

MÔ HÌNH CHẤM ĐIỂM TÍN DỤNG TRONG HOẠT ĐỘNG QUẢN TRỊ RỦI RO CỦA NGÂN HÀNG THƯƠNG MẠI TẠI VIỆT NAM: ỨNG DỤNG MÔ HÌNH LOGISTIC

Quách Hải Yến

Ngân hàng TMCP Quân đội, Hà Nội, Việt Nam

Vũ Minh Tuấn¹

Tập đoàn Công nghiệp - Viễn thông Quân đội, Hà Nội, Việt Nam

Lê Phương Hà

Viện Chiến lược Chuyển đổi số, Hà Nội, Việt Nam

Ngày nhận: 07/10/2022; Ngày hoàn thành biên tập: 25/10/2023; Ngày duyệt đăng: 30/10/2023

DOI: <https://doi.org/10.38203/jiem.vi.102023.1089>

Tóm tắt: Bài viết nghiên cứu ứng dụng mô hình logistic trong việc xây dựng mô hình chấm điểm tín dụng trong hoạt động quản trị rủi ro của ngân hàng thương mại tại Việt Nam. Nghiên cứu sử dụng dữ liệu từ một ngân hàng thương mại của Việt Nam cho 19751 người vay là khách hàng cá nhân năm 2021. Bằng phương pháp nghiên cứu định lượng, kết quả nghiên cứu chỉ ra các biến như: kỳ hạn của khoản vay, lãi suất thỏa thuận của khoản vay, tỷ lệ giữa tổng số tiền thanh toán nợ hàng tháng của bên vay và thu nhập hàng tháng của bên vay, trình độ học vấn của người vay, tình trạng sở hữu nhà của người vay, thời gian làm việc của người vay, tình trạng hôn nhân của người vay, số tài khoản các loại mà người vay mở trong 6 tháng, số ngày quá hạn trả nợ lớn nhất trong 6 tháng, tỷ lệ giữa tổng số dư hiện tại trên hạn mức tín dụng cao nhất cho tất cả các tài khoản thẻ ngân hàng của người vay có quan hệ cùng chiều với xác suất vỡ nợ. Từ đó, bài viết đề xuất xây dựng mô hình chấm điểm tín dụng cho tập khách hàng và cung cấp bằng chứng khoa học quan trọng cho nhà quản trị rủi ro, góp phần hỗ trợ nhà quản trị ra quyết định trong hoạt động kinh doanh.

Từ khóa: Mô hình Logistic, Chấm điểm tín dụng, Quản trị rủi ro, Ngân hàng thương mại

¹ Tác giả liên hệ, Email: tuam.vm88@gmail.com

A CREDIT SCORING MODEL IN RISK MANAGEMENT OF COMMERCIAL BANKS IN VIETNAM: AN APPLICATION OF LOGISTIC MODEL

Abstract: This article studies the application of the logistic model in constructing a credit scoring model for risk management in commercial banks in Vietnam. The research uses data from a Vietnamese commercial bank for 19,751 individual borrowers in 2021. Using quantitative research methods, the research examines variables such as loan term, agreed interest rate of the loan, the ratio between the total debt payments and the monthly income, the education level, home ownership status, working time, marital status, number of accounts of all types opened by the Borrower in 6 months, maximum number of days overdue for payment in 6 months, the ratio between the total current balance and the highest credit rating for all the Borrower's bank card accounts. These variables are found to be positively related to the probability of default. As a results, the article proposes the development of a scoring model for customer, provides important scientific evidence for risk managers and supports managers in making decisions in business activities.

Keywords: Logistic Model, Credit Scoring, Risk Management, Commercial Banking

1. Giới thiệu

Hoạt động tín dụng của Ngân hàng thương mại (NHTM) là phương thức cung cấp vốn cho các hoạt động của nền kinh tế. Cuối năm 2021, tỷ lệ tín dụng trên GDP của Việt Nam đã đạt ngưỡng 124% (Ngân hàng Nhà nước, 2022). Đây là nghiệp vụ trọng tâm và là hoạt động mang lại thu nhập chính cho ngân hàng, đồng thời thúc đẩy sự tăng trưởng kinh tế của một đất nước. Tuy nhiên, tín dụng cũng là hoạt động tiềm ẩn nhiều rủi ro nhất đối với hoạt động kinh doanh của các NHTM. Vì vậy, quản trị rủi ro là công tác cần thiết để các ngân hàng có thể đảm bảo hoạt động kinh doanh của mình diễn ra một cách hiệu quả và an toàn. Đặc biệt, trong công tác quản trị rủi ro, các ngân hàng cần đo lường và lượng hóa được rủi ro để đánh giá chính xác được tình hình, từ đó đưa ra được những cơ chế giám sát phù hợp.

