

Mã số: 461

Ngày nhận: 27/11/2017

Ngày gửi phản biện lần 1: /2017

Ngày gửi phản biện lần 2:

Ngày hoàn thành biên tập: 29/1/2018

Ngày duyệt đăng: 29/1/2018

ỨNG DỤNG MÔ HÌNH LOGISTIC CHẤM ĐIỂM KHÁCH HÀNG CÁ NHÂN NỘP HỒ SƠ VAY TRÊN LENDINGCLUB

Nguyễn Thị Thúy Quỳnh¹

Trần Thị Xuân Anh²

Bùi Lê Trà Linh³

Tóm tắt: Bài nghiên cứu ứng dụng mô hình logistic chấm điểm 235.629 khách hàng cá nhân nộp hồ sơ vay vốn tại Lendingclub - một trong những tổ chức cho vay ngang hàng (Peer - to - Peer) đầu tiên tại Mỹ. Dữ liệu nghiên cứu được thực hiện trong hai năm 2014-2015 bao gồm 111 đặc điểm của khách hàng vay vốn. Nhóm nghiên cứu đã sử dụng mô hình Logistic để đo lường xác suất một khách hàng vay vốn là khách hàng tốt. Trên cơ sở đó, tổ chức tín dụng sẽ xác định được mức rủi ro tín dụng của khách hàng vay vốn nhằm đưa ra quyết định cho vay phù hợp. Nếu người đi vay có xác suất là khách hàng tốt thấp hơn mức xác suất đưa ra thì sẽ bị từ chối khoản vay, ngược lại người đi vay có xác suất là khách hàng tốt cao hơn mức xác suất đưa ra thì được chấp nhận khoản vay đó tùy vào mức độ sẵn sàng chấp nhận rủi ro tín dụng từ tổ chức cho vay.

Từ khóa: chấm điểm khách hàng, logistic, xếp hạng tín dụng, xếp hạng tín nhiệm.

Abstract: The paper applied logistics model for rating 235.629 individual borrowers in lendingclub - one of the first peer - to - peer lending institution in US. The research is implemented based on date of two years from 2014 to 2015, including 111 characteristics of clients. The paper used the logistic model to measure the probability that a customer is a good one. Accordingly, the credit institution will determine the level of credit risk of the borrower to make appropriate loan decision. If the borrower has a good probability of being well below the given probability, then the loan will be denied. In contrast, if the borrower has a probability of being a good

¹ Học viện Tài chính, Email: nguyenthithuyquynh@hvtc.edu.vn

² Học viện Ngân hàng, Email: ttxahvnh@gmail.com

³ Hiệp hội Doanh nghiệp Châu Âu tại Việt Nam, Email: tralinhbuile@gmail.com

client higher than the given probability, then the loan is accepted. The given probability depends on the willingness to accept credit risk from the lender.

Keywords: client rating, logistic, credit rating.

1. Giới thiệu về chấm điểm khách hàng và xếp hạng tín dụng

Tín dụng là một trong những hoạt động chính mang lại lợi nhuận cho các ngân hàng thương mại(NHTM). Trong quá trình cấp tín dụng, các ngân hàng thường đối mặt với nguy cơ rủi ro từ nhiều nguyên nhân khác nhau và thường chung một hệ quả là khách hàng không thực hiện được hoặc không thực hiện đầy đủ các nghĩa vụ tài chính khi đến hạn. Những nguy cơ rủi ro khó loại trừ hoàn toàn mà chỉ có thể hạn chế và phòng ngừa. Một trong những biện pháp hạn chế rủi ro tín dụng hiệu quả và phổ biến hiện nay là xây dựng mô hìnhchấm điểm khách hàng nhằm xếp hạng tín dụng, phân loại khách hàng, hỗ trợ công tác ra quyết định và quản lý tín dụng.

Xếp hạng tín dụng (XHTD)phát triển mạnh từ nhiều năm nay ở Mỹ và các nước Châu Âu. Ba công ty đánh giá tín dụng lớn nhất trên thế giới hiện nay là công ty Standard & Poor's (S&P), Moody's và Fitch Group. S&P và Moody's có trụ sở ở Mỹ, Fitch trụ sở tại cả Mỹ và Anh và do FIMALAC của Pháp kiểm soát.

Trong quá khứ, các tổ chức tín dụng thường sử dụng các mô hình chấm điểm khách hàng và XHTD để đánh giá và phân loại khách hàng vay nợ. Abdou và Pointon (2011) đã hệ thống hoá 2 phương pháp XHTD chính được sử dụng gồm phương pháp chuyên gia và phương pháp thống kê. Phương pháp chuyên gia là phương pháp thu thập và xử lý những đánh giá dự báo bằng cách tập hợp và hỏi ý kiến các chuyên gia giỏi trong lĩnh vực tài chính ngân hàng để xác định rủi ro và chất lượng của khoản tín dụng. Phương pháp này mất nhiều chi phí và thời gian do cần số lượng chuyên gia lớn tham gia đánh giá. Phương pháp thống kê dựa trên các số liệu thực tiễn như mức độ nợ, khả năng trả nợ... và phương pháp kiểm định thống kê để phát hiện các biến số ảnh hưởng tới rủi ro tín dụng. Sự phù hợp của mô hình thống kê phụ thuộc rất lớn vào chất lượng của bộ dữ liệu thực nghiệm. Bộ dữ liệu cần đủ lớn và chính xác thì mô hình thống kê sẽ đưa ra kết quả có ý nghĩa.

