

MÔ HÌNH HỒI QUY LOGISTIC TRONG ĐO LƯỜNG XÁC SUẤT VỠ NỢ KHÁCH HÀNG TÍN DỤNG CÁ NHÂN

Hoàng Thanh Hải¹, Trần Đình Chúc²,
Nguyễn Quỳnh Hoa³

Tóm tắt

Việc đánh giá rủi ro tín dụng là không thể bỏ qua trong hoạt động tín dụng của các ngân hàng. Có nhiều nhân tố ảnh hưởng đến việc lượng hóa rủi ro tín dụng, trong đó xác suất vỡ nợ của khách hàng là yếu tố đầu tiên và rất quan trọng để ngân hàng đánh giá và ước lượng các nhân tố khác. Bài báo này ứng dụng hồi quy logistics để xây dựng một mô hình dự báo xác suất vỡ nợ của khách hàng tín dụng cá nhân và đánh giá tác động của các nhân tố đến xác suất này.

Từ khóa: Rủi ro tín dụng, xác suất vỡ nợ, hồi quy logistic.

A LOGISTIC REGRESSION MODEL FOR ESTIMATING THE PROBABILITY OF DEFAULT OF RETAIL CUSTOMERS

Abstract

In banks' credit activities, the assessment of credit risk is of paramount importance. Among a variety of factors used to quantify credit risk, the probability of default is the key one. In this paper, a logistic regression was employed to construct a model predicting the probability of default of credit card clients and evaluating regressors' influences to this probability.

Key words: credit risk, the probability of default, logistic regression.

1. Giới thiệu

Để đưa ra quyết định cho một khách hàng có được vay hay không và với mức lãi suất là bao nhiêu, các tổ chức tín dụng cần phải phân loại, xếp hạng tín dụng các khách hàng. Việc phân loại phụ thuộc vào xác suất mà khách hàng không trả được nợ đúng hạn, xác suất này gọi là xác suất vỡ nợ (probability of default). Không trả đúng hạn có thể là không trả gốc hoặc trả lãi đúng hạn hoặc cả hai [1]. Tính toán được xác suất vỡ nợ là công việc đầu tiên trong đánh giá tín dụng và xác định chính sách lãi suất. Đối với một khách hàng cá nhân, xác suất vỡ nợ chịu tác động của nhiều nhân tố như trình độ học vấn, độ tuổi, giới tính hay các nhân tố về tài chính như tình trạng trả nợ định kỳ, mức chi tiêu.

Có nhiều mô hình thống kê đã được sử dụng để ước lượng xác suất vỡ nợ như mô hình phân tích chuyên biệt, hồi quy logistic, hồi quy probit, mô hình cây phân loại, mô hình mạng nơ-ron. Mỗi mô hình có ưu và nhược điểm riêng, trong đó mô hình hồi quy logistic là mô hình được sử dụng khá phổ biến vì sự đơn giản của mô hình và độ chính xác trong phân loại cũng tương đương với các phương pháp khác [8].

Xét biến phụ thuộc nhị phân Y ($Y = 1$ nếu khách hàng vỡ nợ, $Y = 0$ nếu ngược lại) và các biến độc lập X_1, X_2, \dots, X_p . Mô hình hồi quy logistic có dạng:

$$p(\mathbf{x}) = P[Y = 1|\mathbf{x}] = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

hay dưới dạng tương đương:

$$\ln \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Các hệ số hồi quy $\beta_0, \beta_1, \dots, \beta_p$ được ước lượng bằng phương pháp hợp lý cực đại.

2. Phương pháp nghiên cứu

2.1 Nguồn số liệu và mô tả biến

Trong bài báo này, tác giả sử dụng bộ dữ liệu tín dụng của các khách hàng cá nhân tại một ngân hàng ở Đài Loan (Trung Quốc) để đánh giá ảnh hưởng của các yếu tố đến xác suất vỡ nợ của khách hàng và xây dựng một mô hình ước lượng xác suất vỡ nợ.