Các NHTM tại Việt Nam ngày càng dành nhiều sự quan tâm tới hoạt động nghiên cứu và xây dựng các mô hình chấm điểm tín dụng để phục vụ cho quá trình quản trị rủi ro. Một số nghiên cứu như Phan (2018), Lê (2017), Trần (2020) đã chỉ ra vai trò quan trọng của mô hình chấm điểm tín dụng ở các NHTM Việt Nam trong hoạt động đo lường rủi ro và đáp ứng các thông lệ quốc tế. Trước đó, các nghiên cứu của Bùi & cộng sự (2016), Phan (2018) cũng đề xuất một số phương pháp, kỹ thuật để xây dựng mô hình chấm điểm tín dụng cho hệ thống ngân hàng ở Việt Nam. Trên thế giới, các nghiên cứu liên quan

đến chủ đề chấm điểm tín dụng chủ yếu tập trung vào phương pháp khác nhau để xây dựng mô hình chấm điểm tín dụng. Trong đó, hai phương pháp được nghiên cứu nhiều nhất là phương pháp thống kê và phương pháp học máy. Nếu như hồi quy Logistic là mô hình được sử dụng phổ biến nhất trong lĩnh vực chấm điểm tín dụng (Crook & cộng sự, 2007) thì các kỹ thuật khác phức tạp hơn về học máy và học sâu vẫn đang tiếp tục được nghiên cứu và chứng minh tính hiệu quả trong những năm trở lại đây. Finlay (2011) thực hiện xây dựng mô hình chấm điểm tín dụng dựa trên hai tập dữ liệu thực tế với năm phương pháp được sử dụng bao gồm: hồi quy Logistic, phân tích phân biệt tuyến tính (LDA), Decision Tree, Neural Network và KNN. Kết quả chỉ ra rằng, mô hình hồi quy Logistic cho kết quả tốt nhất.

Kruppa & cộng sự (2013) thực hiện xây dựng dự đoán xác suất vỡ nợ của các khoản cho vay tiêu dùng bằng phương pháp Random Forest, KNN và hồi quy Logistic. Kết quả ước lượng các mô hình cho thấy thuật toán Random Forest có hiệu quả vượt trội so với phương pháp KNN và hồi quy Logistic truyền thống. Gần đây, Hovdenakk (2021) sử dụng dữ liệu từ một ngân hàng ẩn danh của Na Uy gồm dữ liệu của các khách hàng đi vay với mục tiêu so sánh tính hiệu quả của các thuật toán khác nhau. Bộ dữ liệu gồm 12.240 quan sát, hơn 50 trường thông tin chứa các thông tin về hồ sơ khách hàng, thông tin khoản vay hiện tại và các khoản vay khác, hành vi thanh toán, sử dụng hạn mức, cũng như hành vi trả nợ. Bài viết sử dụng các phương pháp để xây dựng và kiểm tra tính hiệu quả mô hình bao gồm: hồi quy Logistic, KNN, Decision Tree, Random Forest và Gradient Boosting. Kết quả chỉ ra rằng, hồi quy Logistic và Gradient Boosting là hai phương pháp dự đoán chính xác nhất cho tập dữ liệu này dựa theo chỉ số AUROC. Tác giả cũng chỉ ra hồi quy Logistic là phương pháp mang lại lợi ích tốt nhất vì độ chính xác và tính dễ giải thích của nó.

Đối với việc chấm điểm tín dụng cho khách hàng doanh nghiệp, Muriuki (2021) đã phát triển một mô hình dự đoán để đánh giá xác suất vỡ nợ của các khách hàng doanh nghiệp vừa và nhỏ từ các NHTM ở Kenya. Mẫu xây dựng mô hình gồm 210 doanh nghiệp vừa và nhỏ được chọn từ 42 NHTM ở Kenya. Bằng cách xây dựng mô hình thông qua mô hình hồi quy Logistic, bài viết đã chỉ ra được các biến lựa chọn theo phương pháp 5C có ý nghĩa trong việc phân loại các khoản vay có xác suất vỡ nợ cao với các khoản vay an toàn của doanh nghiệp vừa và nhỏ.

Nhìn chung, các nghiên cứu đi trước đã đưa ra rất nhiều phương pháp để xây dựng mô hình chấm điểm tín dụng, trong đó hồi quy Logistic là một phương pháp phổ biến, có tính giải thích cao và mang lại kết quả khá tốt. Song các công trình nghiên cứu này chủ yếu tập trung vào quá trình tạo ra một thể điểm để ứng dụng vào hoạt động kinh doanh nên vẫn tồn tại những hạn chế. Các công trình đi trước phần lớn tập trung chủ yếu vào cấu phần mô hình dự đoán xác suất vỡ nợ và xác định

phương pháp tối ưu tính chính xác của mô hình. Do đó, mục tiêu của bài viết này là sử dụng phương pháp Logistic cho việc xây dựng mô hình chấm điểm tín dụng cho khách hàng cá nhân và từ đó tạo ra thể điểm phù hợp để phục vụ cho hoạt động kinh doanh đối với trường hợp của NHTM Việt Nam.

Bài viết có cấu trúc gồm 5 phần, sau phần giới thiệu là cơ sở lý luận về thể điểm cân bằng. Phương pháp nghiên cứu được thực hiện ở phần 3. Phần 4 là kết quả nghiên cứu và thảo luận. Cuối cùng, kết luận và hàm ý rút ra được thể hiện trong phần 5.

