

MÔ HÌNH HỒI QUY LOGISTICS VÀ MÔ HÌNH COX TRONG ƯỚC LƯỢNG XÁC SUẤT VỠ NỢ TÍN DỤNG, PHÂN NHÓM KHÁCH HÀNG THEO NGUY CƠ VỠ NỢ

Tô Thị Vân Anh*

Trong quá trình hoạt động, rủi ro xuất phát từ các khoản vay đang là một trong những mối quan tâm hàng đầu của ngân hàng. Dựa trên các đặc điểm của hồ sơ vay nợ, ngân hàng có thể đưa ra ước lượng cho xác suất vỡ nợ (PD) sau đó phân chia khách hàng vào các nhóm có cùng nguy cơ từ đó đưa ra quyết định cho vay nhằm hạn chế rủi ro và tối đa lợi nhuận.

• Từ khóa: mô hình hồi quy, xác suất vỡ nợ tín dụng, lợi nhuận, ngân hàng thương mại.

In the current banking context, risks from loans are one of the bank's top concerns. Probability of default has much significance as it is one of the core parts for improved allocation of capital, pricing, client judgment, regulatory compliance and, finally, monitoring of high-risk customers. Due to these significant reasons, based on the information of the loan profile, the bank can give an estimate of the probability of default (PD) based on two models is Logistic regression model and Cox regression model. The objective of these two regression methods is to estimate credit risk and extract important variables in predicting credit risk, then divide the customer into groups at the same risk from which to make a loan decision risk and maximum profitability.

• Keywords: regression model, probability of credit default, profit, commercial bank.

Ngày nhận bài: 5/8/2019

Ngày chuyển phân biên: 7/8/2019

Ngày nhận phân biên: 15/8/2019

Ngày chấp nhận đăng: 22/8/2019

I. Giới thiệu

Ngày nay các ngân hàng thương mại đóng vai trò rất quan trọng trong lĩnh vực kinh tế của nước ta dưới hình thức chuyển nhượng quyền sử dụng vốn của mình cho khách hàng cá nhân hoặc hộ gia đình, sử dụng trong một thời hạn

nhất định phải hoàn trả cả gốc và lãi với mục đích phục vụ đời sống hoặc phục vụ sản xuất kinh doanh. Cùng với đó, hệ thống ngân hàng thương mại Việt Nam cũng đã từng bước đổi mới và được coi là một trong các tổ chức tài chính quan trọng nhất của nền kinh tế với hoạt động chính là huy động vốn để sử dụng nhằm thu lợi nhuận, trong đó hoạt động tín dụng là hoạt động sinh lời lớn nhất, tuy nhiên, rủi ro là điều không thể tránh khỏi. Vì vậy, việc nhận dạng và phân nhóm khách hàng theo rủi ro tín dụng là việc làm cấp bách. Đáp ứng đòi hỏi từ thực tiễn đó, cần nghiên cứu mô hình cho phép ước lượng xác suất không trả được nợ và phương pháp phân nhóm khách hàng theo nguy cơ không trả được nợ tín dụng của khách hàng cá nhân.

Xét trong mối quan hệ tín dụng ngân hàng, “khả năng trả nợ của khách hàng” là việc đánh giá khách hàng có thực hiện đầy đủ và đúng hạn nghĩa vụ nợ cho bên cấp tín dụng trong toàn bộ thời gian quan hệ tín dụng hoặc trong một khoảng thời gian xác định hay không. Phương pháp xác định khả năng trả nợ của khách hàng thường được dựa trên một tiêu chuẩn nhất định do ngân hàng lựa chọn như dựa trên đặc điểm của khách hàng, năng lực tài chính, thiện chí trả nợ của khách hàng khi chưa phát sinh nghĩa vụ nợ và dựa trên đặc điểm của khoản nợ như lịch sử thanh toán nợ, tình trạng trả nợ thực tế của khách hàng. Bài toán tập trung vào dịch vụ cho