Nghiên cứu sử dụng dữ liệu trả nợ của 30.000 khách hàng tín dụng cá nhân tại một ngân hàng ở Đài Loan (Trung Quốc) tháng 10, 2005 [9]. Biến phụ thuộc là biến nhị phân Y ($Y = 1$ nếu khách hàng vỡ nợ, $Y = 0$ nếu ngược lại) và 23 biến giải thích bao gồm thông tin cá nhân và dữ liệu trả nợ của khách hàng:

X_1 : hạn mức tín dụng (đơn vị: Đô la Đài Loan).
 X_2 : Giới tính khách hàng (1 = Nam; 2 = Nữ).
 X_3 : Trình độ học vấn (1 = sau đại học; 2 = đại học; 3 = phổ thông; 4 = khác).
 X_4 : Tình trạng hôn nhân (1 = đã kết hôn; 2 = độc thân; 3 = khác).
 X_5 : độ tuổi (tuổi).
 $X_6 - X_{11}$: Tình trạng trả nợ hàng tháng (từ tháng 4 đến tháng 9 năm 2005): X_6 = tình trạng trả nợ tháng 9, 2005; X_7 = tình trạng trả nợ tháng 8, 2005; ...; X_{11} = tình trạng trả nợ tháng 4, 2005. Các mức bao gồm: -1 = trả nợ đúng hạn; 1 = trả nợ chậm 1 tháng; 2 = trả nợ chậm hai tháng; ...; 8 = trả nợ chậm 8 tháng; 9 = trả nợ chậm chín tháng trở lên.
 $X_{12} - X_{17}$: Lượng chi tiêu hàng tháng (đô la Đài Loan). X_{12} = lượng chi tiêu tháng 9, 2005; X_{13} = lượng chi tiêu tháng 8, 2005; ...; X_{17} = lượng chi tiêu tháng 4, 2005.
 $X_{18} - X_{23}$: Lượng trả nợ hàng tháng (đô la Đài Loan). X_{18} = lượng trả nợ tháng 9, 2005; X_{19} = lượng trả nợ tháng 8, 2005; ...; X_{23} = lượng trả nợ tháng 4, 2005.

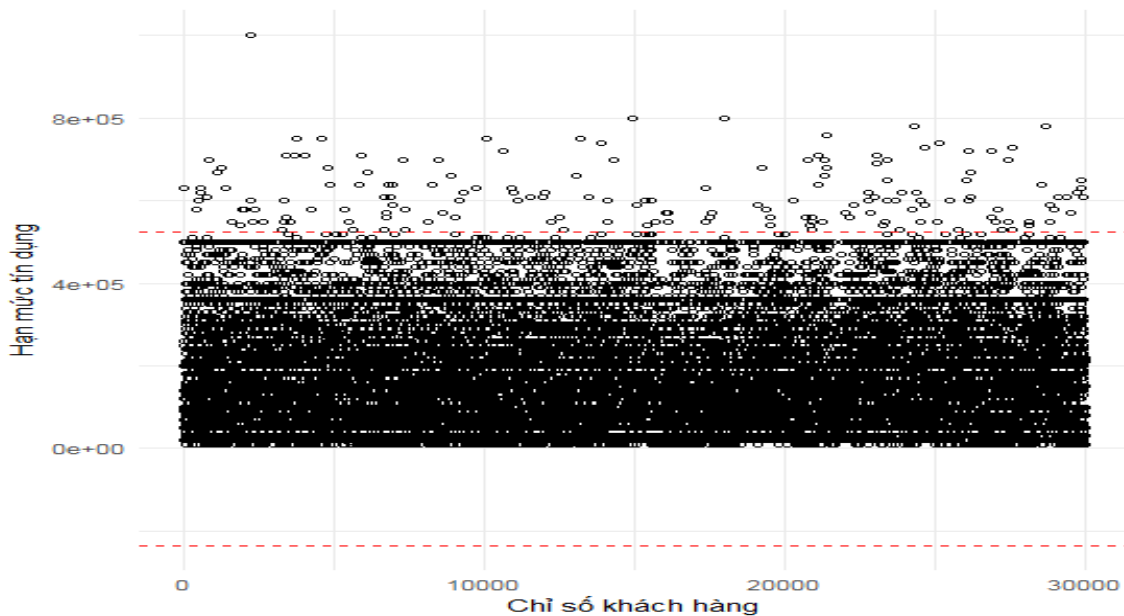
2.2 Phân tích dữ liệu khám phá

Tác giả sử dụng phần mềm R 3.3.1 để làm sạch dữ liệu và dùng các công cụ hình ảnh và kiểm định để đánh giá tổng quan về tác động của các biến độc lập đến biến phụ thuộc Y.

Sử dụng biểu đồ tần số đối với các biến tình trạng trả nợ, có hai mức không được định nghĩa là -2, 0, hơn nữa có tới 25.939 quan sát trong tổng số 30.000 (chiếm 86,5%) quan sát có hai giá trị này, vì vậy tác giả giữ nguyên các mức để phân tích thay vì gán cho chúng giá trị không xác định NA (not available). Tương tự, có một vài quan sát của biến trình độ học vấn và tình trạng hôn nhân có các mức không được định nghĩa là 0, 5 và 6. Các quan sát này được gán cho giá trị NA.