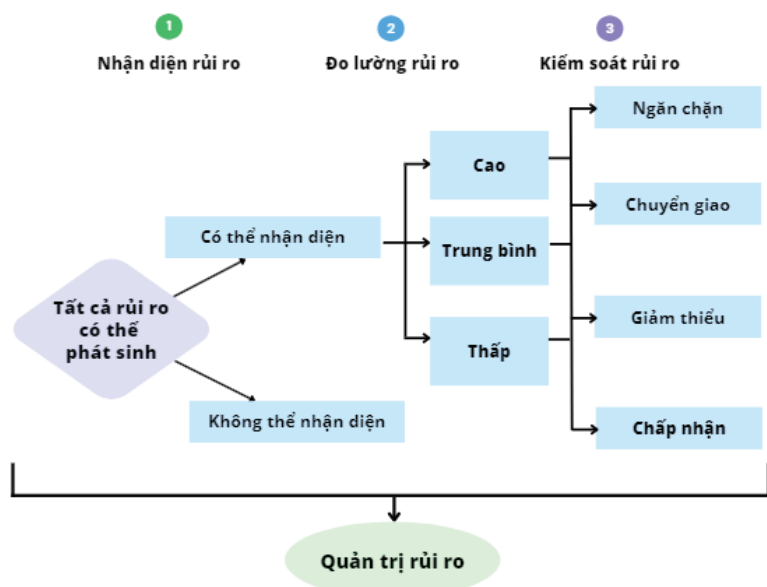
2. Cơ sở lý thuyết về quản trị rủi ro trong ngân hàng thương mại và mô hình chấm điểm tín dụng

2.1 Quản trị rủi ro trong ngân hàng thương mại

Quản trị rủi ro là một chuỗi hành động bao gồm việc xác định rủi ro và kiểm soát rủi ro ở một mức có thể chấp nhận được. Mức độ chấp nhận này khác nhau giữa các tổ chức. Vì vậy, hoạt động quản trị rủi ro yêu cầu sự phối hợp nhuần nhuyễn giữa các bộ phận và nhân viên trong ngân hàng để đảm bảo rằng tất cả các bên tham gia vào hoạt động quản trị rủi ro đều hiểu rõ vấn đề rủi ro của ngân hàng. Mục tiêu cơ bản của quản lý rủi ro là tìm cách tối đa hóa lợi nhuận và tối ưu hóa các quỹ vốn để đảm bảo khả năng thanh toán dài hạn của tổ chức ngân hàng (Gestel & Baesens, 2009).

Đối với mỗi loại rủi ro khác nhau, các ngân hàng sẽ có những cách tiếp cận và phản ứng khác nhau. Tuy nhiên, những hành động này về bản chất đều dựa trên một quy trình gồm các bước liệt kê trong Hình 1. Trong đó, bước nhận diện rủi ro là quan trọng nhất (Rosman, 2009). Trong quá trình kinh doanh, ngân hàng luôn phải đối mặt với những rủi ro tiềm ẩn. Tuy nhiên, thường có những dấu hiệu để nhận biết những rủi ro đó. Nhận dạng rủi ro là một quá trình kết hợp nhiều phương pháp nhằm nhận dạng liên tục và có hệ thống.

Ví dụ, đối với hoạt động tín dụng, các rủi ro có thể xuất hiện bất cứ thời điểm nào, kể từ khi ngân hàng thẩm định khoản vay cho đến khi khách hàng hoàn trả cả gốc và lãi. Bất kỳ khoản vay nào cũng có thể xảy ra sự cố, việc phát hiện sớm vấn đề và sử dụng các biện pháp theo dõi kịp thời, chuyên nghiệp có thể giúp hạn chế tổn thất ở mức thấp nhất. Một số phương pháp dùng để nhận diện rủi ro tín dụng có thể kể đến là: phân tích tình hình tài chính của khách hàng, kiểm tra thực tế, phân tích hợp đồng, thu thập thông tin (từ khách hàng hoặc Trung tâm Thông tin tín dụng CIC). Trong quá trình nhận diện rủi ro, các ngân hàng thường phối hợp nhiều phương pháp với nhau để tận dụng được ưu điểm của mỗi phương pháp, qua đó giảm thiểu khả năng bỏ sót các rủi ro chưa được phát hiện.



Hình 1. Quy trình quản trị rủi ro trong ngân hàng

Nguồn: Gestel & Baesens (2009)

Sau khi xác định và phân tích các nguyên nhân rủi ro mà ngân hàng có thể gặp phải trong hoạt động kinh doanh, ngân hàng tiến hành đánh giá và đo lường rủi ro. Đo lường rủi ro bao gồm việc tính toán những con số cụ thể về mức độ rủi ro mà ngân hàng gặp phải và tính toán những tổn thất do rủi ro gây ra (Rosman, 2009). Để đo lường rủi ro, các nhà quản lý ngân hàng thường phải sử dụng các mô hình đo lường rủi ro. Trong đó, hai loại mô hình phân tích rủi ro tiêu biểu và được sử dụng phổ biến là mô hình định lượng và mô hình định tính. Các mô hình phân tích rủi ro này không loại trừ lẫn nhau; ngân hàng có thể đồng thời phối hợp nhiều mô hình để bổ sung cho nhau hoặc đánh giá mức độ rủi ro theo các cách tiếp cận khác nhau (Brown & Moles, 2014).

Theo Nguyễn (2009), kiểm soát rủi ro là việc sử dụng các biện pháp, kỹ thuật, công cụ, chiến lược nhằm kiểm soát và hạn chế đến mức thiệt hại thấp nhất mà tổn thất từ rủi ro có thể gây ra cho các ngân hàng. Ngân hàng sẽ thực hiện kiểm soát rủi ro thông qua việc soạn thảo và ban hành hệ thống chính sách, sau đó triển khai đến tất cả các bộ phận của ngân hàng. Theo Phung (2017), các biện pháp kiểm soát rủi ro thường được sử dụng gồm: ngăn chặn rủi ro, phòng ngừa rủi ro, giảm thiểu tổn thất, chuyển giao rủi ro, đa dạng hóa.

