở Bảng 4.

Từ kết quả trên, có thể thấy rằng mô hình dự đoán đúng 2.269 trường hợp thực sự không thực hiện nghĩa vụ đúng hạn và 64.169 trường hợp không không thực hiện nghĩa vụ đúng hạn. Tỷ lệ dự báo chính xác tổng thể đạt 80,8%, với khoảng tin cậy 95% là (80,53%; 81,07%). So với các nghiên cứu gần đây, ví dụ Sifrain (2023) sử dụng dữ liệu LendingClub với các mô hình logistic regression, random forest và neural network ghi nhận AUC khoảng 0,90 và độ chính xác (accuracy) khoảng 90%. Nghiên cứu của Sam'an và cộng sự (2024) với dữ liệu năm 2023 cũng chỉ ra rằng dù nhiều biến được thêm, mô hình nâng cao thường đạt accuracy từ khoảng 80- 90%. Do đó, tỷ lệ dự báo chính xác tổng thể của mô hình (80,8%) tuy chưa đạt mức cao nhất nhưng nằm trong khoảng chấp nhận được.

4.4. Kiểm định độ phù hợp của mô hình

Kết quả mô hình cho thấy các yếu tố tài chính và tín dụng là những chỉ báo mạnh nhất cho hành vi trả nợ của người vay. Việc sử dụng các biến như lãi suất, DTI, lịch sử tín dụng và kỳ hạn vay nên được ưu tiên trong các mô hình chấm điểm tín dụng, nhằm nâng cao khả năng phát hiện rủi ro và đưa ra quyết định cho vay hiệu quả hơn. Để đánh giá hiệu suất phân loại của mô hình hồi quy logistic, đường cong ROC đã được xây dựng (Hình 1). Kết quả cho thấy đường cong ROC nằm hoàn toàn phía trên đường chéo ngẫu nhiên, phản ánh năng lực phân biệt rõ ràng giữa hai nhóm người vay: hoàn trả đúng hạn và không hoàn trả đúng hạn. Diện tích dưới đường cong ROC (AUC) đạt giá trị 0,7814 > 0,7, cho thấy mô hình có khả năng phân loại đúng trong khoảng 78,14% các cặp quan sát ngẫu nhiên được chọn. Giá trị AUC này được ghi nhận trên bộ dữ liệu quy mô lớn (1.048.575 khoản vay) và giai đoạn dài 2007- 2023. Theo thang đo phổ biến, giá trị AUC này phản ánh hiệu suất phân loại ở được chấp nhận trong các mô hình phân loại ứng dụng thực tiễn. Do đó, mô hình có thể được xem là phù hợp để hỗ trợ đánh giá rủi ro tín dụng



Nguồn: Kết quả phân tích từ RStudio

Hình 1. Đường cong ROC

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay

ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

trong bối cảnh cho vay ngang hàng và có thể sử dụng để đối chiếu về hiệu năng của mô hình logistic khi áp dụng cho dữ liệu LendingClub trong phạm vi nghiên cứu.

4.5. Thảo luận kết quả nghiên cứu

Nghiên cứu này đã sử dụng mô hình hồi quy logistic để đánh giá các yếu tố ảnh hưởng đến khả năng trả nợ đúng hạn của người vay trên nền tảng P2P, dựa trên tập dữ liệu LendingClub hơn 1 triệu khoản vay giai đoạn 2007- 2023.

Về yếu tố tài chính, lãi suất vay, tỷ lệ nợ/ thu nhập (DTI) và khoản thanh toán hàng kỳ đều tác động tiêu cực đến khả năng trả nợ đúng hạn, phù hợp với lý thuyết tài chính tiêu dùng và kết quả của Serrano-Cinca và cộng sự (2015); Emekter và cộng sự (2015). Các biến này phản ánh gánh nặng trả nợ, chi phí vay và áp lực nghĩa vụ tài chính càng cao thì xác suất trả đúng hạn càng giảm. Kết quả góp phần củng cố tính ổn định của các quan hệ này trên mẫu lớn và giai đoạn dài.

