

TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN  
KHOA TOÁN KINH TẾ



## CHUYÊN ĐỀ THỰC TẬP

<b>Chuyên ngành:</b>	Toán Kinh tế
<b>Đề tài:</b>	Ứng dụng, đánh giá, và so sánh một số mô hình phân loại vào việc phân loại khách hàng thẻ tín dụng
<b>Sinh viên thực hiện:</b>	Nguyễn Đức Hiếu
<b>Mã sinh viên:</b>	11131371
<b>Lớp:</b>	Toán Kinh tế 55
<b>Giảng viên hướng dẫn:</b>	PGS. Nguyễn Thị Minh

---

Hà Nội, Ngày 24 tháng 3 năm 2017

## LỜI MỞ ĐẦU

Đối với các ngân hàng việc chấm điểm tín dụng và phân loại các khách hàng là yếu tố thiết yếu cho lợi nhuận của ngân hàng. Phương pháp truyền thống của việc ra quyết định có cho một cá nhân cụ thể vay hay không là dựa trên đánh giá cảm tính dựa trên kinh nghiệm cá nhân. Tuy nhiên, sự phát triển về quy mô của nền kinh tế đã tạo ra sức ép về nhu cầu vay, đi kèm với đó là sự cạnh tranh giữa các ngân hàng và công nghệ máy tính ngày càng phát triển đã khiến cho việc sử dụng các mô hình thống kê trong việc phân loại các khách hàng tín dụng là bắt buộc đối với các ngân hàng trên thế giới mà ở Việt Nam cũng không phải là ngoại lệ.

Vậy, phương pháp ước lượng nào có thể giúp chúng ta xây dựng được hệ thống chấm điểm tín dụng chính xác nhất? Đã có một số nghiên cứu mang tính chất so sánh hiệu năng giữa các mô hình (Baesens et al., 2003; Xiao et al., 2006; Lessmann et al., 2015). Sự khác biệt về hiệu năng của các phương pháp khác nhau là có, tuy nhiên hầu như là không đáng kể, và không phải các mô hình hiệu quả hơn đều là các mô hình mới và tân tiến. Theo Thomas (2010), cách hiệu quả để xây dựng một hệ thống lượng định hiệu quả là phối hợp nhiều mô hình khác nhau thay vì tìm kiếm một mô hình toàn diện có thể áp dụng với tất cả các ngân hàng.

Trong bài này, chúng ta sẽ tiếp cận đến một số phương pháp phân loại các khách hàng tín dụng phổ biến hiện nay và rút ra một số kết luận về việc sử dụng các phương pháp khác nhau sao cho hợp lý. Bài viết này được bố cục như sau:

- **Chương 1** đưa ra một cái nhìn tổng quan về lĩnh vực quản trị rủi ro tín dụng trong ngân hàng và đưa ra một số vấn đề của việc chấm điểm tín dụng tại các ngân hàng Việt Nam.
- Các mô hình được thực hiện trong bài này sẽ được giới thiệu ở **Chương 2**, đi kèm với đó là một số chỉ tiêu sẽ được dùng để đánh giá mô hình trong bài này.

- Trong **Chương 3**, chúng ta sẽ ứng dụng các phương pháp được giới thiệu ở **Chương 2** trong một bộ số liệu mẫu về các khách hàng thẻ tín dụng trong một ngân hàng ở Đài Loan.
- Kết quả của các mô hình sẽ được thảo luận ở **Chương 4**, cùng với một số kết luận rút ra được sau khi áp dụng mô hình.

Đề tài này được soạn thảo bằng  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  kết hợp với Sweave và knitr (Xie, 2015). Tất cả phân tích được thực hiện trên phần mềm thống kê R version 3.3.3 (2017-03-06) (R Core Team, 2017), các phân tích cụ thể được thực hiện sử dụng các gói mở rộng ggplot2 (Wickham, 2009)... Trong quá trình thực hiện các mô hình có bao gồm các lệnh tạo giá trị ngẫu nhiên, để đảm bảo khả năng tái lập các kết quả trong mô hình, ta gieo seed cho R:

```
set.seed(0)
```

Em xin cảm ơn giáo viên hướng dẫn, cô Nguyễn Thị Minh, trưởng khoa Toán Ứng dụng trong Kinh tế, cùng với các thầy cô giáo khác trong khoa đã tạo điều kiện cho em thực hiện đề tài này.