Ngoài các bước chính trên, một số nghiên cứu cho rằng, quy trình quản trị rủi ro trong ngân hàng bao gồm cả bước xử lý tổn thất từ rủi ro. Theo Phung (2017), xử lý tổn thất là việc sử dụng các biện pháp bù đắp rủi ro hiện có, nhằm đảm bảo cho hoạt động của ngân hàng không bị ảnh hưởng. Trong các NHTM, các biện pháp xử lý rủi ro được cụ thể hóa bởi tính chất, mức độ tổn thất và nguồn vốn để xử lý rủi

ro. Cuối cùng, sau khi đã thiết lập được một khung quy trình như Hình 1, hoạt động quản trị rủi ro được triển khai trên phạm vi toàn ngân hàng và đòi hỏi sự tham gia của nhiều bên như bộ phận cấp tín dụng, bộ phận kế toán tài chính để đảm bảo rủi ro được duy trì ở mức chấp nhận được.

2.2 Vấn đề mô hình chấm điểm tín dụng

Ý tưởng chấm điểm tín dụng đã được giới thiệu cách đây hơn 60 năm dựa trên những cơ sở về lý thuyết thống kê (Thomas & cộng sự, 2002). Cơ sở lý thuyết cho việc chấm điểm tín dụng (CĐTD) được bắt nguồn từ kỹ thuật thống kê phân tích biệt thức do Ronald A. Fisher phát triển. Phân tích biệt thức là một kỹ thuật thống kê được sử dụng để phân biệt các nhóm trong tổng thể thông qua các thuộc tính có thể đo lường được. Năm 1941, David Durand nhận ra rằng phương pháp của Fisher có thể sẽ mang lại hiệu quả khi áp dụng vào bài toán phân biệt các khoản nợ tốt và xấu. Từ đó, kỹ thuật CĐTD này được đưa vào trong hoạt động ngân hàng.

Việc xây dựng một mô hình tính điểm đòi hỏi các nhà phát triển phải có đủ dữ liệu lịch sử bao gồm cả điều kiện kinh tế tốt lẫn xấu và phản ánh hiệu suất của các khoản vay đã thực hiện trước đó để xác định những đặc điểm nào của người vay là hữu ích trong việc dự đoán liệu khoản vay đó có hoạt động tốt hay không. Các dữ liệu như thu nhập hàng tháng của người nộp đơn, nợ chưa thanh toán, tài sản tài chính, người nộp đơn đã làm công việc hiện tại trong bao lâu, liệu người nộp đơn có bị vỡ nợ hoặc đã từng quá hạn đối với khoản vay trong quá khứ hay không, liệu người nộp đơn sở hữu hay thuê nhà và loại tài khoản ngân hàng mà người nộp đơn sử dụng là tất cả các yếu tố tiềm ẩn có thể liên quan đến tỷ lệ vỡ nợ của người vay và có thể xem xét sử dụng để dự báo xác suất vỡ nợ của người vay tiềm năng. Sau đó, các nhà phát triển sẽ đưa ra một mô hình tính điểm tín dụng. Đây là một mô hình toán học có thể được sử dụng để kết hợp giá trị của các đặc điểm đã chọn. Mô hình CĐTD sẽ cho ra kết quả là điểm của mỗi người đi vay dựa theo mức độ rủi ro.

Theo Frankel (2006), điểm tín dụng thấp hơn có liên quan một cách có hệ thống đến xác suất vỡ nợ cao hơn đối với các khoản vay thế chấp. Một mô hình CĐTD tốt nên phân biệt rõ ràng giữa những người vay tốt, những người luôn trả nợ và những người vay xấu không trả được nợ. Do đó, một mô hình được thiết kế tốt nên chỉ định nhiều ứng viên có điểm cao và thấp hơn, đồng thời giảm thiểu số lượng ứng viên có điểm ở mức giữa. Dựa theo chức năng và vị trí trong vòng đời của khách hàng, mô hình CĐTD có thể chia thành ba loại như sau: mô hình chấm điểm tín dụng khoản vay mới (Application Score – Ascore), mô hình chấm điểm tín dụng hành vi (Behavioral Score – Bscore), mô hình chấm điểm tín dụng thu hồi nợ (Collection Score – Cscore). Nghiên cứu này xây dựng mô hình chấm điểm tín dụng hành vi.

3. Phương pháp nghiên cứu

3.1 Mô hình Logistic và ứng dụng trong việc thiết kế thẻ điểm

Mô hình Logistic hay hồi quy Logistic là một phương pháp phổ biến được sử dụng trong các bài toán phân loại, đặc biệt là trong các lĩnh vực kinh tế. Hồi quy Logistic là phương pháp phổ biến được sử dụng để xây dựng các mô hình dự báo rủi ro, trong đó biến kết quả luôn là biến nhị phân (“tốt hoặc xấu”). Kết quả của hồi quy Logistic là xác suất xảy ra của một sự kiện; tương ứng với sự tăng lên của một đơn vị trong biến dự báo, xác suất xảy ra của sự kiện cũng tăng/giảm.