Đối với các biến liên tục, tác giả sử dụng biểu đồ tán xạ để quan sát các điểm bất thường (outliers) trong dữ liệu. Thông thường, một giá trị x được gọi là giá trị bất thường trong mẫu nếu $x > Q_3 + 1,5.IQR$ hoặc $x < Q_1 - 1,5.IQR$ trong đó Q_1, Q_3 là tứ phân vị thứ 1 và thứ 3 của mẫu, $IQR = Q_3 - Q_1$. Hình 1 là biểu đồ tán xạ của biến hạn mức tín dụng. Giá trị $Q_3 + 1,5.IQR$ trong trường hợp này là 525.000, tuy nhiên từ biểu đồ ta thấy giá trị đó hơi thấp để có thể loại quan sát. Tác giả sử dụng mức 750.000 làm căn cứ xác định điểm bất thường, có 6 quan sát thuộc loại này.

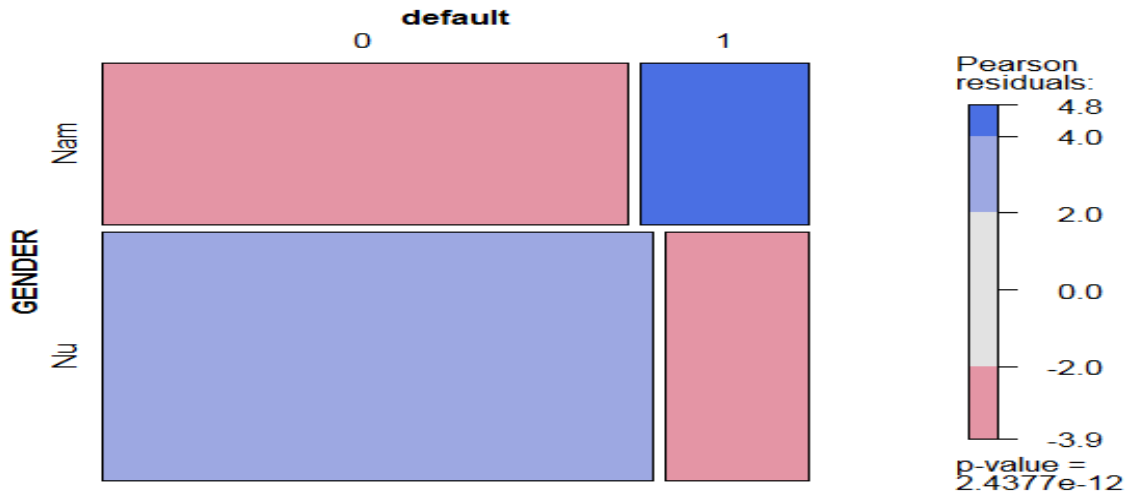
Sau quá trình xác định outliers của tất cả các biến liên tục, có 1283 chỉ số tương ứng với 457 khách hàng bị loại khỏi dữ liệu gốc ban đầu. Bởi vậy, dữ liệu dùng để xây dựng mô hình còn lại có kích thước 29.543.



Hình 1: Biểu đồ tán xạ biến hạn mức tín dụng

Để đánh giá tác động của các biến độc lập lên biến phụ thuộc, tác giả sử dụng biểu đồ