## MỤC LỤC

Lời mở đầu . . . . .	1
Mục lục . . . . .	3
Danh sách bảng . . . . .	4
Danh sách hình . . . . .	5
<b>1 Tổng quan về quản trị rủi ro tín dụng đối với khách hàng cá nhân</b>	<b>6</b>
1.1 Một số khái niệm . . . . .	6
1.2 Thực trạng của việc chấm điểm tín dụng tại Việt Nam . . . . .	6
1.3 Kết luận . . . . .	6
<b>2 Các phương pháp phân loại khách hàng vay thẻ tín dụng</b>	<b>7</b>
2.1 Các mô hình phân loại . . . . .	7
2.1.1 Mô hình logistic . . . . .	7
2.1.2 Mô hình phân loại tuyến tính . . . . .	7
2.1.3 Mô hình SVM (Support Vector Machine) . . . . .	7
2.2 Đánh giá mô hình . . . . .	7
2.2.1 Đường ROC và phần diện tích dưới đường cong (AUC) . . . . .	7
2.2.2 Thang đo H . . . . .	7
<b>3 Tình huống nghiên cứu</b>	<b>8</b>
3.1 Số liệu và các biến số . . . . .	8
3.2 Ứng dụng mô hình logit . . . . .	9
3.3 Ứng dụng mô hình phân loại tuyến tính . . . . .	9
3.4 Ứng dụng mô hình SVM . . . . .	9
<b>4 Kết luận</b>	<b>11</b>
Tài liệu tham khảo . . . . .	12

## **DANH SÁCH BẢNG**

## DANH SÁCH HÌNH VẼ

3.1	Ma trận hệ số tương quan giữa các biến trong bộ số liệu. . . . .	10
-----	--	----

## **CHƯƠNG 1**

# **TỔNG QUAN VỀ QUẢN TRỊ RỦI RO TÍN DỤNG ĐỐI VỚI KHÁCH HÀNG CÁ NHÂN**

### **1.1 MỘT SỐ KHÁI NIỆM**

### **1.2 THỰC TRẠNG CỦA VIỆC CHẤM ĐIỂM TÍN DỤNG TẠI VIỆT NAM**

### **1.3 KẾT LUẬN**

## CHƯƠNG 2

# CÁC PHƯƠNG PHÁP PHÂN LOẠI KHÁCH HÀNG VAY THẺ TÍN DỤNG

### 2.1 CÁC MÔ HÌNH PHÂN LOẠI

#### 2.1.1 Mô hình logistic

Mô hình hồi quy Logistic được dùng để nghiên cứu mối quan hệ giữa xác suất của các biến nhị phân hoặc phân loại và các biến giải thích khác. Hướng tiếp cận của mô hình Logistic cho bài toán phân loại là bằng cách ước lượng giá trị xác suất  $P(y = 1|X)$  như sau:

$$P(y = 1|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}$$

Với  $y$  là biến dùng để phân loại, chỉ nhận hai giá trị 0 hoặc 1,  $X$  là các vector của biến độc lập,  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  là các hệ số cần ước lượng. Các hệ số  $\beta$  thường được ước lượng bằng phương pháp ước lượng hợp lý tối đa ([Hosmer Jr et al., 2013](#)).

#### 2.1.2 Mô hình phân loại tuyến tính

#### 2.1.3 Mô hình SVM (Support Vector Machine)

### 2.2 ĐÁNH GIÁ MÔ HÌNH

#### 2.2.1 Đường ROC và phần diện tích dưới đường cong (AUC)

#### 2.2.2 Thang đo H



## CHƯƠNG 3

### TÌNH HUỐNG NGHIÊN CỨU

#### 3.1 SỐ LIỆU VÀ CÁC BIẾN SỐ

Chúng ta thực hành trên bộ số liệu mẫu bao gồm 30000 quan sát và 25 biến bao gồm tình trạng trả nợ, các thông tin nhân khẩu học cơ bản cùng với số liệu về tín dụng và tình trạng hồ sơ của các khách hàng thẻ tín dụng ở Đài Loan từ tháng 4 năm 2005 đến tháng 9 năm 2005.