Ở Việt Nam, hoạt động XHTD được phát triển từ năm 2002. Trước đó, việc xem xét cấp tín dụng thông thường dựa vào đánh giá chủ quan và mang cảm tính lớn của những người xét cấp tín dụng nên dễ dẫn đến những rủi ro khó lường hoặc có thể mất cơ hội của người đi vay. Những rủi ro gặp phải có thể dẫn đến sự đổ vỡ của ngân hàng...

Năm 1988, Ủy ban Basel về giám sát ngân hàng đã ban hành hiệp ước Basel II nhằm chuẩn mực hóa hoạt động ngân hàng theo xu hướng toàn cầu, thiết lập một hệ thống ngân hàng quốc tế ổn định, thống nhất, bình đẳng, giảm cạnh tranh không lành mạnh đã tạo ra sự thay đổi về quản trị rủi ro trong các ngân hàng, giúp các ngân hàng hoạt động an toàn hơn.

Ở Việt Nam, việc áp dụng Basel II tại các NHTM đang gặp không ít khó khăn về chi phí tài chính, hệ thống cơ sở dữ liệu và các quy định của ngân hàng nhà nước trong việc hiệu chỉnh các quy định Basel II phù hợp với điều kiện thực tế. Vì vậy, việc tìm kiếm mô hình, phương thức đánh giá rủi ro tín dụng của khách hàng vay tại các NHTM là cần thiết, đặc biệt trong bối cảnh tự do hoá tài chính hiện nay.

Bài viết trình bày kết quả nghiên cứu ứng dụng mô hình Logistic trên cơ sở dữ liệu khách hàng vay vốn tại Lendingclub (LC), một tổ chức cho vay ngang hàng lớn nhất trên thế giới hiện nay, nhằm đưa ra minh chứng thực nghiệm về tính hiệu quả, khách quan của mô hình XHTD nói chung, mô hình Logistic nói riêng, từ đó khuyến nghị đối với các tổ chức tín dụng tại Việt Nam về việc áp dụng mô hình logistic xếp hạng khách hàng cá nhân vay vốn.

2. Cơ sở lý thuyết của mô hình Logistic

Mô hình Logistic (Maddala [12], 1992) được ứng dụng rộng rãi trong phân tích rủi ro tín dụng, ở đó dựa trên các yếu tố ảnh hưởng đến mức độ tín nhiệm của khách hàng dự báo xác suất (mức độ) xảy ra rủi ro tín dụng, quy ra mức điểm tương ứng nhằm XHTD của khách hàng và làm cơ sở xác định khoản vay phù hợp.

Một khách hàng i sẽ có thông tin đặc trưng bởi một vec tơ các biến độc lập $X_i = (X_{1i}, X_{2i}, \dots, X_{ki})$. Mục tiêu là với những thông tin thu thập được về khách hàng cần dự báo khả năng vỡ nợ của khách hàng (khách hàng là xấu ($Bads_i$)) bằng xác suất được xác định bởi công thức $P(Y = 1 | X = X_i)$ (probability of default - PD) và xác suất để khách hàng là tốt ($Goods_i$) được xác định bởi công thức $P(Y = 0 | X = X_i)$. Xác suất vỡ nợ càng cao thì điểm số tín dụng của khách hàng càng thấp. Để có bài toán tỷ lệ thuận giữa xác suất và điểm số tín dụng, bài viết sẽ đi tìm xác suất để khách hàng là tốt, tức là

$$\text{ước lượng: } p_i = P(Y = 0 | X = X_i) = \frac{1}{1 + e^{-(b_0 + bX_i)}}, \quad b_0 \in R; b = (b_1, b_2, \dots, b_k); \quad (1).$$

$$\text{Công thức (1) tương đương: } L_i = \ln \frac{p_i}{1 - p_i} = \ln(Odds_i) = Z_i = b_0 + bX_i \quad (2).$$

Đặt $f_i(Y_i) = p_i^{1 - Y_i} (1 - p_i)^{Y_i}$ là hàm phân phối xác suất của biến cố ($Y_i = 0$). Khi đó hàm phân phối xác suất đồng thời (Likelihood function - LF) của mẫu quan sát độc lập

$$\{Y_1, Y_2, \dots, Y_n\} \text{ được xác định: } f(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f_i(Y_i) = \prod_{i=1}^n p_i^{1 - Y_i} (1 - p_i)^{Y_i}.$$