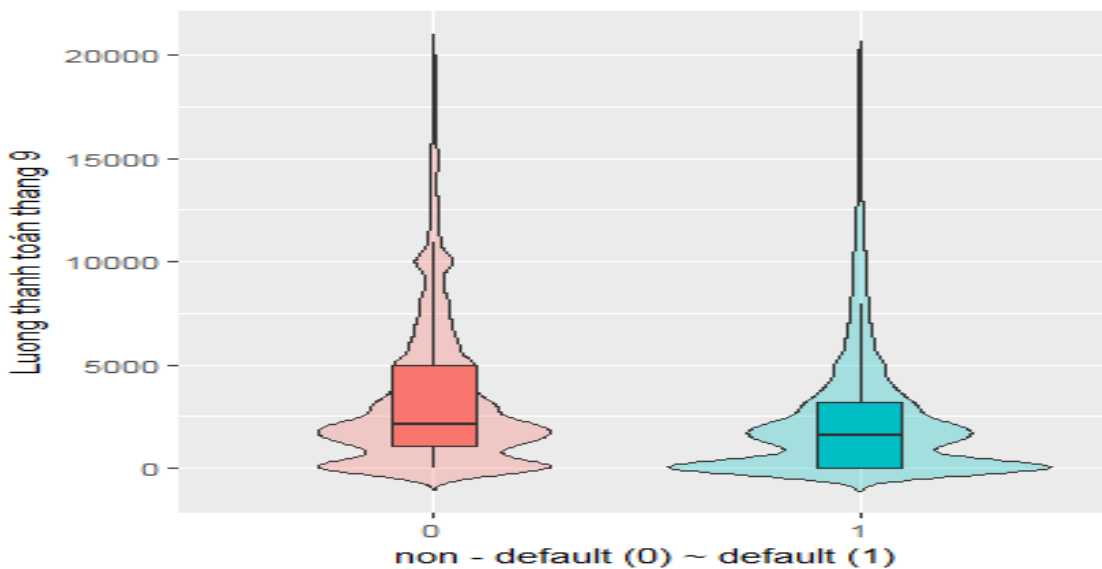
mosaic đối với biến định tính và biểu đồ violin, biểu đồ boxplot đối với biến định lượng.



Hình 2: Biểu đồ mosaic quan hệ giữa giới tính và vỡ nợ

Hình 2 là biểu đồ mosaic mô tả mối quan hệ giữa giới tính và tình trạng vỡ nợ. Quan sát biểu đồ có thể thấy khách hàng nam có xác suất vỡ nợ cao hơn khách hàng nữ. Giá trị $p \approx 2,44.10^{-12}$ của kiểm định Pearson cho thấy sự phụ thuộc giữa hai biến.

Hình 3 là biểu đồ violin của lượng thanh toán tháng 9, 2005. Từ biểu đồ có thể nhận xét nhìn chung lượng thanh toán càng thấp thì khả năng vỡ nợ càng cao.



Hình 3: Biểu đồ violin lượng thanh toán tháng 9, 2005

Bằng phương pháp tương tự đối với các biến giải thích khác, ta thu được kết quả đánh giá

khái quát tác động của các biến đến khả năng vỡ nợ như sau:

X_1 : Hạn mức tín dụng càng thấp, khả năng vỡ nợ càng cao.

X_2 : Nam có khả năng vỡ nợ cao hơn nữ

X_3 : Trình độ học vấn càng cao xác suất vỡ nợ càng thấp.

X_4 : Khách hàng đã kết hôn có xác suất vỡ nợ cao hơn.

X_5 : Nhóm có xác suất vỡ nợ lớn nhất là nhóm khách hàng dưới 25 tuổi, nhóm có xác suất vỡ nợ thấp nhất là nhóm trong độ tuổi 25 -34.

$X_6 - X_{11}$: Càng trả nợ chậm, khả năng vỡ nợ càng cao.

$X_{12} - X_{17}$: Nhìn chung không có sự khác biệt giữa nhóm vỡ nợ và nhóm không vỡ nợ về lượng chi tiêu hàng tháng.

$X_{18} - X_{23}$: Lượng trả nợ càng thấp khả năng vỡ nợ càng cao.

2.3 Tiêu chuẩn lựa chọn mô hình

Dữ liệu sau khi làm sạch, loại bỏ các giá trị NA và các điểm outliers được chia ngẫu nhiên thành hai nhóm, dữ liệu huấn luyện và dữ liệu kiểm định với tỷ lệ 50:50. Dữ liệu huấn luyện (kích thước 14.586) dùng để xây dựng mô hình

và dữ liệu kiểm định (kích thước 14.563) dùng để đánh giá mô hình.

Trên bộ dữ liệu huấn luyện, tác giả chọn biến đưa vào mô hình sử dụng tiêu chí AIC (Akaike's Information Criterion), được định nghĩa bởi:

$$AIC = \frac{-2 \ln L + 2K}{N}$$

Trong đó: L là giá trị hàm hợp lý của mô hình, K là số tham số trong mô hình ($K = p + 1$ nếu mô hình logistic có p biến giải thích) và N là kích thước của mẫu. Mô hình càng sát thực tế (L lớn) và sử dụng ít biến (K nhỏ) thì AIC càng thấp. Bởi vậy, mô hình có AIC càng nhỏ được coi là mô hình càng tối ưu.

3. Kết quả nghiên cứu

3.1 Mô hình

Sử dụng tiêu chí AIC, tác giả xác định được các biến sau đây đưa vào mô hình $X_1, X_2, X_4, X_6, X_8, X_9, X_{11}, X_{12}, X_{13}, X_{15}, X_{18}, X_{19}, X_{20}, X_{21}$.