Các yếu tố tín dụng, số lần chậm trả trong 2 năm gần nhất (DEL) và số bản ghi nợ công khai (PUB) có tác động bất lợi đến khả năng trả nợ đúng hạn, cho thấy lịch sử tín dụng kém làm gia tăng rủi ro không thực hiện nghĩa vụ đúng hạn.

Yếu tố kỳ hạn vay (TERM) có ảnh hưởng tiêu cực rất mạnh, nhóm vay 60 tháng có rủi ro không thực hiện nghĩa vụ đúng hạn cao hơn nhóm 36 tháng, phù hợp với Xu và cộng sự (2021) và gợi ý TERM là biến cần ưu tiên trong chấm điểm rủi ro P2P.

Đáng chú ý là số tiền vay (AMNT) có tác động dương rất nhỏ nhưng ý nghĩa thống kê cao. Kết quả này có thể phản ánh cơ chế thẩm định/sàng lọc chặt hơn đối với khoản vay lớn và chất lượng hồ sơ người vay tốt hơn, nên xác suất trả nợ cao hơn dù mức ảnh hưởng không lớn. Phát hiện

này khác với Serrano-Cinca và cộng sự (2015) nhưng một số nghiên cứu gần đây trên LendingClub cũng ghi nhận AMNT không làm xấu chất lượng tín dụng khi đã kiểm soát các đặc điểm rủi ro khác (Sifrain, 2023; Nguyen và cộng sự, 2021). Do đó, hàm ý ở đây là nhấn mạnh vai trò của cơ chế thẩm định và phân tầng khách hàng theo mức vay, hơn là khuyến khích mở rộng các khoản vay giá trị lớn.

Thâm niên làm việc trên 10 năm (EMP_length 10+ years) có ảnh hưởng tích cực, hàm ý ổn định nghề nghiệp gắn với khả năng trả nợ tốt hơn. Hầu hết các mục đích vay không có ý nghĩa thống kê, ngoại trừ nhóm vay cho kinh doanh nhỏ (small_business) có rủi ro cao hơn, phù hợp với bản chất dòng tiền biến động và nhạy cảm chu kỳ, kết quả này tương đồng với các nghiên cứu trên Lending Club như Sifrain (2023), Yi (2023), Nguyen và cộng sự (2021). Nhìn chung, kết quả nhấn mạnh vai trò của nhóm chỉ báo tài chính tín dụng và đặc điểm khoản vay trong đánh giá rủi ro trả nợ trên P2P.

Để kiểm tra hiệu quả phân loại của mô hình, đường cong ROC đã được sử dụng. Kết quả cho thấy mô hình đạt $AUC = 0,7814$, cho thấy mô hình phân loại khá tốt giữa nhóm trả đúng hạn và nhóm không thực hiện nghĩa vụ đúng hạn. Kết quả này phù hợp với Serrano-Cinca và cộng sự (2015), Emekter và cộng sự (2015). So với một số nghiên cứu gần đây dùng random forest hay neural network như Faturhman và cộng sự (2024), Núñez Mora và cộng sự (2023) và Sifrain (2023), kết quả cho thấy logistic vẫn có thể đạt hiệu suất đáng tin cậy khi dữ liệu và có lợi thế khi triển khai. Vì vậy, các biến như lãi suất, tỷ lệ nợ trên thu nhập, lịch sử tín dụng và kỳ hạn vay nên được ưu tiên trong các mô hình chấm điểm tín dụng, nhằm nâng cao khả năng phát hiện rủi ro và đưa ra quyết định cho vay hiệu quả hơn.