Lấy logarit tự nhiên hai vế:

$$LLF = \ln f(Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^n [(1 - Y_i) \ln p_i + Y_i \ln (1 - p_i)]$$

$$= \prod_{i=1}^n [\ln p_i + Y_i \ln(1 - p_i) - Y_i \ln p_i] = \prod_{i=1}^n \left[\ln \frac{(1 - p_i)^{Y_i} p_i^{1 - Y_i}}{p_i} \right] = \prod_{i=1}^n \ln p_i$$

Thay $p_i = \frac{1}{1 + e^{-(b_0 + bX_i)}}$ thu được: $\ln f(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n \ln[1 + e^{-(b_0 + bX_i)}] - \prod_{i=1}^n Y_i(b_0 + bX_i)$

Bài toán: Tìm cực đại hàm LLF (hay LF) theo biến là các tham số b_0, b khi đã biết các giá trị $X_i, i = \overline{1, n}$. Giá trị ước lượng \hat{b}_0, \hat{b} để hàm LFF đạt cực đại, ứng với mỗi

$X_i, i = \overline{1, n}$ được thay vào công thức (1) thu được $p_i^u = \frac{1}{1 + e^{-(\hat{b}_0 + \hat{b}X_i)}}$ là xác suất dự

đoán để khách hàng i là khách hàng tốt.

Lựa chọn biến độc lập: Biến độc lập được lựa chọn dựa trên bộ dữ liệu thu thập về các đặc trưng được cho là có tác động ảnh hưởng đến việc trả nợ của khách hàng.

Thông tin cá nhân	Tình trạng việc làm, nghề nghiệp, thu nhập, tình trạng nhà ở, bản ghi về bản án và số lượng người phụ thuộc...
Lịch sử tín dụng gần đây	Chiều dài của lịch sử tín dụng, số lượng và giá trị của quá khứ vay vốn, số lượng và giá trị của các khoản vay trễ hạn trong quá khứ thường được cung cấp bởi các tổ chức cung cấp thông tin tín dụng.
Dữ liệu hành vi	Lịch sử sử dụng của tín dụng trên các sản phẩm trước đó: Số tiền chi tiêu, việc trả nợ thực tế....

Hai công cụ sử dụng lựa chọn các biến độc lập nhằm đảm bảo phân loại khách hàng tốt và khách hàng xấu là: The Weight of Evidence (*WOE*) và Information Value (*IV*). *WOE* mô tả mối quan hệ giữa một biến giải thích và biến phụ thuộc nhị phân; *IV* đo lường sức mạnh của mối quan hệ đó. Cụ thể

$$WOE_i = \ln \frac{P_i}{1 - P_i} \text{ và } IV = \sum_{i=1}^n (Distr\ Goods_i - Distr\ Bads_i) \cdot WOE_i$$

Trong đó: *Distr Goods_i* bằng tỷ số phần trăm giữa tổng khách hàng tốt của nhóm biến với tổng số khách hàng tốt trong tổng thể; *Distr Bads_i* bằng tỷ số phần trăm giữa tổng khách hàng xấu của nhóm biến với tổng số khách hàng xấu trong tổng thể.

Theo Siddiqi ([13]), nếu *IV* < 0.02 thì biến độc lập không có quan hệ với biến phụ thuộc; *IV* từ 0.02 đến 0.1 thì biến độc lập không có quan hệ quá chặt chẽ với biến phụ thuộc; *IV* từ 0.1 đến 0.3 thì biến độc lập có mối quan hệ khách chặt chẽ với biến phụ thuộc; *IV* ≥ 0.3 thì biến độc lập rất chặt chẽ với biến độc lập.

Đánh giá sự phù hợp của mô hình: Đối với mô hình Logistic thông thường có các phương pháp kiểm định tỷ số hàm hợp lý (LR), đo độ phù hợp qua tỷ lệ phần trăm dự báo đúng, kiểm định sự phù hợp Goodness of Fit test,... Tuy nhiên có thể sử dụng đường

cong Receiver Operating Characteristic (ROC) và hệ số Gini. Hệ số Gini từ 0.8-1 cho biết mô hình rất tốt; từ 0.6-0.8 là mô hình tốt; từ 0.4-0.6 là mô hình khá; từ 0.2-0.4 là mô hình trung bình; từ 0.0-0.2 là mô hình yếu.

3. Ứng dụng mô hình Logistic chấm điểm khách hàng cá nhân nộp hồ sơ vay

Trên LC người đi vay chỉ cần điền thông tin vào đơn và nộp trực tuyến. Hệ thống của LC sẽ tự động phân tích dữ liệu, đánh giá rủi ro, chấm điểm tín dụng và đưa ra mức lãi suất phù hợp. Sau đó công ty kết nối người đi vay với nhà đầu tư để nhà đầu tư lựa chọn khoản đầu tư thích hợp dựa trên những thông tin về người đi vay như XHTD, mục đích vay tiền, lịch sử tín dụng,.... Đây là hình thức cho vay mà người có tiền và người cần tiền được kết nối trực tiếp với nhau mà không cần thông qua ngân hàng.