Kết quả hồi quy trên dữ liệu huấn luyện được thể hiện trên bảng 1.

Bảng 1: Kết quả hồi quy logistic

Biến	Hệ số	Sai số chuẩn	OR	Đơn vị so sánh	Khoảng tin cậy 95% của OR	p
Hạn mức tín dụng	-0,094	0,0094	0,911	+50.000	0,894; 0,928	$< 2 * 10^{-16}$
Giới tính	-0,186	0,0359	0,830	Nam	0,774; 0,891	$2,54 * 10^{-12}$
Tình trạng hôn nhân				Đã kết hôn		
Độc thân	-0,174	0,0358	0,841		0,784; 0,902	$1,26.10^{-6}$
Khác	0,024	0,1623	1,024		0,745; 1,408	0,882
Lượng chi tiêu tháng 9	-0,176	0,0731	0,839	+50.000	0,727; 0,968	0,016
Lượng chi tiêu tháng 8	0,186	0,0869	1,204	+50.000	1,016; 1,429	0,032
Lượng chi tiêu tháng 6	0,104	0,0498	1,109	+50.000	1,006; 1,223	0,037
Lượng trả nợ tháng 9	-0,179	0,0305	0,836	+10.000	0,788; 0,888	$5,00 * 10^{-9}$
Lượng trả nợ tháng 8	-0,134	0,0282	0,875	+10.000	0,828; 0,924	$1,97 * 10^{-6}$
Lượng trả nợ tháng 7	-0,055	0,0234	0,947	+10.000	0,904; 0,991	0,019
Lượng trả nợ tháng 6	-0,051	0,0219	0,950	+10.000	0,910; 0,992	0,020

Bảng 1: Kết quả hồi quy logistic (tiếp)

Biến	Hệ số	Sai số chuẩn	OR	Đơn vị so sánh	Khoảng tin cậy 95% của OR	p
Tình trạng trả nợ tháng 9				Trả đúng hạn		
X6 = -2	- 0,403	0,0989	0,668		0,550; 0,811	$4,59 \cdot 10^{-5}$
X6 = 0	-0,637	0,0682	0,529		0,463; 0,605	$< 2 \cdot 10^{-16}$
Chậm 1 tháng	0,341	0,0668	1,406		1,234; 1,603	$3,25 \cdot 10^{-7}$
Chậm 2 tháng	1,570	0,0776	4,809		4,131; 5,599	$< 2 \cdot 10^{-16}$
Chậm 3 tháng	1,494	0,1704	4,454		3,189; 6,220	$< 2 \cdot 10^{-16}$
Chậm 4 tháng	0,892	0,2808	2,439		1,407; 4,229	0,002
Chậm 5 tháng	0,156	0,4842	1,169		0,453; 3,019	0,747
Chậm 6 tháng	0,842	0,7551	2,321		0,528; 10,196	0,265
Chậm 7 tháng	2,199	1,2813	9,020		0,732; 111,139	0,086
Chậm 8 tháng	-12,351	247,3855	0,000		0,000; $+\infty$	0,960
Tình trạng trả nợ tháng 7				Trả đúng hạn		
X8 = -2	0,063	0,1065	1,065		0,865; 1,313	0,552
X8 = 0	0,215	0,0835	1,240		1,052; 1,460	0,010
Chậm 1 tháng	-12,276	377,9862	0,000		0,000; $+\infty$	0,974
Chậm 2 tháng	0,570	0,0863	1,769		1,494; 2,095	$1,88 \cdot 10^{-11}$
Chậm 3 tháng	0,365	0,1918	1,440		0,989; 2,097	0,057
Chậm 4 tháng	0,103	0,3920	1,108		0,514; 2,389	0,794
Chậm 5 tháng	0,096	0,7570	1,101		0,250; 4,854	0,899
Chậm 6 tháng	14,183	247,3858	$1,444 \cdot 10^6$		0,000; $+\infty$	0,954
Chậm 7 tháng	-0,147	0,8638	0,863		0,159; 4,692	0,865
Chậm 8 tháng	-12,266	179,2412	0,000		0,000; $+\infty$	0,945
Tình trạng trả nợ tháng 6				Trả đúng hạn		
X9 = -2	0,104	0,1043	1,120		0,905; 1,361	0,318
X9 = 0	0,074	0,0806	1,077		0,920; 1,261	0,357
Chậm 1 tháng	1,045	655,392	2,844		0,000; $+\infty$	0,999
Chậm 2 tháng	0,459	0,0942	1,582		1,315; 1,903	$1,13 \cdot 10^{-6}$
Chậm 3 tháng	0,476	0,2350	1,609		1,015; 2,551	0,043
Chậm 4 tháng	0,307	0,4306	1,360		0,585; 1,198	0,475
Chậm 5 tháng	-1,200	0,7044	0,030		0,076; 1,200	0,088
Chậm 6 tháng	-15,808	247,3907	0,000		0,000; $+\infty$	0,949
Chậm 7 tháng	11,473	179,2366	$9,612 \cdot 10^4$		0,000; $+\infty$	0,949
Chậm 8 tháng	-13,742	343,0832	0,000		0,000; $+\infty$	0,968
Tình trạng trả nợ tháng 4				Trả đúng hạn		
X11 = -2	0,186	0,0760	1,204		1,038; 1,398	0,014
X11 = 0	-0,183	0,0658	0,833		0,732; 0,947	0,005
Chậm 2 tháng	0,293	0,0825	1,340		1,140; 1,575	0,000
Chậm 3 tháng	0,609	0,2232	1,839		1,187; 2,849	0,006
Chậm 4 tháng	0,773	0,4713	2,165		0,860; 5,453	0,101
Chậm 5 tháng	0,491	0,7671	1,634		0,363; 7,347	0,522
Chậm 6 tháng	1,863	0,8055	6,444		1,329; 3,125	0,021
Chậm 7 tháng	-10,391	179,2357	0,000		0,000; $+\infty$	0,954
Chậm 8 tháng	38,125	517,0116	$3,610 \cdot 10^{16}$		0,000; $+\infty$	0,941

