

TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN
KHOA TOÁN KINH TẾ



CHUYÊN ĐỀ THỰC TẬP

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

Hà Nội, Ngày 22 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.

Đã có rất nhiều nghiên cứu với các phương pháp khác nhau dành cho việc chấm điểm tín dụng. Hiệu quả của các phương pháp khác nhau là có sự khác biệt, phụ thuộc vào việc ứng dụng các mô hình và trường hợp cụ thể của từng bộ số liệu. Trong bài này, chúng ta sẽ tiếp cận đến một số phương pháp phân loại các khách hàng tín dụng và rút ra một số kết luận về việc sử dụng các phương pháp thống kê khác nhau sao cho hợp lý.

Văn bản này được soạn thảo bằng \LaTeX , 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),...

Em xin cảm ơn giáo viên hướng dẫn, cô Nguyễn Thị Minh, bộ 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
1 Tổng quan	6
1.1 Khái niệm	6
1.2 Một số mô hình chấm điểm tín dụng	6
1.3 Thực trạng sử dụng các mô hình tín dụng tại Việt Nam	6
1.4 Kết luận	6
2 Các phương pháp phân loại khách hàng sử dụng thẻ tín dụng	7
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
3 Tình huống nghiên cứu	8
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

4 Kết luận	10
Tài liệu tham khảo	11

DANH SÁCH BẢNG

DANH SÁCH HÌNH VẼ

CHƯƠNG 1

TỔNG QUAN

1.1 KHÁI NIỆM

1.2 MỘT SỐ MÔ HÌNH CHẤM ĐIỂM TÍN DỤNG

1.3 THỰC TRẠNG SỬ DỤNG CÁC MÔ HÌNH TÍN DỤNG TẠI VIỆT NAM

1.4 KẾT LUẬN

CHƯƠNG 2

CÁC PHƯƠNG PHÁP PHÂN LOẠI KHÁCH HÀNG SỬ DỤNG THẺ TÍN DỤNG

2.1 CÁC MÔ HÌNH PHÂN LOẠI

2.1.1 Mô hình logistic

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)

3.2 ỨNG DỤNG MÔ HÌNH LOGIT

3.3 ỨNG DỤNG MÔ HÌNH PHÂN LOẠI TUYẾN TÍNH

3.4 ỨNG DỤNG MÔ HÌNH SVM

CHƯƠNG 4

KẾT LUẬN

TÀI LIỆU THAM KHẢO

- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.