* *Đại học Khoa học Tự nhiên, ĐHQG Hà Nội*

vay tín dụng. Mỗi một khách hàng khi có nhu cầu giao dịch sẽ được yêu cầu cung cấp thông tin khách hàng, các thông tin đó có thể là dữ liệu cá nhân như giới tính, tuổi tác, nghề nghiệp,... thông tin lịch sử vay tín dụng trong quá khứ như thời hạn vay. Dữ liệu hành vi như lịch sử sử dụng khoản vay trên các sản phẩm. Thông tin này biểu diễn dưới dạng vector $X = (X_1, \dots, X_m)$.

1) Ước lượng PD hiện tại dựa trên hai mô hình là mô hình hồi quy Logistic và mô hình hồi quy Cox. Mục tiêu của hai phương pháp hồi quy này là ước tính rủi ro tín dụng và trích xuất các biến quan trọng trong dự đoán rủi ro tín dụng.

2) Phân nhóm khách hàng theo nguy cơ không trả được nợ, các khách hàng trong cùng một nhóm sẽ có nguy cơ rủi ro như nhau.

3) Phân nhóm khách hàng theo các chỉ tiêu phân nhóm, biểu diễn dưới dạng cây quyết định.

II. Mô hình logistic trong ước tính xác suất vỡ nợ

Mô hình hồi quy logistics xem xét mối liên hệ giữa biến phụ thuộc (Y) và tất cả các biến còn lại là biến độc lập (X), thể hiện các nhân tố thông tin của khách hàng. Biến Y là biến nhị phân chỉ nhận hai giá trị 0 hoặc 1. Cụ thể:

$$Y = \begin{cases} 1 & \text{Nếu sự kiện vỡ nợ xảy ra} \\ 0 & \text{Nếu sự kiện vỡ nợ không xảy ra} \end{cases} \quad (1)$$

Giả sử π_i là xác suất trả được nợ của khách hàng thứ i , ta có mô hình ước lượng cho như sau:

$$\begin{aligned} \pi_i(x) &= P(Y = 0 | X = x) \\ &= \frac{e^{\beta_0 + \beta^T x}}{1 + e^{\beta_0 + \beta^T x}} = \frac{1}{1 + e^{-s(x)}} \end{aligned} \quad (2)$$

Hàm $s(x) = \beta_0 + \beta^T x$ được gọi là điểm log-odds

$$s(x) = \log\left(\frac{\pi_i(x)}{1 - \pi_i(x)}\right)$$

Nếu một khách hàng có m thông tin được mô tả bằng biến $X = (X_1, X_2, \dots, X_m)$ và giá trị cụ thể là $x = (x_1, \dots, x_m)$ có β_j là hệ số của thông tin x_j thì $s(x)$ là điểm tín dụng của khách hàng x . Tham số được ước lượng trong mô hình hồi quy logistic là β_0 và $\beta = (\beta_1, \beta_2, \dots, \beta_m)$, ký hiệu tương ứng là $\hat{\beta}_0$ và $\hat{\beta}$. Tính toán ước lượng dựa trên phương pháp MLE.

III. Quyết định điểm cắt phân lớp khách hàng

Khi một người nộp đơn vay tín dụng, họ sẽ cung cấp đầy đủ thông tin để người cho vay xây dựng hồ sơ. Sau đó với mô hình ước lượng, điểm số $s(x) = \beta_0 + \beta^T x$ sẽ được gán cho mỗi cá nhân. Điểm số cao thì ít rủi ro hơn nên những hồ sơ có điểm số cao sẽ được chấp nhận. Do đó, điểm cắt được đưa ra.

