

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 trong 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 24 tháng 4 năm 2017

MỤC LỤC

Mục lục	1
Danh sách bảng	2
Danh sách hình	3
Lời mở đầu	4
CHƯƠNG I: Tổng quan về quản trị rủi ro tín dụng đối với khách hàng cá nhân.	6
1 Một số khái niệm	6
2 Thực trạng của việc chấm điểm tín dụng tại Việt Nam	6
3 Kết luận	6
CHƯƠNG II: Các phương pháp phân loại khách hàng vay thẻ tín dụng	7
1 Các mô hình phân loại	7
1.1 Mô hình logit	7
1.2 Mô hình phân loại tuyến tính	7
1.3 Mô hình SVM (Support Vector Machine)	7
2 Đánh giá mô hình	7
2.1 Đường ROC và phần diện tích dưới đường cong (AUC)	7
2.2 Thang đo H	7
CHƯƠNG III: Tình huống nghiên cứu.	8
1 Số liệu và các biến số	8
2 Ứng dụng mô hình logit	8
3 Ứng dụng mô hình phân loại tuyến tính	8
4 Ứng dụng mô hình SVM	8
CHƯƠNG IV: Kết luận.	9
PHỤ LỤC A: Thông tin về phiên làm việc trên R	10

Tài liệu tham khảo.	13
--------------------------------------	-----------

DANH SÁCH BẢNG

DANH SÁCH HÌNH VẼ

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, Zhao, and Fei 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 3.4.0 (2017-04-21) (R Core Team 2017), các phân tích cụ thể được thực hiện sử dụng các gói mở rộng caret (Jed Wing et al. 2016), tidyverse (Wickham 2017)... Mô hình logit được thực hiện với gói glmnet (Friedman, Hastie, and Tibshirani 2010). Mô hình SVM được thực hiện với gói kernlab (Karatzoglou et al. 2004), là một giao diện của phần mềm LIBSVM (Chang and Lin 2011) trong môi trường R.

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

CHƯƠNG I

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

- 1 MỘT SỐ KHÁI NIỆM**
- 2 THỰC TRẠNG CỦA VIỆC CHẤM ĐIỂM TÍN DỤNG TẠI VIỆT NAM**
- 3 KẾT LUẬN**

CHƯƠNG II

CÁC PHƯƠNG PHÁP PHÂN LOẠI KHÁCH HÀNG VAY THẺ TÍN DỤNG

1 CÁC MÔ HÌNH PHÂN LOẠI

1.1 Mô hình logit

1.2 Mô hình phân loại tuyến tính

1.3 Mô hình SVM (Support Vector Machine)

2 ĐÁNH GIÁ MÔ HÌNH

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

2.2 Thang đo H

CHƯƠNG III

TÌNH HUỐNG NGHIÊN CỨU

- 1 SỐ LIỆU VÀ CÁC BIẾN SỐ**
- 2 ỨNG DỤNG MÔ HÌNH LOGIT**
- 3 ỨNG DỤNG MÔ HÌNH PHÂN LOẠI TUYẾN TÍNH**
- 4 ỨNG DỤNG MÔ HÌNH SVM**

CHƯƠNG IV

KẾT LUẬN

PHỤ LỤC A

THÔNG TIN VỀ PHIÊN LÀM VIỆC TRÊN R

```
## R version 3.4.0 (2017-04-21)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.2 LTS
##
## Matrix products: default
## BLAS: /usr/lib/libblas/libblas.so.3.6.0
## LAPACK: /usr/lib/lapack/liblapack.so.3.6.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=vi_VN             LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=vi_VN        LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=vi_VN           LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=vi_VN     LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets
## [6] methods    base
##
## other attached packages:
##  [1] glmnet_2.0-5    foreach_1.4.3  Matrix_1.2-8
##  [4] ggfortify_0.4.1 GGally_1.3.0    caret_6.0-76
##  [7] ggplot2_2.2.1   lattice_0.20-35 broom_0.4.2
## [10] dplyr_0.5.0     tidyr_0.6.1     readr_1.1.0
## [13] knitr_1.15.1
##
```

```
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.10      RColorBrewer_1.1-2
## [3] nloptr_1.0.4      compiler_3.4.0
## [5] plyr_1.8.4        iterators_1.0.8
## [7] tools_3.4.0       lme4_1.1-13
## [9] evaluate_0.10     tibble_1.3.0
## [11] gtable_0.2.0      nlme_3.1-131
## [13] mgcv_1.8-17       psych_1.7.3.21
## [15] DBI_0.6-1         parallel_3.4.0
## [17] SparseM_1.77      gridExtra_2.2.1
## [19] stringr_1.2.0     MatrixModels_0.4-1
## [21] hms_0.3           stats4_3.4.0
## [23] grid_3.4.0        nnet_7.3-12
## [25] reshape_0.8.6     R6_2.2.0
## [27] foreign_0.8-67    minqa_1.2.4
## [29] reshape2_1.4.2    car_2.1-4
## [31] magrittr_1.5      splines_3.4.0
## [33] scales_0.4.1      codetools_0.2-15
## [35] ModelMetrics_1.1.0 MASS_7.3-47
## [37] assertthat_0.2.0  pbkrtest_0.4-7
## [39] mnormt_1.5-5      colorspace_1.3-2
## [41] quantreg_5.33     stringi_1.1.5
## [43] lazyeval_0.2.0    munsell_0.4.3
```

TÀI LIỆU THAM KHẢO

1. R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
2. Wickham, Hadley (2017). *tidyverse: Easily Install and Load 'Tidyverse' Packages*. R package version 1.1.1. URL: <https://CRAN.R-project.org/package=tidyverse>.
3. Jed Wing, Max Kuhn. Contributions from et al. (2016). *caret: Classification and Regression Training*. R package version 6.0-73. URL: <https://CRAN.R-project.org/package=caret>.
4. Lessmann, Stefan et al. (2015). “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research”. In: *European Journal of Operational Research* 247.1, pp. 124–136.
5. Xie, Yihui (2015). *Dynamic Documents with R and knitr*. 2nd. ISBN 978-1498716963. Boca Raton, Florida: Chapman and Hall/CRC. URL: <http://yihui.name/knitr/>.
6. Chang, Chih-Chung and Chih-Jen Lin (2011). “LIBSVM: A library for support vector machines”. In: *ACM Transactions on Intelligent Systems and Technology* 2 (3). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 27:1–27:27.
7. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1, pp. 1–22. URL: <http://www.jstatsoft.org/v33/i01/>.

8. Thomas, Lyn C (2010). “Consumer finance: Challenges for operational research”. In: *Journal of the Operational Research Society* 61.1, pp. 41–52.
9. Xiao, Wenbing, Qian Zhao, and Qi Fei (2006). “A comparative study of data mining methods in consumer loans credit scoring management”. In: *Journal of Systems Science and Systems Engineering* 15.4, pp. 419–435.
10. Karatzoglou, Alexandros et al. (2004). “kernlab – An S4 Package for Kernel Methods in R”. In: *Journal of Statistical Software* 11.9, pp. 1–20. URL: <http://www.jstatsoft.org/v11/i09/>.
11. Baesens, Bart et al. (2003). “Benchmarking state-of-the-art classification algorithms for credit scoring”. In: *Journal of the operational research society* 54.6, pp. 627–635.