

MACHINE LEARNING

MÔ HÌNH SCORECARD



Vũ Thị Phương Thảo
Trần Thế Khải
Trần Đức Dũng



OUTLINE

1. Mô hình Scorecard (Xếp hạng tín nhiệm) là gì?
2. Xây dựng mô hình Scorecard trong quản trị rủi ro Ngân hàng
3. Đưa mô hình lên web app



MÔ HÌNH SCORECARD LÀ GÌ?

Mô hình Scorecard được sử dụng để tính điểm tín nhiệm cho một cá nhân hoặc một tổ chức cụ thể dựa trên khả năng xảy ra của một sự kiện nào đó như vỡ nợ, vi phạm luật.

Dựa trên điểm tín nhiệm này, các tổ chức tài chính hoặc chính phủ có thể cung cấp các sản phẩm, dịch vụ tốt hơn nếu chủ thẻ có điểm tín nhiệm cao và thấp hơn đối với chủ thẻ có điểm tín nhiệm thấp.



MÔ HÌNH SCORECARD TRONG QTRR NGÂN HÀNG

1. Phương pháp chuyên gia

Dựa trên ý kiến thẩm định của các chuyên gia, căn cứ vào các thông tin khách hàng như điều kiện tài chính, tài sản đảm bảo của chủ thể vay, điều kiện và xu hướng thị trường,...



2. Phương pháp mô hình

Dùng mô hình học máy:

- Tăng năng suất
- Giảm thiểu chi phí lao động
- Kết quả nhất quán
- Không bị hạn chế



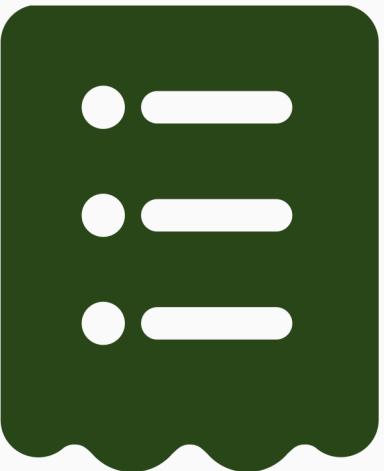
THÔNG TIN KHÁCH HÀNG



Nhân khẩu học



Lịch sử tín dụng



Thông tin giao dịch



Thông tin tài sản đảm bảo

PP XÂY MÔ HÌNH

Có nhiều thuật toán được sử dụng để xây dựng mô hình Scorecard như hồi quy Logistic, Neural Networks, Random Forests, và SVM,... Tuy nhiên, bài nghiên cứu này sẽ tập trung vào phương pháp thông dụng nhất là **hồi quy Logistic**.

Quá trình hồi quy sẽ tiếp nhận các thuộc tính đầu vào (input features) theo phương pháp **trọng số dấu hiệu**, hay còn gọi là **WOE** - Weight of Evidence.



Weight of Evidence

$$WOE_i = \ln \left(\frac{\% \text{ Good}_i}{\% \text{ Bad}_i} \right)$$

PP WOE

WOE là một trong những kỹ thuật trích chọn thuộc tính (feature engineering)

Ý tưởng của WOE là chia mẫu thành các nhóm có số quan sát bằng nhau/gần bằng nhau với biến liên tục hoặc thành các nhóm định danh với biến định danh, theo từng thuộc tính một. Các nhóm được chia ra được gọi là **bin**. Và trên từng nhóm, ta sẽ tính ra tỷ lệ %Good, %Bad của nhóm đó so với toàn bộ mẫu.

$$\% \text{ Good}_i = \frac{\text{Số lượng hồ sơ tốt ở bin thứ } i}{\text{Tổng số lượng hồ sơ tốt trên tất cả các bin}}$$

$$\% \text{ Bad}_i = \frac{\text{Số lượng hồ sơ xấu ở bin thứ } i}{\text{Tổng số lượng hồ sơ xấu trên tất cả các bin}}$$

PP WOE

Tiêu chuẩn xếp hạng thuộc tính được tính từ WOE

Tiêu chuẩn xếp hạng ở đây là **chỉ số giá trị thông tin IV - Information value**, được tính toán từ phương pháp WOE.

$$IV = \sum_i (\%Good_i - \%Bad_i) \times WOE_i$$

VD TÍNH WOE & IV

Giả sử với 1 thuộc tính cụ thể là số năm kinh nghiệm trong nghề (Years on Job)

Bin	Số quan sát	Số lượng hồ sơ tốt	Số lượng hồ sơ xấu	%Good	%Bad
0-1	100	66	34	0.165	0.34
1-4	100	75	25	0.1875	0.25
4-8	100	82	18	0.205	0.18
8-15	100	87	13	0.2175	0.13
15+	100	90	10	0.225	0.1
Tổng		400	100		

Bước 1:

Tính %Good, %Bad cho Bin 0-1, làm tương tự cho các bin còn lại

1

$$\%Good_{0-1} = \frac{66}{400}$$

$$\%Bad_{0-1} = \frac{34}{100}$$

VD TÍNH WOE & IV

Giả sử với 1 thuộc tính cụ thể là số năm kinh nghiệm trong nghề (Years on Job)