5. Kết luận

Đóng góp chính của nghiên cứu là kiểm định thực nghiệm các yếu tố tài chính- tín dụng trên bộ dữ liệu lớn (hơn 1 triệu khoản vay LendingClub), qua đó củng cố bằng chứng về vai trò của các chỉ báo rủi ro cốt lõi trong chấm điểm tín dụng P2P. Bài viết đồng thời chuẩn hoá biến kết quả (loan outcome) từ trường loan_status theo quy tắc minh bạch, có thể lặp lại, giúp tăng tính nhất quán khi so sánh giữa các nghiên cứu sử dụng dữ liệu P2P. Mô hình logistic cho thấy hiệu suất phân loại khá tốt, có thể là cơ sở cho các hệ thống đánh giá tín dụng bán tự động. Việc khai thác bộ dữ liệu chuẩn hoá quốc tế cũng giúp làm rõ một số quan hệ còn chưa thống nhất trong các nghiên cứu trước (ví dụ số tiền vay, lịch sử tín dụng, thâm niên nghề nghiệp).

Về hàm ý thực tiễn, kết quả cho thấy các nền tảng P2P nên ưu tiên tích hợp các biến lãi suất, DTI, lịch sử tín dụng và kỳ hạn vay vào cơ chế phê duyệt và quản trị rủi ro, thay vì dựa chủ yếu vào thông tin nhân khẩu học. Nhóm khách hàng có DTI cao, kỳ hạn dài và lịch sử chậm trả cần được kiểm soát chặt hơn thông qua cảnh báo sớm và hạn mức tín dụng thích ứng. Trong bối cảnh Việt Nam thiếu dữ liệu chuẩn hoá, kết quả có thể được sử dụng như khung tham chiếu về cách chuẩn hoá biến kết quả và lựa chọn nhóm chỉ báo rủi ro khi thiết kế mô hình chấm điểm. Đối với nhà hoạch định chính sách, nghiên cứu nhấn mạnh nhu cầu minh bạch hoá và chuẩn hoá dữ liệu tín dụng, phát triển kho dữ liệu tập trung và cơ chế

chia sẻ dữ liệu (kể cả dữ liệu thay thế), qua đó hỗ trợ xây dựng chuẩn đánh giá rủi ro và hoàn thiện khung pháp lý cho vay P2P tại Việt Nam.

Mặc dù mô hình hồi quy logistic đã cho thấy hiệu quả phân loại khá tốt, nghiên cứu vẫn tồn tại một số hạn chế nhất định. Thứ nhất, dữ liệu sử dụng chủ yếu dựa trên thông tin có sẵn từ nền tảng cho vay ngang hàng, chưa tích hợp dữ liệu thay thế (hành vi giao dịch điện tử, mạng xã hội, dữ liệu thanh toán số), nên có thể bỏ sót một số tín hiệu rủi ro. Thứ hai, hồi quy logistic mang tính tuyến tính trên thang logit, có thể hạn chế trong việc phản ánh quan hệ phi tuyến hoặc tương tác phức tạp giữa các biến.

Ngoài ra, nghiên cứu hiện tại chỉ tập trung vào khả năng phân loại nhị phân (đúng hạn-không đúng hạn), chưa đi sâu vào mức độ trễ hạn, khả năng hồi phục khoản vay hoặc thời gian xảy ra sự kiện, vốn quan trọng cho phân tích rủi ro toàn diện. Do thiết kế nhị phân của biến phụ thuộc, nghiên cứu gộp các trạng thái trễ hạn (Late 16-30, Late 31-120) cùng với Charged Off/Default vào nhóm không đúng hạn. Cách mã hoá này hỗ trợ nhận diện sớm rủi ro, nhưng có thể phát sinh sai lệch phân loại do một số khoản vay trễ hạn có thể hồi phục. Các nghiên cứu tiếp theo có thể tách riêng default/charged off và delinquency và mở rộng dữ liệu để nâng cao tính khái quát hóa và cân nhắc bổ sung phân tích chi phí- lợi ích để đánh giá tính khả thi kinh tế khi triển khai mô hình chấm điểm tín dụng tự động trong thực tiễn. ■

Tài liệu tham khảo

- Alpha AM. (2025). *Thị trường P2P Lending Việt Nam 2025: Quy mô, Pháp Lý & Chiến Lược Đầu Tư*. <https://alphaam.vn/quy-mo-thi-truong-p2p-lending-viet-nam-2025/>
- Ayal, S., Bar-Haim, D., & Ofir, M. (2019). Behavioral Biases in Peer-to-Peer (P2P) Lending. In *Behavioral Finance* (pp. 367–400). World Scientific. https://doi.org/10.1142/9789813279469_0011
- Balogun, E.D., & Alimi, A. (1988). Loan delinquency among small farmers in developing countries: a case study of the small-farmer credit programme in Lagos State of Nigeria. *CBN Economic and Financial Review*, 26(3), 36-44.