Hồi quy Logistic có dạng:

$$\text{prob}(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}.$$

Để ước lượng, mô hình Logistic được chuyển thành dạng tuyến tính như sau:

$$\ln = \left(\frac{\text{prob}(x)}{1 + \text{prob}(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

Mô hình Logistic sử dụng phương pháp ước lượng hợp lý cực đại (Maximum Likelihood Estimation) để tìm ra các hệ số từ β_0 đến β_k .

3.2 Ứng dụng mô hình Logistic trong thiết kế thẻ điểm

Mục đích xây dựng mô hình CĐTD không dừng lại ở việc ước lượng tham số và đưa ra dự đoán xác suất vỡ nợ của một khoản vay. Kết quả cuối cùng mà mô hình đem lại phải chỉ ra được điểm số của mỗi khách hàng để từ đó ngân hàng có thể đưa ra ngưỡng điểm phê duyệt khoản vay phù hợp với mục tiêu phê duyệt và tỷ lệ nợ xấu trong các chiến lược được định trước. Ví dụ, ngân hàng đặt ra chiến lược và mục tiêu kinh doanh thận trọng hơn so với năm trước nên sẽ điều chỉnh ngưỡng điểm để chấp nhận cho vay cao hơn. Điều này yêu cầu ngân hàng phải tạo ra được một thẻ điểm. Trên thực tế, việc thiết kế thẻ điểm sẽ mang lại cho ngân hàng một số lợi ích như: tạo cơ sở để giải thích cơ chế đồng ý và từ chối dựa trên hạng điểm cho khách hàng, kiểm toán viên, quản lý cấp cao và các nhân viên khác bằng các thuật ngữ kinh doanh đơn giản. Quá trình phát triển cho các thẻ điểm được tiếp cận một cách rõ ràng và dễ thực hiện. Như vậy, thẻ điểm có thể dễ dàng đáp ứng bất kỳ yêu cầu quy định nào về tính minh bạch của phương pháp CĐTD.

Thẻ điểm rất dễ theo dõi và giám sát mà không yêu cầu về kiến thức thống kê hoặc lập trình. Điều này làm cho thẻ điểm trở thành một công cụ hiệu quả để quản lý rủi ro. Bảng dưới đây minh họa một thẻ điểm truyền thống với 4 chỉ tiêu chấm điểm. Dựa trên các đặc điểm thông tin của khách hàng A, điểm tín dụng của khách hàng sẽ là 81 điểm. Ngân hàng sẽ dựa trên số điểm này để quyết định cấp hạn mức cho khách hàng hay không trên cơ sở các chính sách tín dụng đã được đặt ra trước.

Bảng 1. Minh họa thẻ điểm truyền thống

Biến	Giá trị	Điểm	Khách hàng A
Tuổi	18-25	5	
	25-32	18	X
	32+	30	
Quan hệ với ngân hàng	Ngắn hơn 1 năm	3	
	Từ 1-2 năm	8	
	Từ 2-5 năm	15	X
Tỷ lệ sử dụng hạn mức	Dưới 60%	34	
	60%-90%	28	X
	Trên 90%	10	
Đã từng quá hạn	Có	0	
	Chưa	20	X
Tổng điểm			81

Nguồn: Tổng hợp của nhóm tác giả

Bài nghiên cứu này sẽ sử dụng một số kỹ thuật để thiết kế ra một thẻ điểm truyền thống (Bảng 1) dựa trên những kết quả ước lượng được từ mô hình. Với phương pháp này, điểm số được tính dựa trên mối quan hệ với tỷ lệ odds vừa được dự đoán từ mô hình Logistic, ví dụ tỷ lệ odds sẽ tăng 2 lần sau mỗi lần giảm 20 điểm. Mối quan hệ giữa tỷ lệ odds và điểm số có thể được trình bày dưới dạng một phép biến đổi tuyến tính:

$$Score = offset + factor * \ln(odds).$$

Khi sử dụng phương pháp này, các tham số được cố định gọi là điểm cơ sở (base point), tỷ lệ odds ở ngưỡng điểm cơ sở và điểm để nhân đôi tỷ lệ odds (pdo). Ý nghĩa của hai tham số này là tại mức điểm base point, cứ giảm thêm một số điểm bằng pdo thì tỷ lệ odds sẽ tăng lên gấp đôi.

Từ các tham số base point và pdo được xác định trước, hệ số và phần bù có thể dễ dàng được tính bằng cách sử dụng hệ phương trình sau:

$$base\ point = offset + factor * \ln(odds),$$

$$base\ point = offset + factor * \ln(2 * odds).$$

Giải hệ phương trình trên, ta được:

$$factor = \frac{pdo}{\ln(2)},$$

$$\alpha = base\ point - [factor * \ln(odds)].$$

Ví dụ, nếu một thẻ điểm được thiết kế trong đó tại mức điểm 500 có tỷ lệ odds là 1 phần 50 và người xây dựng mô hình muốn tỷ lệ odds tăng gấp đôi sau mỗi lần giảm 20 điểm (tức là pdo bằng -20), hệ số tự do và hệ số chặn sẽ là:

$$factor = \frac{pdo}{\ln(2)} = -28,8539,$$

$$offset = 500 + \left[28,8539 * \ln \frac{1}{50} \right] = 387,1229.$$

Từ đó, mức điểm tương ứng với mỗi tỷ lệ odds sẽ được tính bằng công thức dưới đây:

$$Score = 387,1229 - 28,8539 * \ln(odds).$$

Công thức này có thể áp dụng để tính toán điểm số cho bất kỳ trường hợp nào, dựa trên mô hình có thể tạo ra xác suất vỡ nợ hoặc tỷ lệ odds. Mặt khác, tỷ lệ odds đã được dự đoán từ các biến độc lập dựa trên phương trình hồi quy Logistic ở phía trên. Vì vậy, ta có thể rút ra một phương trình thể hiện mối quan hệ giữa điểm số và các biến độc lập như sau:

$$Score = offset + factor * (\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k),$$

$$Score = (offset + factor * \beta_0) + \sum_{i=1}^k factor * \beta_i * x_i.$$

Tóm lại, việc quy đổi điểm từ mô hình yêu cầu người xây dựng cần xác định ba tham số base point, tỷ lệ odds tại base point và pdo. Từ các tham số đó, kết hợp với kết quả ước lượng từ mô hình, chúng ta có thể tính điểm số cho mỗi khách hàng và phân tích được nếu giá trị một biến thay đổi thì sẽ làm ảnh hưởng đến điểm số như thế nào.

3.3 Biến số và thước đo

Biến phụ thuộc là trạng thái khoản vay (loan_satus), nhận giá trị bằng 1 nếu đó là khoản vay tốt và nhận giá trị bằng 0 nếu đó khoản vay xấu. Việc xác định khoản vay của khách hàng là tốt hay xấu được căn cứ vào trạng thái khoản vay, nếu khoản vay được thanh toán đầy đủ thì đó là khoản vay tốt. Ngược lại, khoản vay nào không thể trả nợ hoặc trả chậm trên 120 ngày thì được cho là khoản vay xấu (Hưng & Trang, 2018).

Biến độc lập gồm có 10 biến là: kỳ hạn của khoản vay (term); lãi suất thỏa thuận của khoản vay (int_rate); tỷ lệ giữa tổng số tiền thanh toán nợ hàng tháng của bên vay và thu nhập hàng tháng của bên vay (dti); trình độ học vấn của người vay (education), một biến định tính gồm có 6 thuộc tính là Trung học, Trung cấp, Cao đẳng, Đại học, Sau đại học và khác; tình trạng sở hữu nhà của người vay (home_ownership), gồm ba thuộc tính là đi thuê, sở hữu và khác; thời gian làm việc của người vay (emp_length)

được tính bằng năm; tình trạng hôn nhân của người vay (marital_status) gồm bốn thuộc tính là độc thân, kết hôn, ly dị và góa; số tài khoản các loại mà người vay mở trong 6 tháng vừa qua (open_acc_6m); số ngày quá hạn trả nợ lớn nhất trong 6 tháng vừa qua của người vay (dpd_last_6mths); tỷ lệ giữa tổng số dư hiện tại trên hạng mức tín dụng cao nhất cho tất cả các tài khoản thẻ ngân hàng của người vay (bc_util).

3.4. Dữ liệu nghiên cứu

Bài viết sử dụng dữ liệu thứ cấp của 19751 người vay là khách hàng cá nhân tại một ngân hàng thương mại tại Việt Nam năm 2021. Do yêu cầu bảo mật nên thông tin chi tiết của ngân hàng thương mại cung cấp số liệu xin phép được bảo mật. Thông tin của người vay được thu thập từ hồ sơ khởi tạo các khoản vay, sau đó các trường thông tin phái sinh khác được tổng hợp dựa trên cơ sở dữ liệu lịch sử của ngân hàng. Các khoản vay được ghi nhận với 150 trường thông tin nhưng bài nghiên cứu này chỉ sử dụng các trường thông tin liên quan đến các biến số được xác định trong mô hình.

4. Kết quả nghiên cứu và thảo luận

Kết quả ước lượng cho mô hình hồi quy Logistic được thể hiện ở Bảng 2. Theo đó, các hệ số góc đều có ý nghĩa thống kê ở mức 1% và mang dấu dương.

Bảng 2. Kết quả ước lượng mô hình hồi quy Logistic

Biến số	Hệ số	Sai số chuẩn	Thống kê t	p-value
Hệ số chặn	-1,1106	0,0470	-23,6440	0,0000
dti	0,6852	0,0560	12,1750	0,0000
open_acc_6m	0,4359	0,1360	3,2020	0,0010
bc_util	0,8488	0,0720	11,8010	0,0000
term	1,6029	0,0830	19,2950	0,0000
emp_length	0,7690	0,1340	5,7570	0,0000
education	0,5598	0,0690	8,1570	0,0000
int_rate	0,5268	0,0330	16,0540	0,0000
marital_status	0,5707	0,1120	5,1110	0,0000
home_ownership	0,3539	0,0780	4,5130	0,0000
dpd_last_6mths	0,9950	0,0210	47,3310	0,0000

Nguồn: Tổng hợp của nhóm tác giả

Từ kết quả ước lượng ở Bảng 2, xác suất vỡ nợ của một khoản vay được xác định là:

$$prob(x) = \frac{e^z}{1 + e^z}.$$

với z được tính bằng:

$$z = \ln \frac{prob(x)}{1 - prob(x)} = -1,1106 + 0,6852 * dti + 1,6029 * term + \\ + 0,4359 * open_acc_6m_ + 0,8488 * bc_util + 0,9950 * dpd_last_6mths + \\ 0,7690 * emp_length + 0,5598 * education + 0,5707 * marital_status + \\ 0,3539 * home_ownership + 0,5268 * int_rate.$$