Các tên biến đã được thay đổi để tiện lợi cho việc đọc hiểu và phân tích, cụ thể như sau:

ID Số ID của mỗi khách hàng tín dụng

LIMIT\_BAL Lượng tín dụng cho vay tính bằng Đô la Đài Loan (bao gồm cả các khoản vay cá nhân và các khoản vay với thẻ tín dụng phụ)

SEX Giới tính (1=Nam, 2=Nữ)

EDUCATION (1=sau đại học, 2=đại học, 3=phổ thông, 4=khác, 5=không rõ, 6=không rõ)

MARRIAGE Trạng thái hôn nhân (1=đã cưới, 2=độc thân, 3=khác)

AGE Số tuổi tính bằng năm

PAY\_0 Tình trạng hồ sơ vào thời điểm tháng 9/2005 (-1=trả đúng hạn, 1=chậm 1 tháng, 2=chậm 2 tháng, ... 8=chậm 8 tháng, 9=chậm 9 tháng hoặc nhiều hơn)

PAY\_2 Tình trạng hồ sơ vào thời điểm tháng 8/2005 (thang điểm như trên)

PAY\_3 Tình trạng hồ sơ vào thời điểm tháng 7/2005 (thang điểm như trên)

PAY\_4 Tình trạng hồ sơ vào thời điểm tháng 6/2005 (thang điểm như trên)

PAY\_5 Tình trạng hồ sơ vào thời điểm tháng 5/2005 (thang điểm như trên)

PAY\_6 Tình trạng hồ sơ vào thời điểm tháng 4/2005 (thang điểm như trên)

BILL\_AMT1 Hóa đơn thanh toán vào thời điểm 9/2005 (Đô la Đài Loan)

BILL\_AMT2 Hóa đơn thanh toán vào thời điểm 8/2005 (Đô la Đài Loan)

BILL\_AMT3 Hóa đơn thanh toán vào thời điểm 7/2005 (Đô la Đài Loan)

BILL\_AMT4 Hóa đơn thanh toán vào thời điểm 6/2005 (Đô la Đài Loan)

BILL\_AMT5 Hóa đơn thanh toán vào thời điểm 5/2005 (Đô la Đài Loan)

BILL\_AMT6 Hóa đơn thanh toán vào thời điểm 4/2005 (Đô la Đài Loan)

PAY\_AMT1 Lượng tiền đã thanh toán vào thời điểm tháng 9/2015 (Đô la Đài Loan)

PAY\_AMT2 Lượng tiền đã thanh toán vào thời điểm tháng 8/2015 (Đô la Đài Loan)

PAY\_AMT3 Lượng tiền đã thanh toán vào thời điểm tháng 7/2015 (Đô la Đài Loan)

PAY\_AMT4 Lượng tiền đã thanh toán vào thời điểm tháng 6/2015 (Đô la Đài Loan)

PAY\_AMT5 Lượng tiền đã thanh toán vào thời điểm tháng 5/2015 (Đô la Đài Loan)

PAY\_AMT6 Lượng tiền đã thanh toán vào thời điểm tháng 4/2015 (Đô la Đài Loan)