Nếu $s(x) \leq c$ thì loại bỏ hồ sơ tín dụng. Nếu $s(x) \geq c$ thì chấp nhận hồ sơ tín dụng. Xác suất cắt cho bởi

$$1 - \text{logit}^{-1}(c) = 1 - \frac{1}{1 + e^{-c}}$$

Quyết định điểm cắt sẽ dựa trên một số yếu tố liên quan đến mục tiêu của người cho vay như tối đa hóa lợi nhuận dự kiến. Nếu quyết định cho vay và người vay trả được nợ thì sẽ thu được một khoản lợi nhuận là g và nếu người vay không trả được thì người cho vay sẽ bị thua lỗ một khoản l . Khi đó $1 - F_0(c) = P(S \leq c | Y = 0)$ được gọi là phân lớp đúng của trường hợp trả được nợ. $F_1(c) = P(S \leq c | Y = 1)$ được gọi là phân lớp sai của trường hợp không trả được nợ, với

$$\gamma = \frac{l}{g}$$

và đặt chi phí là:

$$e_\gamma(c) = F_0(c)(1 - p_1) + \gamma(1 - F_1(c))p_1 \quad (3)$$

Điểm cắt được tìm là điểm cắt tối ưu theo chi phí

$$c = \log\left[\frac{\gamma/(1 + \gamma)}{1 - \gamma/(1 + \gamma)}\right] = \log \gamma \quad (4)$$

IV. Mô hình Cox trong ước tính xác suất vỡ nợ

Một mô hình khác được dùng là mô hình phân tích sống sót để tính điểm tín dụng, biến quan tâm là thời gian xảy ra sự kiện. Mô hình này ước lượng được xác suất sống sót trên toàn bộ tập dữ liệu. Ưu điểm của mô hình phân tích sống sót là có thể kết hợp với dữ liệu kiểm duyệt.

A. Hàm sống sót

Hàm sống sót được định nghĩa bởi $s(x)$ là xác suất mà một cá thể sống sót vượt quá thời gian t .

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx \quad (5)$$

B. Hàm rủi ro (Hazard function)

Do khả năng thất bại tại thời điểm t biết rằng đối tượng đã sống sót qua một số thời điểm t:

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = f(t)/S(t) \quad (6)$$

Δt là khoảng thời gian rất nhỏ và là hàm mật độ xác suất của t.

Mối quan hệ giữa hàm sống sót và hàm rủi ro cho bởi:

$$h(t) = - \left[\frac{dS(t)/dt}{S(t)} \right] S(t)$$

$$\Rightarrow S(t) = \exp \left[- \int_0^t h(u)du \right] \quad (7)$$

Với dữ liệu tín dụng trong mô hình Cox, xác suất vỡ nợ (PD) của khách hàng thứ i tại thời điểm t cho bởi:

$$\pi_i(X_i, t) = 1 - S(X_i, t)$$

$$= 1 - \exp \left[- \int_0^t h(X_i, u) du \right] \quad (8)$$

V. Cây quyết định

Cây quyết định là một kiểu mô hình dự báo (predictive model), nghĩa là một ánh xạ từ các quan sát về một sự vật/hiện tượng tới các kết luận về giá trị mục tiêu của sự vật/hiện tượng. Dữ liệu được cho dạng $(x, y) = (x_1, x_2, x_3, \dots, x_n, y)$. Biến phụ thuộc (dependant variable) là biến mà chúng ta cần tìm hiểu, biến phân loại là x_1, x_2, \dots là các biến sẽ giúp ta thực hiện công việc đó.

Giả sử S là tập dữ liệu hiện tại gồm s mẫu dữ liệu, tập nhãn lớp có m giá trị $\{C_1, \dots, C_m\}$, S_i là số lượng mẫu của S_i trong lớp C_i . Để phân loại một mẫu ta sử dụng khái niệm entropy.

$$Entropy(S) = - \sum_{i=1}^m p_i \log_2(p_i)$$

với $p_i = \frac{S_i}{S}$

là xác suất mẫu thuộc lớp C_i .

Thuộc tính V gồm n giá trị $\{v_1, \dots, v_n\}$ được sử

dụng để chia S thành nhiều tập con S_1, \dots, S_n . S_i là các mẫu trong S có giá trị thuộc tính V là v_i .