Với các kết quả ước lượng thu được ở Bảng 2, nghiên cứu thực hiện các phép biến đổi để tạo ra thẻ điểm với các điểm thành phần đóng góp từ các biến. Dựa theo phương pháp tạo thẻ điểm đề cập phía trên, bài nghiên cứu cố định các tham số base point là 500 điểm, tỷ lệ odds tại base point là 1:50, số điểm nhân đôi tỷ lệ odds pđo bằng -20. Khi đó,

$$Score = (offset + factor * \beta_0) + \sum_1^k factor * \beta_i * x_i.$$

Với $offset + factor * \beta_0$ là điểm cơ sở cho sẵn đối với tất cả khách hàng, $factor * \beta_i * x_i$ là điểm thành phần từ mỗi biến, thẻ điểm được xây dựng như Bảng 3.

Bảng 3. Kết quả thẻ điểm dựa trên kết quả ước lượng mô hình

Biến	Nhóm giá trị	Điểm	Hệ số ước lượng
Điểm cơ sở		355	
dti	Dưới 13%	9	0,6852
dti	13-18%	4	0,6852
dti	18-23%	0	0,6852
dti	23-32%	-6	0,6852
dti	32% +	-15	0,6852
open_acc_6m	0	2	0,4359
open_acc_6m	1	0	0,4359
open_acc_6m	2	-1	0,4359
open_acc_6m	3	-3	0,4359
open_acc_6m	4+	-5	0,4359
bc_util	Dưới 30%	10	0,8488
bc_util	30-45%	5	0,8488
bc_util	45-75%	-1	0,8488
bc_util	75% +	-8	0,8488
term	Từ dưới 12 tháng	9	1,6029
term	Trên 12 tháng	-22	1,6029

Biến	Nhóm giá trị	Điểm	Hệ số ước lượng
emp_length	Trên 10 năm	3	0,7690
emp_length	5-10 năm	1	0,7690
emp_length	1-4 năm	-1	0,7690
emp_length	Khác	-10	0,7690
education	Đại học & Sau Đại học	6	0,5598
education	Cao đẳng & Trung cấp	0	0,5598
education	Trung học & Khác	-6	0,5598
int_rate	Dưới 8%	21	0,5268
int_rate	8-11%	9	0,5268
int_rate	11-17%	-2	0,5268
int_rate	17-22%	-11	0,5268
int_rate	Trên 22%	-17	0,5268
marital_status	Kết hôn	3	0,5707
marital_status	Độc thân	-2	0,5707
marital_status	Ly dị & Góa	-5	0,5707
marital_status	Khác	-8	0,5707
home_ownership	Thế chấp; Khác	3	0,3539
home_ownership	Sở hữu	0	0,3539
home_ownership	Thuê	-3	0,3539
dpd_last_6mths	10+	-62	0,9950
dpd_last_6mths	1 - 10	106	0,9950
dpd_last_6mths	0	179	0,9950

Nguồn: Tính toán của nhóm tác giả

Có thể thấy, số điểm thành phần cho mỗi thuộc tính của từng biến tuân theo đúng xu hướng mối quan hệ giữa biến đó và rủi ro vỡ nợ của khách hàng. Ví dụ, đối với biến số năm kinh nghiệm làm việc *emp_length*, các khách hàng có càng nhiều năm kinh nghiệm làm việc thì càng có điểm cao hơn. Cụ thể, đối với các khách hàng có từ 1-4 năm kinh nghiệm sẽ nhận -1 điểm, các khách hàng có từ 5-10 năm kinh nghiệm sẽ nhận 1 điểm, các khách hàng có nhiều hơn 10 năm kinh nghiệm sẽ nhận 3 điểm. Đặc biệt, đối với các khách hàng không điền thông tin về số năm kinh nghiệm sẽ nhận số điểm thấp nhất là -10. Điều này có thể giải thích rằng những khách hàng này đang cố che giấu một tín hiệu nào đó có thể mang xu hướng rủi ro.

Từ thẻ điểm, với mỗi khách hàng, ta sẽ cộng điểm cơ sở với điểm thành phần từ các biến để cho ra tổng điểm cuối cùng. Khách hàng có tổng điểm càng cao thì rủi

ro càng thấp, ngược lại, khách hàng có điểm càng thấp thì rủi ro vỡ nợ càng cao. Từ đó, người sử dụng thẻ điểm có thể phân tích phân bố điểm của cả danh mục để đưa ra những chính sách phê duyệt tín dụng dựa theo ngưỡng điểm phù hợp.

5. Kết luận

Việc xây dựng mô hình CĐTD đối với khách hàng cá nhân có vai trò quan trọng trong hoạt động kinh doanh của ngân hàng nói chung và đặc biệt hoạt động quản trị rủi ro của các ngân hàng nói riêng. Kết quả nghiên cứu cho thấy, chất lượng của bảng CĐTD sẽ ảnh hưởng lớn đến quyết định các khoản vay của khách hàng cá nhân. Do đó, để có quyết định chính xác và hạn chế rủi ro, nhà quản trị cần chủ động trong việc thu thập và lưu trữ dữ liệu khách hàng. Chất lượng, độ chính xác và độ sâu của dữ liệu của khách hàng sẽ quyết định giá trị của thông tin của khách hàng, góp phần tăng hiệu quả của mô hình CĐTD.