Các yếu tố ảnh hưởng khả năng trả nợ của người vay trên các nền tảng cho vay
ngang hàng (P2P lending): Bằng chứng từ dữ liệu của LendingClub

- <https://dc.cbn.gov.ng/efr/vol26/iss3/5/>
- Basha, S., Elgammal, M. M., & Abuzayed, B. M. (2021). Online peer-to-peer lending: A review of the literature. *Electronic Commerce Research and Applications*, 48, 101069. <https://doi.org/10.1016/j.elerap.2021.101069>
- Bùi Quang Trường. (2025). Tối ưu hóa quy trình cho vay trực tuyến trong ngân hàng thông qua tích hợp mô hình định hướng mục tiêu i* và BPMN. *Tạp chí Kinh Tế - Luật và Ngân hàng*, 27(10), 90–104. <https://doi.org/10.59276/JELB.2025.09.2839>
- CGAP. (1999). Measuring Microcredit Delinquency: Ratios Can be harmful to your health. *Occasional Paper*, 3, 1–20. <http://www.cgap.org>
- Chong, P. S., Labadin, J., & Meziane, F. (2022). Credit Risk Prediction for Peer-To-Peer Lending Platforms: An Explainable Machine Learning Approach. *Journal of Computing and Social Informatics*, 1(2), 1–16. <https://doi.org/10.33736/jcsi.4761.2022>
- Do, H. L. (2022). P2P Lending's Credit Risk in Vietnam: Determinants and Discussions. *Journal of Economics, Finance And Management Studies*, 05(04). <https://doi.org/10.47191/jefms/v5-i4-18>
- Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, 47(1), 54–70. <https://doi.org/10.1080/00036846.2014.962222>
- Faturohman, T., Wiryono, S. K., Khilfah, H. L. N., Andri, A., Hamzah, M. A., Saputra, O., & Indrayana, G. G. (2024). Peer-to-peer lending default prediction model: a credit scoring application with social media data. *International Journal of Monetary Economics and Finance*, 17(2/3), 189–200. <https://doi.org/10.1504/IJMEF.2024.139027>
- Freedman, S., & Jin, G. Z. (2017). The information value of online social networks: Lessons from peer-to-peer lending. *International Journal of Industrial Organization*, 51, 185–222. <https://doi.org/10.1016/j.ijindorg.2016.09.002>
- Ho, K.-C., Gu, Y., Yan, C., & Gozgor, G. (2024). Peer effects in the online peer-to-peer lending market: Ex-ante selection and ex-post learning. *International Review of Financial Analysis*, 92, 103056. <https://doi.org/10.1016/j.irfa.2023.103056>
- Kaggle. (2024). LendingClub_data_Defaults. <https://www.kaggle.com/code/danofner/lendingclub-data-defaults>
- Ko, P.-C., Lin, P.-C., Do, H.-T., & Huang, Y.-F. (2022). P2P Lending Default Prediction Based on AI and Statistical Models. *Entropy*, 24(6), 801. <https://doi.org/10.3390/e24060801>
- Ma, Z., Hou, W., & Zhang, D. (2021). A credit risk assessment model of borrowers in P2P lending based on BP neural network. *PLOS one*, 16(8), e0255216. <https://doi.org/10.1371/journal.pone.0255216>
- Monje, L., Carrasco, R. A., & Sánchez-Montañés, M. (2025). Machine Learning XAI for Early Loan Default Prediction. *Computational Economics*. <https://doi.org/10.1007/s10614-025-10962-9>
- Nguyen, N., & Ngo, D. (2025). Comparative analysis of boosting algorithms for predicting personal default. *Cogent Economics & Finance*, 13(1). <https://doi.org/10.1080/23322039.2025.2465971>