Information Gain (IG) đánh giá khả năng của một thuộc tính khi được dùng để phân lớp các mẫu dựa vào số entropy. IG cho biết mức độ giảm của entropy khi phân nhánh mẫu.

$$Gain(S, V) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (9)$$

Thuộc tính làm Gain (S, V) lớn nhất sẽ được chọn để phân mảnh.

Khi có một số lượng lớn các giá trị, ta sử dụng Gain Ratio tính thông qua thông tin tiềm năng (potential information) của mỗi phân hoạch

$$P(S, V) = - \sum_{i=1}^m \frac{|S_i|}{S} \log_2 \left(\frac{|S_i|}{S} \right) \quad (10)$$

Thuộc tính được chọn nếu tỉ số

$$Gainratio = \frac{Gain(S, V)}{P(S, V)}$$

đạt giá trị lớn nhất.

Khách hàng sẽ được phân nhóm theo các quy luật phân nhóm được biểu diễn dưới dạng cây quyết định (cây quyết định) với các nút mô tả chỉ tiêu và các nhánh mô tả kết quả của chỉ tiêu. Kết quả cuối cùng của việc phân nhánh cây là việc tạo ra các nhóm có nguy cơ vỡ nợ khác nhau có ý nghĩa.

VI. Kết quả thực nghiệm

Dữ liệu đầu vào bài toán là dữ liệu tín dụng trong hai năm 2015 - 2016. Bộ dữ liệu chuẩn của mô hình Logistic gồm 21 biến, bao gồm: Giới tính, nghề nghiệp, thu nhập, tổng hạn mức vay, tổng giá trị vay, chiều dài thời gian vay, số lần vay, số lần trễ hạn, tổng giá trị do trễ hạn, tổng giá trị đã vay, tổng giá trị ứng trước, thời gian trả nợ trung bình, số tiền trả nợ trung bình, số tiền đã tiêu, tổng số giao dịch, tổng thời gian trễ hạn, trung bình tháng, số dư lần cuối, tỷ lệ số dư cuối/ số dư đầu, trạng thái vỡ nợ (Y=1 nếu vỡ nợ, Y=0 nếu không vỡ nợ).

Biến thu nhập được chia làm hai nhóm: Khách hàng có thông tin và không có thông tin

về thu nhập, kết quả như sau: Trong số 385.013 khách hàng có 374.574 khách hàng không có thông tin thu nhập với số vỡ nợ là 9.050 (chiếm tỷ lệ 0.024), 10.439 khách hàng có thông tin thu nhập với số vỡ nợ là 87 (chiếm tỷ lệ 0.008).

Phân tích trước tiên dựa trên nhóm khách hàng không có thông tin thu nhập, sau đó sẽ so sánh với nhóm khách hàng có thông tin thu nhập. Cách phân chia thực tế do thu nhập ảnh hưởng trực tiếp tới việc xác định tổng hạn mức vay và giá trị vay.

Dựa trên các thông tin khách hàng có 2 biến liên quan tới điều kiện vay là: tổng hạn mức vay và tổng giá trị vay.

Chia biến tổng giá trị vay thành 5 nhóm miền giá trị là <276M, <800M, <2B, <4.7B <1073B. Chia biến tổng hạn mức vay thành 5 nhóm có miền giá trị sau: <366M, <2.67B, <5.8B, <28.5B, >28.5B.

Có tất cả 21 nhóm biến, tính PD bằng phương pháp hồi quy Logistic cho các nhóm này, các giá trị PD gần nhau được gộp lại, kết quả còn lại 11 nhóm.

Xử lý với nhóm khách hàng có thông tin thu nhập: Tính giá trị thu nhập trung bình chia nhóm khách hàng vào 11 nhóm có tương đồng thu nhập trung bình với các khoảng tổng giá trị khoản vay và tổng hạn mức, sau đó gộp toàn bộ dữ liệu của khách hàng theo 11 nhóm đã biết.