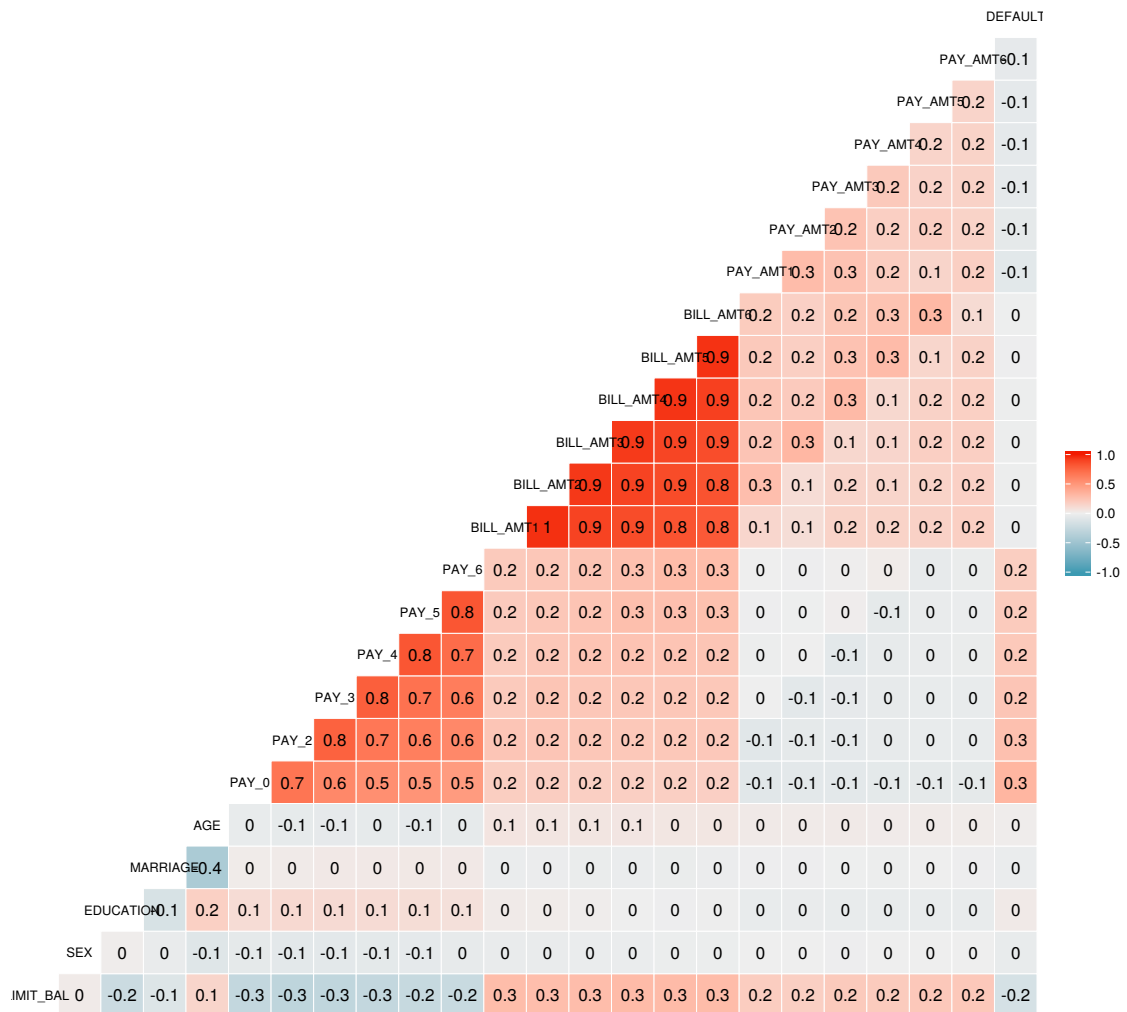
default Có trả nợ hay không (1=có, 0=không)

Hình 3.1 mô tả ma trận hệ số tương quan giữa các biến số trong bộ số liệu

### **3.2 ỨNG DỤNG MÔ HÌNH LOGIT**

### **3.3 ỨNG DỤNG MÔ HÌNH PHÂN LOẠI TUYẾN TÍNH**

### **3.4 ỨNG DỤNG MÔ HÌNH SVM**



Hình 3.1: Ma trận hệ số tương quan giữa các biến trong bộ số liệu.

## **CHƯƠNG 4**

## **KẾT LUẬN**

## TÀI LIỆU THAM KHẢO

- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54(6):627–635.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.
- Lessmann, S., Baesens, B., Seow, H.-V., and Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1):124–136.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Thomas, L. C. (2010). Consumer finance: Challenges for operational research. *Journal of the Operational Research Society*, 61(1):41–52.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Xiao, W., Zhao, Q., and Fei, Q. (2006). A comparative study of data mining methods in consumer loans credit scoring management. *Journal of Systems Science and Systems Engineering*, 15(4):419–435.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.