Bài viết sử dụng bộ số liệu khách hàng cá nhân trong hai năm 2014 và 2015 trên website www.lendingclub.com để mô tả các bước ứng dụng mô hình logistic nhằm đánh giá khả năng khách hàng là tốt, chấm điểm khách hàng, chừa điểm cắt làm căn cứ giúp các nhà đầu tư ra quyết định khoản cho vay. Các bước cụ thể như sau:

❖ Phân chia biến thành các nhóm

Dữ liệu gồm 235.629 hồ sơ khách hàng đi vay, mỗi khách hàng kê khai 111 đặc điểm tương ứng với 111 biến độc lập.

Bảng 1. Mô tả nhóm biến và các biến đưa vào mô hình

Các biến trước vay	Các biến giống Id	id, member_id, địa chỉ, ...	5 biến
	Thông tin cá nhân	Tình trạng nhà ở, công việc, thu nhập, tỷ lệ nợ/ thu nhập, mô tả khoản vay,...	11 biến
	Hồ sơ tín dụng	<ul style="list-style-type: none"> - Tháng sớm nhất/muộn nhất đánh giá, mở hạn mức tín dụng, lần điều tra hồ sơ... - Số tháng kể từ lần công khai hồ sơ cuối cùng, lần quá hạn cuối cùng, cuộc điều tra gần đây, lần mở thẻ, tài khoản cuối cùng,... - Số lần điều tra, số tài khoản quay vòng mở, số tài khoản thẻ mở,... trong khoảng 6 tháng/12 tháng/24 tháng - Số lượng tài khoản trả góp, số tài khoản quay vòng, số tài khoản thẻ... - Tổng số hoặc trung bình tổng số dư hiện tại, tổng mức quay vòng tín dụng cao, tổng dư nợ tín dụng, ... - Các biến tỷ lệ sử dụng quay vòng tín dụng, tỷ lệ các tài khoản thẻ quá 75% hạn mức... 	74 biến
LC đánh giá	Gồm 2 biến grade và sub_grade thể hiện đánh giá của LC về mức độ rủi ro đối với mỗi bộ hồ sơ vay dựa trên thông tin cá nhân và thông tin hồ sơ tín dụng của người vay. Grade có các giá trị A,B,C,D,..., còn sub_grade có các giá trị A1,A2...,B1,B2,... càng gần A1 càng được đánh giá là tốt và lãi suất vay càng thấp.		2 biến

Các biến sau vay	Thông tin ban đầu khoản vay	Các biến thông tin ban đầu của khoản vay, ngay sau khi khoản vay được tài trợ, như: tổng số tiền cho vay, lãi suất, khoản tiền phải thanh toán hàng tháng,...	6 biến
	Nhà đầu tư theo dõi khoản vay	Các thông tin về tình trạng khoản vay cho đến thời điểm quan sát: Dư nợ gốc, khoản tiền nhà đầu tư nhận đến nay, tiền gốc đến nay, tiền lãi, chi phí thu hồi nợ, phí trả trễ, ...	13 biến

Nhóm tác giả thực hiện

Sử dụng phân tích định tính bài nghiên cứu lựa chọn 89 biến độc lập vào mô hình.

Tiếp tục loại bỏ các biến thiếu quan sát và xóa bỏ một vài quansát; Các biến có $IV < 0.02$ sẽ bị loại, biến có $IV > 0.02$ sẽ được chọn vào mô hình, các biến có tương quan lớn hơn 0.7 thì sẽ dựa trên IV của biến nào cao hơn sẽ giữ lại. Sử dụng Stepwise lựa chọn các biến đưa vào mô hình. Có 17 biến được lựa chọn và tính giá trị WOE tương ứng.

Bảng 2: Danh sách các biến được lựa chọn làm biến độc lập

Step	Effect Entered	DF	Score Chi-Square	Pr > Chi_Square
1	dti3w (Tỷ lệ giá trị khoản vay/thu nhập khả dụng)	1	1003.639	<.0001
2	acc_open_past_24mths (số TK giao dịch mở trong 24 tháng trước ngày nộp hồ sơ vay)	1	995.2945	<.0001
3	bc_open_to_buyw (số tiền còn lại có thể vay trong các tài khoản thẻ ngân hàng quay vòng)	1	973.1144	<.0001
4	inq_last_6mths (số lần điều tra trong 6 tháng qua)	1	525.435	<.0001
5	mo_sin_old_rev_tl_op (số tháng kể từ khi mở tài khoản đầu tiên)	1	264.2194	<.0001
6	Purposew (mục đích vay tiêu dùng, đảo nợ, tín dụng..)	1	208.7095	<.0001
7	tyle_loanw (Giá trị khoản vay trên Lending Club/ Tổng dư nợ hiện tại)	1	124.3415	<.0001
8	Termw (thời hạn vay của khoản vay)	1	121.9971	<.0001
9	mo_sin_old_il_acctw (Số tháng kể từ tài khoản ngân hàng mở đầu tiên)	1	85.1731	<.0001
10	mths_since_recent_bc (số tháng từ khi mở tài khoản thẻ gần nhất đến nay)	1	78.2574	<.0001
11	percent_bc_gt_75w (tỷ lệ phần trăm số tài khoản thẻ vượt qua 75% hạn mức tín dụng chia cho tổng số tài khoản thẻ)	1	65.4251	<.0001
12	annual_incw (thu nhập hàng năm của người vay)	1	70.8672	<.0001
13	Verification_statusw (trạng thái xác minh thu nhập của khách hàng)	1	84.6592	<.0001
14	Dtiw (% số thu nhập dành vào việc trả nợ hàng tháng)	1	35.484	<.0001
15	revol_utilw (các khoản tín dụng của người vay được sử dụng liên quan đến tất cả các tín dụng quay vòng có sẵn)	1	27.2376	<.0001