Kết quả có 11 nhóm được chia như sau:

Nhóm	Tổng giá trị vay	Tổng hạn mức vay
N1	< 276M	< 366M
N2	< 276M	< 2.67B
N3	< 276M	< 5.8B
N4	< 276M	< 28.5B
N5	< 276M	> 28.5B
N6	< 800M	< 2.67B
N7	< 800M	> 2.67B
N8	< 2B	> 366M
N9	< 4.7B	< 2.67B
N10	< 4.7B	> 2.67B
N11	> 4.7B	> 2.67B

Bảng 1: Kết quả AUC và PD cho 11 nhóm

	AUC	Accuracy	Sensitivity	Specificity	Cutoff	PD
N1	0.98	0.92	0.94	0.92	0.023	0.03
N2	0.98	0.96	0.95	0.96	0.013	0.0307
N3	0.99	0.96	0.96	0.96	0.056	0.0771
N4	0.92	0.88	0.88	0.88	0.002	0.052
N5	NUL	NUL	NUL	NUL	NUL	NUL
N6	0.99	0.95	0.96	0.95	0.005	0.0058
N7	0.94	0.96	0.94	0.96	0.007	0.0148
N8	0.97	0.97	0.97	0.97	0.011	0.0103
N9	NUL	NUL	NUL	NUL	NUL	NUL
N10	0.99	0.97	1	0.97	0.017	0.0123
N11	0.96	0.98	0.9	0.98	0.078	0.0215

Kết quả chạy riêng cho mô hình nhóm N4
Tập dữ liệu training gồm 1658 khách hàng

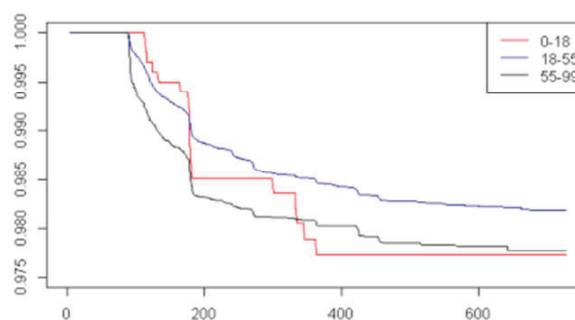
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.6058	1.0730	-7.088	1.36e-12 ***
HanMucVay	1.5925	0.4727	3.369	0.000754 ***
SoLanDaVay	-2.0506	0.6389	-3.209	0.001330 **
TongGiaTriungTruoc	-2.0953	0.3579	-5.854	4.80e-09 ***
TongSoTienDaTieu	6.3851	2.0633	3.095	0.001971 **
TongSoGiaoDich	-5.4980	1.9112	-2.877	0.004019 **
SoduLanCuoi	-2.7063	0.4723	-5.730	1.00e-08 ***

Kết luận mô hình:

$$s(x) = -7.6058 + 1.5925 \times \text{hanmucvay} - 2.0506 \times \text{solandavay} - 2.0953 \times \text{Tonggiatriungtruoc} + 6.3851 \times \text{tongsotientatieu} - 5.4980 \times \text{tongsotgiaodich} - 2.7063 \times \text{sodulancuoi}$$

Hình 1. Đường sống sót cho 3 nhóm tuổi (0,18], (18,55] và (55,99]



Kết quả cho mô hình Cox: Dữ liệu tín dụng theo dõi trong 2 năm: 01/01/2015- 31/12/2016.

Các biến gồm có: Tuổi (tuổi khách hàng), giới tính, giá trị vay, T_1 (khoảng thời gian từ lúc mở hồ sơ vay đến khi trễ hạn lần 1), Y_1 (nếu xảy ra trễ lần 1 thì $Y_1 = 1$, nếu không thì $Y_1 = 0$), T_2 là thời gian xảy ra sự kiện trễ hạn lần 2, Y_2 nhận giá trị 0 hoặc 1 như với Y_1 , T (khoảng thời gian từ lúc mở hồ sơ vay đến khi vỡ nợ), Y ($Y = 1$ nếu có vỡ nợ, $Y = 0$ nếu không có vỡ nợ).