- Nguyen, T. T. A., Pham, T. M. H., & Vu, T. L. T. (2021). Default in the us peer-to-peer market with covid-19 pandemic update: An empirical analysis from lending club platform. *International Journal of Entrepreneurship*, 25(7), 1–19. <https://www.abacademies.org/abstract/default-in-the-us-peertopeer-market-with-covid19-pandemic-update-an-empirical-analysis-from-lending-club-platform-12003.html>
- Nigmonov, A., Shams, S., & Urbonas, P. (2024). Estimating probability of default via delinquencies? Evidence from European P2P lending market. *Global Finance Journal*, 63, 101050. <https://doi.org/10.1016/j.gfj.2024.101050>
- Núñez Mora, J. A., Moncayo, P., Franco, C., Madrazo-Lemarro, P., & Beltrán, J. (2023). Loan Default Prediction: A Complete Revision of LendingClub. *Revista Mexicana de Economía y Finanzas*, 18(3), 1–13. <https://doi.org/10.21919/remef.v18i3.886>
- Precedence Research. (2025). *Peer to Peer (P2P) Lending Market Size, Share, and Trends 2025 to 2034*. <https://www.precedenceresearch.com/peer-to-peer-lending-market>
- Nguyễn Thị Thúy Quỳnh, Trần Thị Xuân Anh, & Bùi Lê Trà Linh. (2018). Ứng dụng mô hình logistic chấm điểm khách hàng cá nhân nộp hồ sơ vay trên LendingClub. *Tạp Chí Kinh Tế Đối Ngoại*, 102. <https://tapchi.ftu.edu.vn/cac-so-tap-chi-ktđn/tap-chi-ktđn-so-101-110/tap-chi-ktđn-so-102/1550-ung-dung-mo-hinh-logistic-cham-diem-khach-hang-ca-nhan-nop-ho-so-vay-trên-lendingclub.html>
- Sam'an, M., Safuan, & Munsarif, M. (2024). Feature selection in P2P lending for default prediction using grey wolf optimization and machine learning. *Bulletin of Electrical Engineering and Informatics*, 13(5), 3609–3615. <https://doi.org/10.11591/eei.v13i5.7651>
- Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of Default in P2P Lending. *PLOS one*, 10(10), e0139427. <https://doi.org/10.1371/journal.pone.0139427>
- Sifrain, R. (2023). Predictive Analysis of Default Risk in Peer-to-Peer Lending Platforms: Empirical Evidence from LendingClub. *Journal of Financial Risk Management*, 12(01), 28–49. <https://doi.org/10.4236/jfrm.2023.121003>
- Statista. (2023). Percentage distribution of household income in the United States in 2023. <https://www.statista.com/statistics/203183/percentage-distribution-of-household-income-in-the-us/>
- Vu, T. T. H. (2025). Bibliometric insights into mobile money and data trends. *Data in Brief*, 60, 111633. <https://doi.org/10.1016/j.dib.2025.111633>

- Weiss, G. N. F., Pelger, K., & Horsch, A. (2010). Mitigating Adverse Selection in P2P Lending – Empirical Evidence from Prosper.com. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1650774>
- Xu, J., Lu, Z., & Xie, Y. (2021). Loan default prediction of Chinese P2P market: a machine learning methodology. *Scientific Reports*, *11*(1), 18759. <https://doi.org/10.1038/s41598-021-98361-6>
- Yi, B. (2023). P2P Investment Data Analytics: A Case Study of Lending Club. *BCP Business & Management*, *38*, 1844–1853. <https://doi.org/10.54691/bcpbm.v38i.3979>
- Zeng, X., Liu, L., Leung, S., Du, J., Wang, X., & Li, T. (2017). A decision support model for investment on P2P lending platform. *PLOS one*, *12*(9), e0184242. <https://doi.org/10.1371/journal.pone.0184242>