16	mths_since_recent_inq (Số tháng kể từ khi hầu hết các cuộc điều tra gần đây)	1	21.9252	<.0001
17	mo_sin_rcnt_tlw (số tháng kể từ khi lần mở tài khoản bất kỳ gần nhất đến hiện tại)	1	16.4708	<.0001

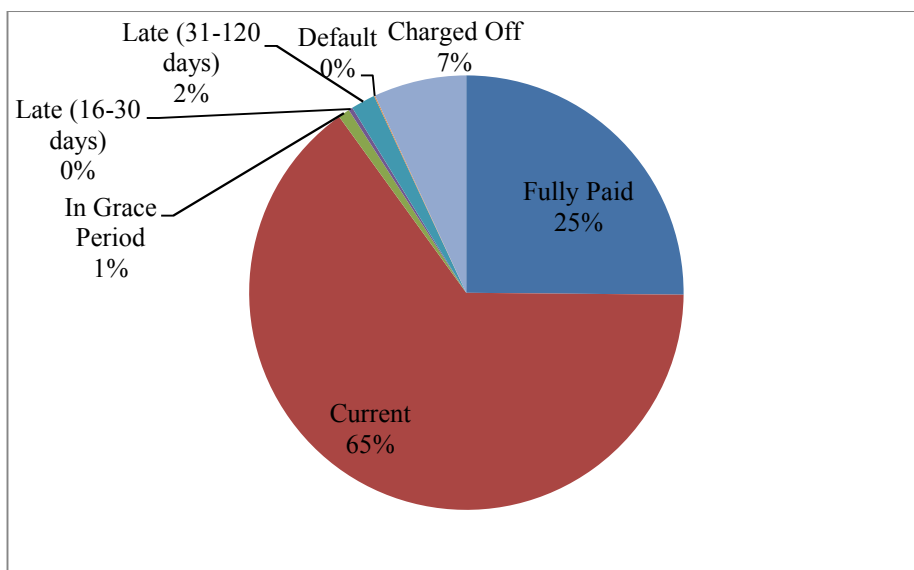
Nguồn: Truy suất kết quả từ phần mềm SAS
Nhóm tác giả thực hiện

❖ **Lựa chọn biến phụ thuộc và định nghĩa khách hàng tốt và khách hàng xấu**

Tháng 2 năm 2016, dựa vào trạng thái của các khoản vay năm 2014 trên LC, lựa chọn Loan status làm biến phụ thuộc Y gồm 7 trạng thái như sau:

- Current: Các khoản vay còn dư nợ nhưng thanh toán đầy đủ các khoản nợ hàng tháng.
- Fully Paid: Khách hàng đã thanh toán đầy đủ khoản vay.
- In Grace Period: Trong thời gian ân hạn, khách hàng được phép thanh toán khoản vay mà không phải nộp phạt (15 ngày tính từ ngày phải thanh toán khoản vay).
- Late (16-30 days): Các khoản vay trả muộn so với hạn trả từ 16 đến 30 ngày.
- Late (31-120 days): Các khoản vay trả muộn so với hạn trả nợ từ 31 đến 120 ngày.
- Default: Khách hàng bị vỡ nợ, quá hạn trên 120 ngày.
- Charged Off: Các khoản vay quá hạn, nợ không thể hoàn trả (chậm quá 150 ngày).

Biểu đồ 1. Các nhóm trong biến phụ thuộc Y trước khi phân lớp



Nhóm tác giả thực hiện

Đến cuối tháng 8 năm 2016, cập nhật trạng thái của các khoản vay và xem lại trạng thái của các khoản vay tại tháng 2 năm 2016.

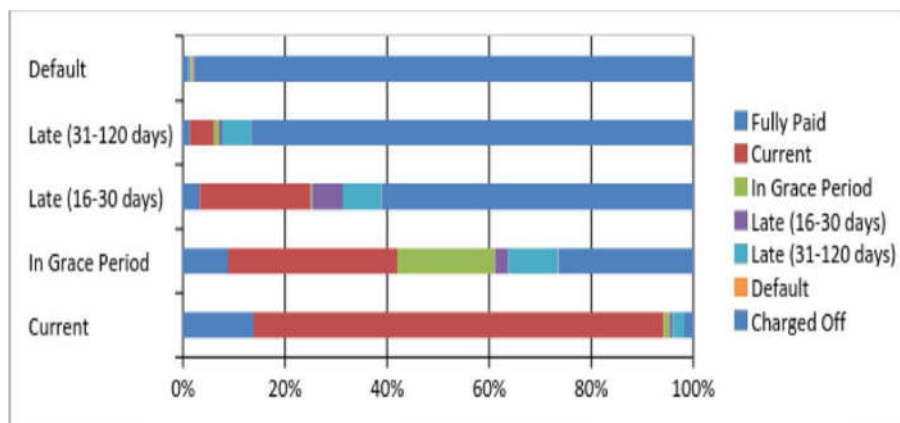
Bảng 3. Bảng tính tỷ lệ thay đổi trạng thái của khách hàng sau 8 tháng