Thống kê dữ liệu gồm 1.099.552 hồ sơ, số hồ sơ vỡ nợ 6917, số hồ sơ trễ hạn lần 1 là 407.248, số hồ sơ trễ hạn lần 2 là 189.335 hồ sơ.

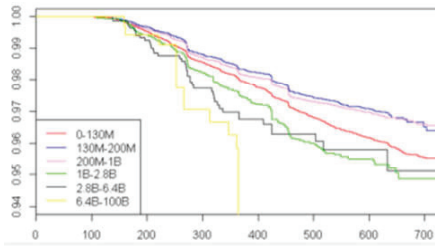
Ta chia làm biến giá trị vay thành 6 nhóm: 0 - 130M, 130M - 200M, 200M - 1B, 1B - 2.8B, 2.8B - 6.4B, 6.4 - 100B.

Với thống kê dữ liệu đã làm sạch, chia nhóm theo hai biến tuổi và giá trị vay, có tất cả 18 nhóm chia. Gộp các nhóm có chung đường sống sót ta thu được 8 nhóm với 8 đường sống sót phân biệt như sau:

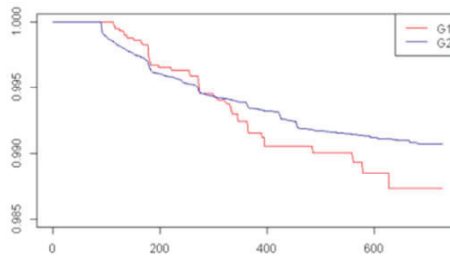
- G1 Tuổi 0 - 18
- G2 Tuổi 55 - 99
- G3 Tuổi 18 - 55, vay 0 - 130 M
- G4 Tuổi 18 - 55, vay 130 M - 200 M
- G5 Tuổi 18 - 55, vay 200 M - 1B
- G6 Tuổi 18 - 55, vay 1B - 2.8B
- G7 Vay 2.8B - 6.4B
- G8 Vay 6.4B - 100B

Đường sống sót 8 nhóm được thể hiện ở hình 3, 4, 5. Kết quả kiểm định cho thấy

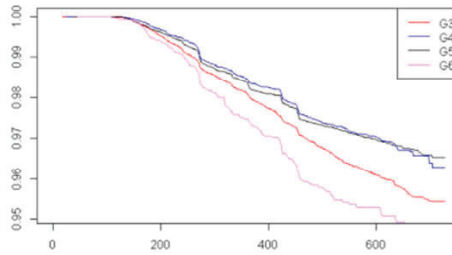
Hình 2. Đường sống sót 6 nhóm vay theo ngày



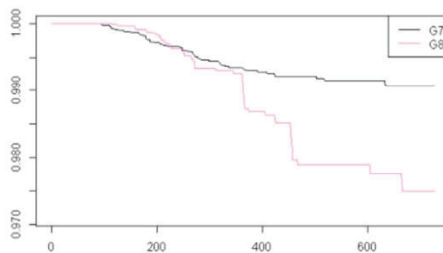
Hình 3. Đường sống sót cho nhóm G1, G2



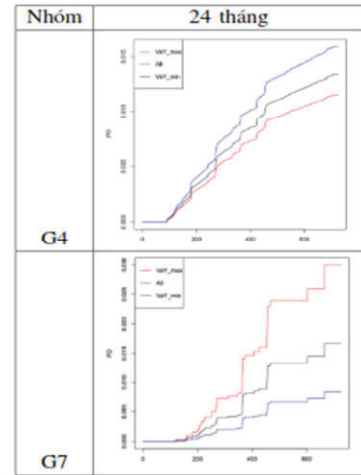
Hình 4. Đường sống sót của 4 nhóm G3, G4, G5, G6