Bài viết này cung cấp bằng chứng khoa học quan trọng cho nhà quản trị rủi ro của các ngân hàng thương mại tại Việt Nam. Các nhà quản trị rủi ro có thể ứng dụng phương pháp, ứng dụng mô hình CĐTD này để xác định điểm tín dụng cho khách hàng của mình, từ đó hỗ trợ cho quá trình ra quyết định.

Bên cạnh những đóng góp, bài viết tồn tại một số hạn chế. *Thứ nhất*, bài viết mới chỉ đề xuất một phương pháp luận (phương pháp hồi quy Logistic) để có thể xây dựng mô hình CĐTD. Trong khi đó, trên thế giới, mô hình CĐTD có thể được xây dựng bằng nhiều phương pháp hiện đại, trong đó có những phương pháp cho phép dự đoán chính xác và hiệu quả hơn so với các phương pháp thống kê truyền thống. *Thứ hai*, bài nghiên cứu này chưa đo lường được giá trị kinh tế mang lại từ các quyết định được đưa ra dựa trên mô hình CĐTD. Điều này xuất phát từ sự hạn chế của dữ liệu nên không có cơ sở để tính toán mô phỏng giá trị kinh tế của các khoản vay, để từ đó chỉ ra được giá trị kinh tế mà mô hình đem lại. Các nghiên cứu tiếp theo, chúng tôi sẽ cố gắng khắc phục những hạn chế này.

Tài liệu tham khảo

- Brown, K. & Moles, P. (2014), *Credit Risk Management*, Edinburgh Business School Heriot-Watt University, United Kingdom.
- Bùi, L.T.L., Hoàng, N.H. & Nguyễn, V.K. (2016), “Ứng dụng mô hình Logistic chấm điểm khách hàng nộp hồ sơ vay tại NHTM”, trong *Kỷ yếu Hội thi NCKH sinh viên toàn quốc “Olympic Kinh tế lượng và ứng dụng”*, tr. 99-131.
- Crook, J.N., Edelman, D.B. & Thomas, L.C. (2007), “Recent developments in consumer credit risk assessment”, *European Journal of Operational Research*, Vol. 183, pp. 1447-1465.
- Finlay, S. (2011), “Multiple classifier architectures and their application to credit risk assessment”, *European Journal of Operational Research*, Vol. 210 No. 2, pp. 368-378.
- Frankel, J. (2006), “The effect of monetary policy on real commodity prices”, *National Bureau of Economic Research*, 12713.

- Gestel, T.V. & Baesens, B. (2009), *Credit Risk Management Basic Concepts*, Oxford University Press Inc., New York.
- Hovdenakk, A.H. (2021), *Machine Learning vs Logistic Regression in Credit Scoring: a Trade-off Between Accuracy and Interpretability*, Master thesis, The University of Bergen.
- Hung, N.T. & Trang, L.T.H. (2018), “Mô hình chấm điểm tín dụng dựa trên sự kết hợp giữa mô hình Cây quyết định, Logit, K láng giềng gần nhất và Mạng thần kinh nhân tạo”, *Tạp chí Khoa học và Đào tạo Ngân hàng*, Số 193, tr. 46-54.
- Kruppa, J., Schwarz, A., Armingier, G. & Ziegler, A. (2013), “Consumer credit risk: Individual probability estimates using machine learning”, *Expert Systems with Applications*, Vol. 40 No. 13, pp. 5125-5131.
- Lê, T.H. (2017), “Kiểm soát rủi ro tín dụng theo Basel II tại các ngân hàng thương mại Việt Nam”, *Tạp chí Tài chính Online*.
- Muriuki, B.M. (2021), *Logistical regression model for predicting small and medium enterprises (SMEs): Credit risk for commercial banks in Kenya*, Master Thesis, KCA University, Kenya.
- Ngân hàng Nhà nước (2022), "Bức tranh nợ xấu của ngành Ngân hàng và một số kiến nghị", <https://sbv.gov.vn/webcenter/portal/vi/links/cm255?dDocName=SBV489213>, truy cập ngày 01/12/2022.
- Nguyễn, M.K. (2009), *Quản trị rủi ro tài chính*, Nhà Xuất bản Thống kê, Việt Nam.
- Phan, X.V. (2018), *Ứng dụng mô hình phân tích sống sót trong đo lường rủi ro tín dụng khách hàng cá nhân tại Ngân hàng TMCP Ngoại thương Việt Nam tại TP.HCM*, Luận văn Thạc sĩ, Trường Đại học Kinh tế Hồ Chí Minh.
- Phung, L.S. (2017), *Credit Risk Management for Personal Banking at Military Bank - Son Tay Town Branch*, Master Thesis, International School – Vietnam National University Hanoi and University de Nantes
- Rosman, R. (2009), “Risk management practices and risk management processes of Islamic banks: a proposed framework”, *International Review of Business Research Papers*, Vol. 5 No. 1, pp. 242-254.
- Thomas, L.C., Edelman, D.B. & Crook, J.N. (2002), “Credit Scoring and Its Applications”, *The Society for Industrial and Applied Mathematics*, Philadelphia, USA.
- Trần, C.C. (2020), “Phương pháp xếp hạng nội bộ dưới hiệp ước Basel II và thực tế áp dụng tại các NHTM Việt Nam”, *Tạp chí Công Thương - Các kết quả nghiên cứu khoa học và ứng dụng công nghệ*, Số 26, tr. 319-325.