Trạng thái trước vay	Fully Paid	Current	In Grace Period	Late (16-30 days)	Late (31-120 days)	Default	Charged Off
Fully Paid	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Current	13.88%	80.40%	1.09%	0.61%	2.29%	0.00%	1.73%

In Grace Period	8.88%	33.12%	19.14%	2.60%	9.76%	0.10%	26.40%
Late (16-30 days)	3.28%	21.79%	0.28%	5.98%	7.55%	0.00%	61.11%
Late (31-120 days)	1.40%	4.73%	0.96%	0.58%	5.60%	0.09%	86.64%
Default	1.05%	0.00%	0.00%	0.00%	0.53%	0.53%	97.89%
Charged Off	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Nhóm tác giả tính toán

Biểu đồ 2. Tỷ lệ thay đổi trạng thái của khách hàng sau 8 tháng



Nhóm tác giả thực hiện

Nhóm khách hàng có trạng thái là “Late (31-120 days)” là khách hàng xấu tương ứng với $Y = 1$. Những quan sát có trạng thái “In Grace Period” và “Late (16-30 days)” ở trạng thái trung tính sẽ được loại bỏ khỏi mẫu.

Lấy ngẫu nhiên 70% mẫu để xây dựng mô hình và 30% mẫu sử dụng để hậu kiểm.

Bảng 4. Kết quả Mô hình

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi_Square	Pr > Chi_Square
Intercept	1	2.3289	0.00912	65272.4514	<.0001
acc_open_past_24mths	1	0.7831	0.0425	339.9262	<.0001
annual_incw	1	0.5267	0.052	102.655	<.0001
dti3w	1	0.5389	0.0455	140.3659	<.0001
dtiw	1	0.259	0.0504	26.4611	<.0001
inq_last_6mthsw	1	0.7158	0.0556	165.7191	<.0001
mo_sin_old_il_acctw	1	0.5404	0.0666	65.8343	<.0001
mo_sin_old_rev_tl_op	1	0.5253	0.0485	117.4692	<.0001
mo_sin_rcnt_tl	1	0.2318	0.0571	16.4552	<.0001
mths_since_recent_bc	1	0.4349	0.0695	39.1241	<.0001
mths_since_recent_inq	1	0.2604	0.0625	17.3486	<.0001
percent_bc_gt_75w	1	0.4587	0.0748	37.5524	<.0001
purposew	1	0.9315	0.0644	209.1187	<.0001
revol_utilw	1	0.5908	0.1057	31.2147	<.0001
termw	1	0.6653	0.053	157.4741	<.0001
tyle_loanw	1	0.535	0.0523	104.5204	<.0001

verification_statusw	1	0.5056	0.0556	82.7335	<.0001
bc_open_to_buyw	1	0.6831	0.0541	159.6187	<.0001

Nguồn: Truy suất kết quả từ phần mềm SAS

Nhóm tác giả thực hiện

4. Chấm điểm khách hàng dựa trên kết quả của mô hình Logistic

Bảng 4 cho thấy mỗi khách hàng trên LCứng với một véc tơ thông tin đầu vào X_i đều tính được xác suất để khách hàng là tốt. Lấy mức xác suất khách hàng tốt nhân với 1000 thu được điểm của khách hàng đó.

Lựa chọn điểm cắt xác suất khách hàng tốt 10% (mức thông thường của các ngân hàng) tương ứng với xác suất 0.842 (tương ứng với mức điểm 842) làm căn cứ cho vay thì khách hàng nào dưới 842 điểm sẽ bị từ chối cho vay.