Hình 5. Đường sống sót của 2 nhóm G7, G8



Hình 6. Đồ thị xác suất vỡ nợ của hai hồ sơ vay có giá trị vay tối đa và tối thiểu trong 2 nhóm G4, G7, thời hạn vay 24 tháng



chỉ có nhóm G4 và G7 là có ý nghĩa, phù hợp với mô hình Cox. Mô hình hồi quy Logistic và mô hình Cox đều tính được xác suất vỡ nợ và phân loại các khách hàng thành các nhóm có cùng mức nguy cơ rủi ro. Hình 7 cho thấy với nhóm G4 (tuổi 18-55, vay 130M-200M), khoản vay tối thiểu có xác suất vỡ nợ cao hơn khoản vay tối đa. Trong nhóm G7 (vay 2.8B - 6.4B), khoản vay tối đa có xác suất vỡ nợ cao hơn khoản vay tối thiểu. Kết quả được coi là phù hợp vì trong nhóm G7 khoản vay lớn thì khả năng vỡ nợ là cao hơn còn trong nhóm

G4 khoản vay nhỏ thuộc về các đối tượng nhỏ lẻ có xác suất vỡ nợ cao hơn khoản vay vừa phải.

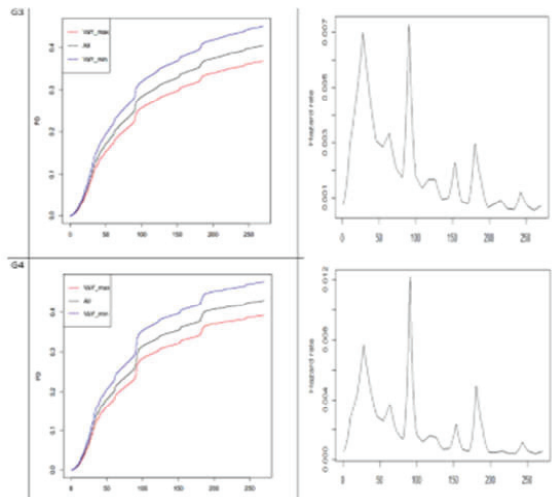
Cây quyết định phân nhóm khách hàng theo các chỉ tiêu phân nhóm.

Dữ liệu được sử dụng là dữ liệu Logistic của 6 nhóm vay (hạn mức vay < 2.67B và giá trị vay từ 276M - 800M) với 76.365 hồ sơ chia làm 2 tập: train set và test set. Các biến quan sát bao gồm: Tổng thời gian trễ hạn (ngày), tổng số tiền đã tiêu (VNĐ),

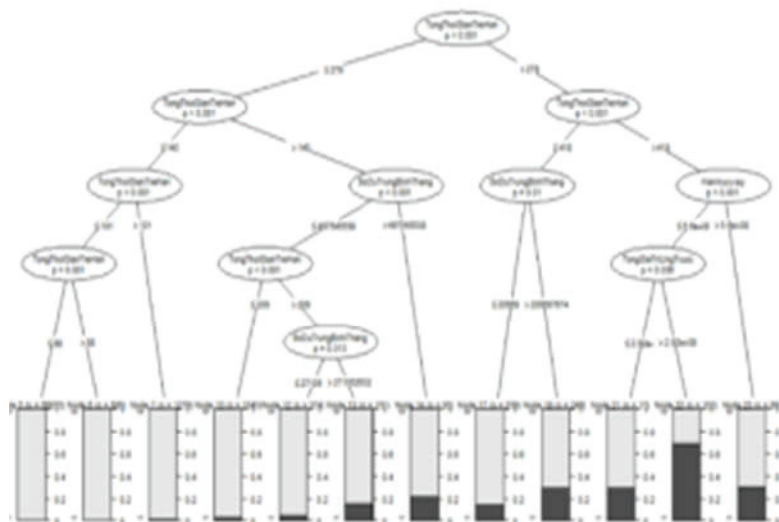
Bảng 2: Kết quả nhóm G4, G7 với sự kiện Y trong 24, 12 và 9 tháng