Nguồn: Tính toán của tác giả trên phần mềm R

Bảng 2: Hệ số phóng đại phương sai (VIF)

Biến	VIF	Biến	VIF
X1	1,212	X12	5,297
X2	1,006	X13	6,082
X4	1,006	X15	2,970
X6	2,060	X18	1,211
X8	2,699	X19	1,122
X9	4,585	X20	1,145
X11	2,343	X21	1,049

Nguồn: Tính toán của tác giả trên phần mềm R

3.2. Kiểm định mô hình

Kiểm tra đa cộng tuyến

Hệ số phóng đại phương sai của các biến trong mô hình cho bởi bảng 2.

Do không có hệ số phóng đại phương sai nào lớn hơn 10 nên có thể cho rằng mô hình không có đa cộng tuyến.

Độ chính xác phân loại

Bảng 3 so sánh độ chính xác trong phân loại của mô hình trên hai bộ dữ liệu huấn luyện và dữ liệu kiểm định với ngưỡng xác suất (cutoff - level) phân loại giữa vỡ nợ và không vỡ nợ là 0,5.

Bảng 3: Ma trận confusion của mô hình

Dữ liệu huấn luyện		Thực tế	
		Vỡ nợ	Không vỡ nợ
Dự báo	Vỡ nợ	1.177	539
	Không vỡ nợ	2.088	10.782
Dữ liệu kiểm định		Thực tế	
		Vỡ nợ	Không vỡ nợ
Dự báo	Vỡ nợ	1.140	535
	Không vỡ nợ	2.112	10.776

Nguồn: Tính toán của tác giả trên phần mềm R

Bảng 4: Thống kê độ chính xác của mô hình

	Dữ liệu huấn luyện	Dữ liệu kiểm định
Độ chính xác	0,820	0,818
Khoảng tin cậy 95%	0,814; 0,826	0,812; 0,825
Kappa	0,376	0,367
McNemar's Test p-value	$<2,2.10^{-16}$	$<2,2.10^{-16}$
Độ nhạy (sensitivity)	0,360	0,351
Độ đặc hiệu (specificity)	0,952	0,953

Nguồn: Tính toán của tác giả trên phần mềm R

Các thống kê về độ chính xác của mô hình trên hai bộ dữ liệu cho bởi bảng 4.