Bin	%Good	%Bad	WOE
0-1	0.165	0.34	-0.723
1-4	0.1875	0.25	-0.288
4-8	0.205	0.18	0.130
8-15	0.2175	0.13	0.515
15+	0.225	0.1	0.811
Tổng			

Bước 2:

Tính WOE cho Bin 0-1, làm tương tự cho các bin còn lại

2

$$WOE_{0-1} = \ln\left(\frac{0.165}{0.34}\right)$$

VD TÍNH WOE & IV

Giả sử với 1 thuộc tính cụ thể là số năm kinh nghiệm trong nghề (Years on Job)

Bin	%Good	%Bad	WOE	IV
0-1	0.165	0.34	-0.723	0.127
1-4	0.1875	0.25	-0.288	0.018
4-8	0.205	0.18	0.130	0.003
8-15	0.2175	0.13	0.515	0.045
15+	0.225	0.1	0.811	0.101
Tổng				0.294

Bước 3:

Tính IV cho bin 0-1, làm tương tự cho các bin còn lại



$$IV_{0-1} = (0.165 - 0.34) \times (-0.723)$$

VD TÍNH WOE & IV

Ta cộng tổng IV của các bin để ra tổng IV của thuộc tính:

Tổng IV (YOJ) = 0.294

Giá trị này có ý nghĩa gì?



TIÊU CHUẨN PHÂN LOẠI THEO IV

Tiêu chuẩn phân loại sức mạnh của biến theo giá trị IV:

- ≤ 0.02 : Biến không có tác dụng trong việc phân loại hồ sơ
- $0.02 - 0.1$: Yếu
- $0.1 - 0.3$: Trung bình
- $0.3 - 0.5$: Mạnh
- $\Rightarrow 0.5$: Biến rất mạnh



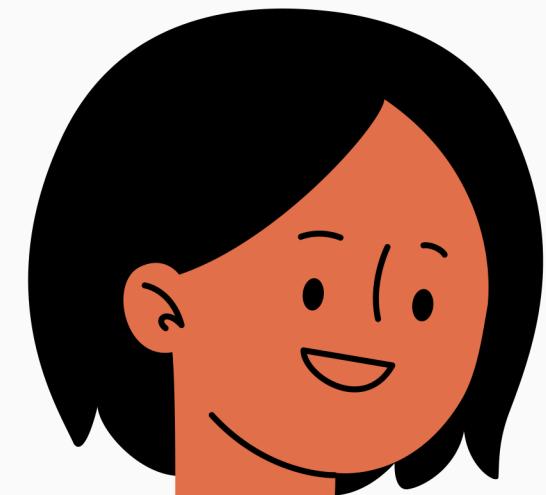
ƯU ĐIỂM & NHƯỢC ĐIỂM CỦA WOE

- Giúp các hệ số trong phương trình hồi qui logistic giải thích được đúng thực tế mối quan hệ giữa biến độc lập với biến phụ thuộc.
- Giúp loại bỏ các outliers
- Phản ánh được ảnh hưởng của từng biến lên biến phụ thuộc
- Đối với các biến quá phân tán thì WOE sẽ nhóm thành các nhóm và hệ số WOE thể hiện thông tin cho toàn bộ nhóm.



- Khi tính toán WOE, rất khó để biết phân chia bao nhiêu bins là phù hợp đối với các biến liên tục hoặc khi nào thì nên nhóm các nhóm với nhau hoặc tách nhóm với các biến định danh.
- Do các biến WOE luôn đơn điệu với biến phụ thuộc nên giữa các biến độc lập luôn có sự tương quan (do cùng tương quan với biến phụ thuộc). Điều này có thể dẫn đến nguy cơ đa cộng tuyến cao ảnh hưởng tới khả năng giải thích của hệ số hồi qui.
- Dễ dàng xảy ra overfitting do có thể hiệu chỉnh ảnh hưởng của biến bằng cách nhóm các biến định danh.

CÁC BƯỚC XÂY MÔ HÌNH



Bước 1: Khảo sát dữ liệu

Bước 2: Chia bin cho các thuộc tính và tính điểm WOE, IV cho từng bin

Bước 3: Train mô hình theo phương pháp hồi qui logistic

Bước 4: Tính điểm tín nhiệm (credit scores) trên các thuộc tính

Đưa mô hình lên web app

Scorecard model

Please upload your file here. Only .csv accepted.

Choose a file

Drag and drop file here
Limit 200MB per file

cust.csv 317.0B

	NAME	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLA
0	Tran Duc Dung	2,300	98,449	117,195	HomeImp	Office	4	0	0	83
1	Vu Thi Phuong Thao	2,900	78,600	100,000	DebtCon	Office	6	1	0	145
2	Tran The Khai	2,400	34,863	47,471	HomeImp	Mgr	12	0	0	73

The credit scores are:

	Name	Credit Score	Rating
0	Tran Duc Dung	693.4812	Good
1	Vu Thi Phuong Thao	721.2913	Good
2	Tran The Khai	578.6594	Very Poor

Thank
you!