Bảng 5: Bảng chấm điểm khách hàng trên mẫu xây dựng mô hình

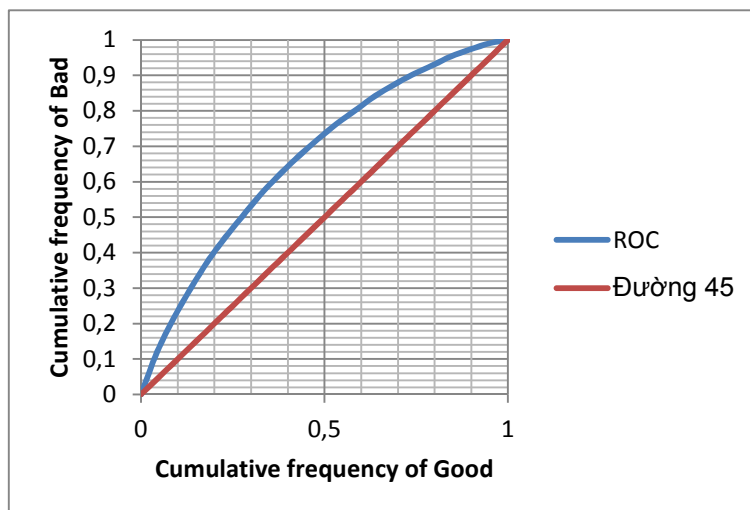
Class	Total N	Percent	Number of Good	Distr of Goods	Cumulative frequency of Good	Number of Bad	Distr of Bad	Cumulative frequency of Bad	Diện tích dưới đường cong
.	163023		148511			14512			0.667017462
...< - 814	8280	5.08%	6534	4.40%	4.40%	1746	12.03%	12.03%	0.002646717
814< - 842	8073	4.95%	6698	4.51%	8.91%	1375	9.47%	21.51%	0.007562939
842< - 860	8226	5.05%	6986	4.70%	13.61%	1240	8.54%	30.05%	0.012126358
860< - 874	8663	5.31%	7472	5.03%	18.65%	1191	8.21%	38.26%	0.017184072
874< - 884	7589	4.66%	6680	4.50%	23.14%	909	6.26%	44.52%	0.018617101
884< - 894	9022	5.53%	8021	5.40%	28.54%	1001	6.90%	51.42%	0.02590869
894< - 902	8284	5.08%	7403	4.98%	33.53%	881	6.07%	57.49%	0.027144782
902< - 909	8147	5.00%	7376	4.97%	38.50%	771	5.31%	62.80%	0.029872711
909< - 915	7542	4.63%	6881	4.63%	43.13%	661	4.55%	67.36%	0.03015398
915< - 921	8054	4.94%	7389	4.98%	48.10%	665	4.58%	71.94%	0.034653214
921< - 927	8328	5.11%	7695	5.18%	53.29%	633	4.36%	76.30%	0.038405526
927< - 933	8888	5.45%	8279	5.57%	58.86%	609	4.20%	80.50%	0.043705776
933< - 938	7873	4.83%	7334	4.94%	63.80%	539	3.71%	84.21%	0.040670305
938< - 943	7922	4.86%	7472	5.03%	68.83%	450	3.10%	87.31%	0.043149997
943< - 948	7770	4.77%	7362	4.96%	73.79%	408	2.81%	90.13%	0.043980196
948< - 954	9166	5.62%	8754	5.89%	79.68%	412	2.84%	92.96%	0.053961276
954< - 959	7351	4.51%	7008	4.72%	84.40%	343	2.36%	95.33%	0.044426119
959< - 965	7940	4.87%	7659	5.16%	89.56%	281	1.94%	97.26%	0.0496618
965< - 973	8583	5.26%	8317	5.60%	95.16%	266	1.83%	99.10%	0.054983796
973< - HIGH	7322	4.49%	7191	4.84%	100.00%	131	0.90%	100.00%	0.048202109

Nhóm tác giả tính toán

Như vậy 30% lượng khách hàng xấu tập trung ở 3 lớp khách hàng trên tổng số 20 lớp khách hàng có xác suất khách hàng tốt thấp nhất. Lớp khách hàng thứ 6 có 51.45% khách hàng xấu chứng tỏ mô hình có khả năng phân loại khách hàng tốt và xấu khá rõ nét. Tỷ lệ phần trăm lượng khách hàng xấu giảm dần theo lớp khách hàng có xác suất khách hàng tốt tăng dần. 3.77% lượng khách hàng xấu thuộc 3 lớp khách hàng có xác suất khách hàng tốt cao nhất.

5. Đánh giá chất lượng mô hình bằng hệ số GINI

Biểu đồ. Đường cong ROC dựa trên 70% mẫu xây dựng mô hình

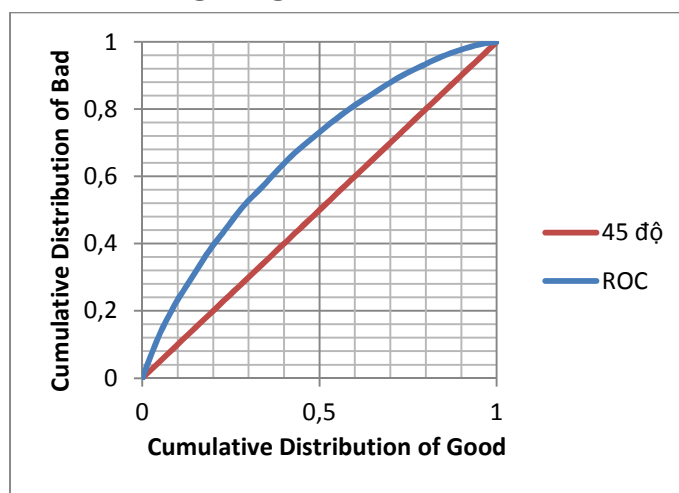


Nhóm tác giả thực hiện

$$Gini = \frac{0.667 - 0.5}{0.5} = 0.33403 \text{ cho thấy mô hình có khả năng dự báo mức trung bình.}$$

Hậu kiểm mô hình: Sử dụng 30% mẫu được giữ lại để tính toán hệ số GINI.