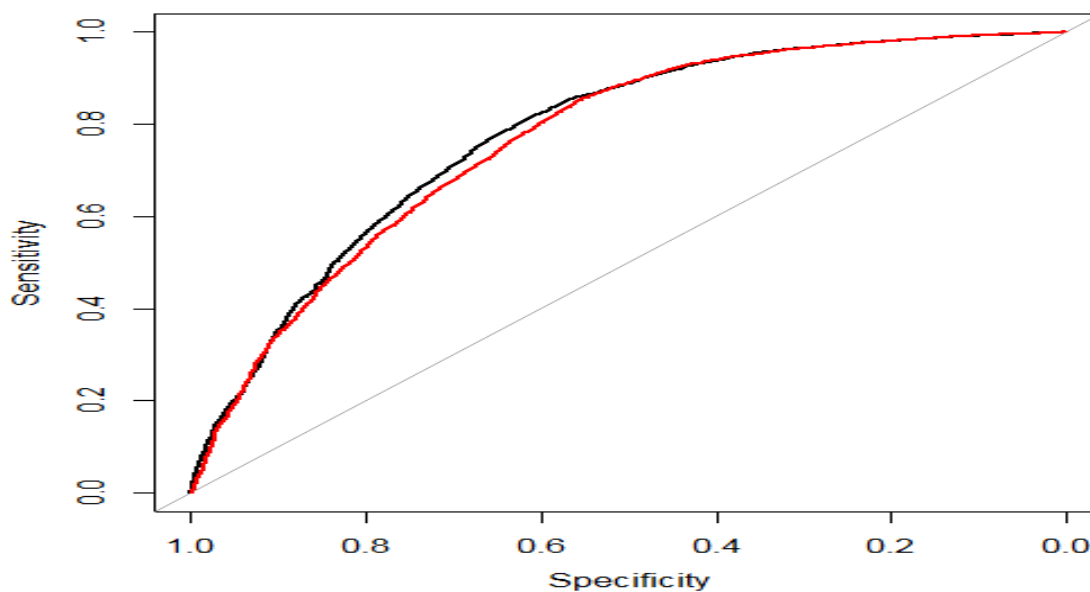
Hình 4 là đường cong ROC của hai mô hình. Từ các kết quả thống kê độ chính xác và

đường cong ROC, có thể thấy mô hình có độ chính xác trong phân loại khá ổn định trên hai mẫu dữ liệu huấn luyện và dữ liệu kiểm định.

Bảng 5: So sánh AUC của mô hình trên hai bộ dữ liệu.

	Dữ liệu huấn luyện	Dữ liệu kiểm định
AUC	0,776	0,766
Khoảng tin cậy 95%	0,766; 0,786	0,756; 0,776

Nguồn: Tính toán của tác giả trên phần mềm R



Hình 4. Đường cong ROC của mô hình trên hai bộ dữ liệu

Giá trị p của kiểm định Delong so sánh hai AUC là 0,148 cho thấy không có sự khác biệt giữa hai bộ dữ liệu về AUC.

3.3. Đánh giá ảnh hưởng của các nhân tố đến xác suất vỡ nợ

Để đánh giá ảnh hưởng của các biến giải thích trong mô hình đến xác suất vỡ nợ, tác giả sử dụng tỷ số odds (odds ratio – OR).

Mô hình hồi quy logistic có thể viết dưới dạng:

$$\ln \text{odds}(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

trong đó $\text{odds}(x) = \frac{P[Y=1|x]}{1-P[Y=1|x]}$ được gọi là odds của khả năng vỡ nợ. Để đánh giá ảnh hưởng của x_k , ta sẽ xác định odds thay đổi như thế nào khi x_k thay đổi một lượng δ , với điều kiện các biến giải thích khác giữ nguyên. Ta có

$$\text{odds}(x, x_k + \delta) = \frac{P[Y = 1|x, x_k + \delta]}{1 - P[Y = 1|x, x_k + \delta]}$$

Đại lượng

$$\text{OR} = \frac{\text{odds}(x, x_k + \delta)}{\text{odds}(x)}$$

được gọi là odds ratio. Dễ thấy, $\text{OR} = e^{\beta_k \delta}$. OR cho biết odds của khả năng vỡ nợ thay đổi một nhân tử là $e^{\beta_k \delta}$ khi x_k thay đổi một lượng là δ , các yếu tố khác không đổi. Từ kết quả hồi quy trong bảng 1 và bảng 2, ta có nhận xét:

X_1 : OR = 0,911 cho thấy khi hạn mức tín dụng tăng 50.000 đô là Đài Loan thì odds vỡ nợ giảm khoảng 10%.

X_2 : OR = 0,830 cho thấy nữ có odds vỡ nợ bằng 83% so với odds vỡ nợ của nam.

X_4 : Khách hàng độc thân có odds vỡ nợ chỉ bằng 84% odds vỡ nợ của khách hàng đã kết hôn; khách hàng có tình trạng hôn nhân “Khác” có odds vỡ nợ không khác biệt có ý nghĩa thống kê với odds vỡ nợ của khách hàng đã kết hôn.