		24m	12m	9m
G4	Tuoi.pca			
	GTVpca	-0.11	-0.13	-0.15
	GioiTinh.pca			
	Kiểm định PH	0.0391	0.0478	0.1045
	C index	0.58	0.61	0.63
G7	Tuoi.pca	-0.41	-0.42	-0.38
	GTVpca			
	GioiTinh.pca	0.26	0.23	0.24
	Kiểm định PH	0.03188	0.02014	0.3364
	C index	0.59	0.63	0.65

Hình 7. Đồ thị xác suất vỡ nợ và đường rủi ro cơ sở của hai hồ sơ vay có giá trị vay tối đa và tối thiểu trong 2 nhóm G3, G4, thời hạn vay 12 tháng



Hình 8. Kết quả Conditional inference Trees



hạn mức vay (VNĐ), tổng giá trị ứng trước (VNĐ), số dư lần cuối (VNĐ), giá trị khoản vay (VNĐ), số dư trung bình tháng (VNĐ), giới tính (Nam, nữ).

Nguy cơ vỡ nợ thấp với nhóm khách hàng có tổng thời gian trễ hạn dưới 3 tháng và nguy cơ vỡ nợ cao với nhóm khách hàng có tổng thời gian trễ hạn cao và hạn mức vay nhỏ hơn 600 triệu và tổng giá trị ứng trước cao hơn 300 triệu.

VII. Kết luận

Hai mô hình hồi quy Logistic và mô hình Cox ước lượng được xác suất vỡ nợ của từng khách hàng và của từng nhóm khách hàng, xác suất vỡ nợ của hồ sơ sẽ thay đổi theo thời gian kể từ thời điểm bắt đầu vay. Mô hình Cox với giả thuyết mỗi nhóm có một nguy cơ riêng là phù hợp để tính xác suất và thời gian trả nợ trễ hạn lần 1. Trong lần trễ hạn đầu tiên, hồ sơ có giá trị vay lớn khả năng trễ hạn lần 1 lại thấp hơn hồ sơ có giá trị vay nhỏ.

Hình 9 cho thấy cây quyết định phân nhóm khách hàng thành 12 nhóm. Trong đó chỉ tiêu phân nhóm giúp phân biệt khách hàng có nguy cơ vỡ nợ thấp và cao là *TongThoiGianTreHan*. Thấy được mối quan hệ chặt chẽ giữa các chỉ tiêu, phân nhóm bằng cây quyết định sử dụng suy diễn cho kết quả tin cậy. Mười hai nhóm khách hàng có đặc điểm phân biệt rõ ràng và có nguy cơ vỡ nợ khác biệt nhau.

Tài liệu tham khảo:

Paul Murrell, "R Graphics", NXB Chapman & Hall/CRC. (2005)

Yanagimoto, Kamakura, "The maximum full and partial like-likelihood estimators in the proportional hazard model", *Annals of the Institute of Statistical Mathematics*, 36, tr.363-373.

COX, D. R, "Regression models and life tables (with discussion)", *Journal of Royal Statistical Society: B*, 34, tr.187-220.

COX, D. R, *Partial likelihood*, *Biometrika*, 62, tr.269-276.

Kaplan, E. L. and Meier, P., *Nonparametric estimation from incomplete observations*, *Journal of American Statistical Association*, 53, tr.457-481 (1958).

Ross Gayler, "Credit Scoring using R. R Pubs".

G. Rodriguez, Revised, "Logit Models for Binary Data".

Badr Missaoui, "Statistics in Retail Finance", Room 545, Huxley Building.

Edward S. Venter, "Probability of Default Calibration for Low Default Portfolios: Revisiting the Bayesian Approach". Master thesis of Commerce in Financial Risk Management, Faculty of Economics and Business Sciences, Stellenbosch University, South Africa (2016).