Biểu đồ 4: Đường cong ROC dựa trên 30% mẫu hậu kiểm



Nhóm tác giả thực hiện

$$Gini = \frac{0.6644 - 0.5}{0.5} = 0.3289 \text{ cho thấy mô hình có khả năng dự báo mức trung bình và hệ số Gini giữa mẫu xây dựng mô hình và mẫu kiểm định khá ổn định, không sai khác nhiều.}$$

6. Đánh giá kết quả ứng dụng mô hình Logistic

Ứng dụng mô hình Logistic trong XHTD và chấm điểm khách hàng là chủ đề không mới, với việc ứng dụng trên cơ sở dữ liệu của LC một lần nữa khẳng định tính khoa học và hữu dụng trong vấn đề XHTD khách hàng cho các tổ chức tín dụng. Cùng với mô hình XHTD hiện nay tại các NHTM Việt Nam, mô hình Logistic nếu được áp dụng trên cơ sở bộ dữ liệu cập nhật, đầy đủ sẽ giúp các ngân hàng đưa ra được đánh giá khách quan hơn đối với khách hàng vay vốn, giảm thiểu rủi ro tín dụng cũng như tỷ lệ nợ xấu trong các ngân hàng hiện nay.

Tài liệu tham khảo

1. Lê Thị Hạnh (2017), “Kiểm soát rủi ro tín dụng theo Basel II tại các NHTM Việt Nam”, *Tap chí Tài chính*, Truy cập ngày 16.6.2017, <http://tapchitaichinh.vn/tai-chinh-kinh-doanh/tai-chinh-doanh-nghiep/kiem-soat-rui-ro-tin-dung-theo-basel-ii-tai-cac-ngan-hang-thuong-mai-viet-nam-100966.html>
2. Trương Thị Hồng, Lê Thị Minh Ngọc (2014), “Xếp hạng tín dụng khách hàng cá nhân tại NHTM Việt Nam, thực trạng và những hạn chế cần hoàn thiện”, *Thị trường Tài chính Tiền tệ*, số 21, trang 17-21.
3. Bùi Lê Trà Linh, Hoàng Ngọc Hà, Nguyễn Văn Khoa, Phạm Thị Mai (2016), “Ứng dụng mô hình Logistic chấm điểm khách hàng nộp hồ sơ vay tại NHTM”, *Kỷ yếu Hội thi NCKH sinh viên toàn quốc “Olympic Kinh tế lượng và ứng dụng” lần 1*, trang 99-131.
4. Lê Văn Triết (2010), “Hoàn thiện hệ thống xếp hạng tín dụng cá nhân của ngân hàng TMCP Á Châu”, Luận văn thạc sĩ kinh tế, trường Đại học Kinh tế Thành phố Hồ Chí Minh, Truy cập 15.6.2017, http://dulieu.tailieuhocTap.vn/books/luan-van-de-tai/luan-van-de-tai-cao-hoc/file_goc_770495.pdf.
5. Lê Thanh Tùng (2014), “Hệ thống xếp hạng tín dụng nội bộ và các ứng dụng trong quản trị rủi ro tín dụng theo Basel II”, *Thị trường Tài chính Tiền tệ*, số 15, trang 18-21.
6. NHNN, Quyết định số 18/2007/QĐ-NHNN về việc sửa đổi, bổ sung một số điều của Quyết định số 493/2005/QĐ-NHNN.
7. NHNN, Quyết định số 493/2005/QĐ-NHNN ban hành quy định về phân loại nợ, trích lập và sử dụng dự phòng để xử lý rủi ro tín dụng trong hoạt động ngân hàng của tổ chức tín dụng.
8. NHNN, Quyết định số 57/2002/QĐ-NHNN về việc triển khai thí điểm đề án phân tích, xếp loại tín dụng doanh nghiệp.
9. NHNN, Thông tư số 02/2013/TT-NHNN quy định về phân loại tài sản có, mức trích, phương pháp trích lập dự phòng rủi ro và việc sử dụng dự phòng để xử lý rủi ro trong hoạt động của tổ chức tín dụng, chi nhánh ngân hàng nước ngoài.
10. Damodar N. Gujarati (2004), *Basic Econometrics*, McGraw-Hill, Fourth Edition.

11. Logistic Regression and Newton's Method, 36-402, *Advanced Data Analysis*, 15 March 2011, Truy cập ngày 15.8.2016, <https://www.stat.cmu.edu/~cshalizi/402/lectures/14-logistic-regression/lecture-14.pdf>.
12. Maddala (1992), *Introduction to Econometrics*, Macmillan Publishing company, New York, Second Edition.
13. Naeem Siddiqi (2006), *Credit Risk Scorecards*, John Wiley & Sons, Inc.
14. Thomas G.Tape, MD, *Interpreting Diagnostic Tests*, truy cập ngày 20.8.2016, <<http://gim.unmc.edu/dxtests/Default.htm>>