X_{12} : Khi lượng chi tiêu trong tháng 9 tăng 50.000 thì odds vỡ nợ giảm khoảng 16%.

X_{13} : Khi lượng chi tiêu tháng 8 tăng 50.000 thì odds vỡ nợ tăng gấp 1,2 lần.

X_{15} : Khi lượng chi tiêu tháng 6 tăng 50.000 thì odds vỡ nợ tăng gấp 1,1 lần.

X_{18} : odds vỡ nợ giảm 16% khi lượng trả nợ tháng 9 tăng 10.000.

X_{19} : odds vỡ nợ giảm 12,5% khi lượng trả nợ tháng 8 tăng 10.000.

X_{20} : odds vỡ nợ giảm 5,3% khi lượng trả nợ tháng 7 tăng 10.000.

X_{21} : odds vỡ nợ giảm 5% khi lượng trả nợ tháng 6 tăng 10.000.

X_6, X_8, X_9, X_{11} : Do phần lớn OR > 1 nên có thể thấy việc trả nợ trễ làm tăng khả năng vỡ nợ của khách hàng.

Như vậy có thể thấy mô hình phản ánh khá chính xác tác động của các biến độc lập đến xác suất vỡ nợ như phần phân tích dữ liệu khám phá đã trình bày.

4. Kết luận

Bài báo ứng dụng mô hình hồi quy logistic để đánh giá tác động của các biến giải thích đến xác suất vỡ nợ của một khách hàng tín dụng cá

nhân và xây dựng một mô hình dự báo xác suất vỡ nợ. Mô hình dự báo có độ chính xác toàn cục trên 80% và AUC $\approx 0,7$ trên cả dữ liệu huấn luyện và dữ liệu kiểm định. Kết quả đánh giá tác động của các biến dựa vào OR của mô hình là phù hợp với kết quả phân tích dữ liệu khám phá sử dụng các công cụ hình ảnh.

Hạn chế của nghiên cứu là chưa đánh giá, phân tích ảnh hưởng tương tác giữa các biến, chưa so sánh được tầm quan trọng của các biến giải thích cũng như chưa so sánh được độ chính xác trong phân loại tín dụng, trong ước lượng xác suất vỡ nợ giữa mô hình hồi quy logistic và các kỹ thuật data mining khác. Những vấn đề trên sẽ được tác giả tiếp tục hoàn thiện trong những nghiên cứu sau này.

TÀI LIỆU THAM KHẢO

- [1]. Nguyễn Quang Dong, Nguyễn Thị Minh. (2012). *Giáo trình Kinh tế lượng*. Nhà xuất bản Đại học Kinh tế Quốc dân, 440 – 443.
- [2]. Harrell, F. E. (2015). *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis*, 2nd edition, Springer – Verlag, Cham.
- [3]. Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, 3rd edition, John Wiley & Sons.
- [4]. Long, J. S. (1997). *Regression models for categorical and limited dependent variables*, Sage Publications.
- [5]. Steenackers, A., Goovaerts, M. J. (1989). A credit scoring model for personal loans. *Insurance: Mathematics and Economics*, 8, 31 – 34.
- [6]. Yeh, I. -C., Lien, C. -h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert systems with applications*, 36, 2473 – 2480.
- [7]. Drugov, V. G., *Default payments of credit card clients in Taiwan from 2005*, https://rstudio-pubs-static.s3.amazonaws.com/281390_8a4ea1f1d23043479814ec4a38dbbfd9.html.
- [8]. Nguyễn Chí Dũng. (2017). *Kinh tế lượng ứng dụng với R*. <http://www.mediafire.com/file/3lg8bsfbu6csq8d/KinhTeLuongUngDungVoiR.rar>
- [9]. <https://archive.ics.uci.edu/ml/machine-learning-databases/00350/>

Thông tin tác giả:

1. Hoàng Thanh Hải

- Đơn vị công tác: Trường ĐH Kinh tế & QTKD
- Địa chỉ email: hoangthanhhai03091988@gmail.com

2. Trần Đình Chúc

- Đơn vị công tác: Trường ĐH Kinh tế & QTKD

3. Nguyễn Quỳnh Hoa

- Đơn vị công tác: Trường ĐH Kinh tế & QTKD

Ngày nhận bài: 05/11/2018

Ngày nhận bản sửa: 12/12/2018

Ngày duyệt đăng: 28/12/2018