

ỨNG DỤNG MÔ HÌNH PHÂN LỚP TRONG ĐÁNH GIÁ TÍN DỤNG DOANH NGHIỆP Ở VIỆT NAM

● NGUYỄN QUỐC ANH

TÓM TẮT:

Bài nghiên cứu sử dụng những chỉ số tài chính phổ biến nhằm tìm ra một mô hình đánh giá tín dụng công ty. Đồng thời kiểm tra xem các chỉ số tài chính nào ảnh hưởng đáng kể đến mức độ đánh giá tín dụng công ty. Với bộ dữ liệu gồm 256 công ty được đánh giá tín dụng bởi Moodys và những chỉ số tài chính, nhóm sử dụng những mô hình phân lớp bằng phương pháp thống kê (Logistic Regression) và phương pháp máy học (SVM, Neural Networks). Kết quả là những mô hình phân lớp có độ chính xác trong khoảng 60-70%. Nhóm cũng thực hiện kiểm định và kết luận được những chỉ số nào có tác động đáng kể cũng như những chỉ số không có tác động đáng kể.

Từ khóa: doanh nghiệp, mô hình phân lớp, tín dụng.

1. Đặt vấn đề

Sau khi trải qua cuộc khủng hoảng kinh tế 2007 - 2009, các doanh nghiệp (DN) Việt Nam đã chuyển mình để phát triển, mở rộng quy mô với một tốc độ nhanh chưa từng thấy. Đặc biệt, các DN xây dựng, kinh doanh và bất động sản đã phát triển cực kỳ mạnh mẽ từ sự kiện khủng hoảng đóng băng thị trường bất động sản năm 2013. Cũng vì thế, xuyên suốt giai đoạn này, nhu cầu về vốn của các DN cực kỳ lớn và sự cạnh tranh thu hút nguồn vốn là cực kỳ khốc liệt (Bùi Phúc Trung, 2019).

Trong suốt thời kỳ 2014 - 2019, nền kinh tế Việt Nam tăng trưởng ấn tượng với mức trung bình trên 5%/năm. Nguồn động lực cho sự tăng trưởng ấy phần lớn đến từ sự phát triển nhanh chóng của những DN. Cũng trong thời gian vừa qua, nhiều DN bất động sản như Tân Hoàng Minh đã vi phạm nghiêm trọng về phát hành trái phiếu, tạo nên một làn sóng khủng hoảng trái phiếu đáo hạn ảnh hưởng đến nhiều công ty, đặc biệt là ngành tài chính - ngân hàng. Ngoài ảnh hưởng đến những DN, những trái chủ, đặc biệt là

những nhà đầu tư cá nhân, chịu ảnh hưởng nặng nề không kém về cả lòng tin và tài chính.

Sự ảnh hưởng tức thời không chỉ dừng lại ở thiệt hại về tài chính, mà còn làm suy giảm lòng tin vào thị trường chứng khoán của mọi người, làm suy giảm đầu tư vào thị trường chứng khoán. Trong giai đoạn nền kinh tế đang đối mặt với áp lực lạm phát cao, suy giảm tăng trưởng, lãi suất vay ngân hàng cao, việc khó tiếp cận nguồn vốn tạo ra áp lực buộc nhiều DN phải xem xét tới việc giảm quy mô phát triển (Nguyễn Thị Thu Thủy và Nguyễn Văn Thuận, 2021).

Đứng trên góc độ là nhà đầu tư cá nhân, việc lựa chọn một trái phiếu tốt để đầu tư dài hạn là một quyết định đầu tư quan trọng. Tuy nhiên, sự hạn chế về khả năng tiếp cận thông tin cũng như phân tích chuyên sâu về tình hình tài chính DN đã khiến nhiều người có những quyết định đầu tư sai lầm. Nghiên cứu này được thực hiện nhằm giải quyết một phần sự hạn chế về tiếp cận cũng như hiểu biết thông tin DN của nhà đầu tư cá nhân. Trong đó, tác

giả thực hiện xác định các yếu tố ảnh hưởng đến Đánh giá tín dụng của các công ty, đề xuất mô hình nghiên cứu, kiểm định thang đo các yếu tố ảnh hưởng đến Đánh giá tín dụng của các công ty. Từ đó, đề xuất những khuyến nghị cho các nhà đầu tư cá nhân, không chuyên trong việc tự thẩm định trái phiếu phát hành của các công ty ở Việt Nam để đưa ra quyết định đầu tư an toàn.

2. Cơ sở lý thuyết

2.1. Tín dụng

Tín dụng trong lĩnh vực tài chính mang nhiều nghĩa, nhưng nghĩa thường dùng nhất là một thỏa thuận hợp đồng - nơi bên đi vay nhận một khoản tiền, hay một vật giá trị và cam kết trả đủ cho bên cho vay khi đến hạn, thường là với một mức lãi suất (Chen & Wu, 2019). Tín dụng cũng có thể là uy tín hay lịch sử tín dụng của một cá nhân hoặc một công ty. Ngoài ra, trong lĩnh vực kế toán, tín dụng còn là một loại bút toán kế toán cụ thể.

2.2. Thang đo mức độ tín dụng

Xếp hạng tín dụng hay xếp hạng tín nhiệm là việc dự báo khả năng trả nợ, đánh giá rủi ro tín dụng cũng như đưa ra những dự báo về khả năng vỡ nợ của đối tượng đi vay thông qua một hệ thống được xếp hạng sẵn (Veeramanikandan & Jeyakarthic, 2019). Định nghĩa này không bó hẹp trong đối tượng là một cá nhân hay tổ chức cụ thể mà áp dụng đối với mọi tổ chức đang có nhu cầu đi vay bao gồm: cá nhân, DN, Nhà nước, Chính phủ.

Xếp hạng tín dụng DN thường được thực hiện bởi các cơ quan chuyên nghiệp. Các cơ quan này sẽ tiến hành xem xét các loại nợ của công ty và đánh

giá ảnh hưởng. Đặc biệt Standard & Poors (S&P)D, Moodys và Fitch là 3 cơ quan xếp hạng tín dụng lớn và uy tín nhất thế giới.

3. Mô hình nghiên cứu

Bản chất của việc phân loại mức đánh giá tín dụng của DN dựa trên những chỉ số tài chính là bài toán phân lớp dữ liệu. Tác giả sử dụng những mô hình phân lớp phổ biến như hồi quy logistic, SVM, Neural Networks để thực hiện bài nghiên cứu này. (Hình 1)

Từ mô hình nghiên cứu, có 15 giả thuyết nghiên cứu được đề xuất như sau:

H1: Chỉ số thanh toán hiện hành Current Ratio có ảnh hưởng đến đánh giá tín dụng.

H2: Chỉ số nợ dài hạn trên tổng nguồn vốn Long-term Debt/Capital có ảnh hưởng đến đánh giá tín dụng.

H3: Chỉ số nợ trên vốn chủ Debt-Equity Ratio có ảnh hưởng đến đánh giá tín dụng.

H4: Chỉ số biên lợi nhuận gộp Gross Margin có ảnh hưởng đến đánh giá tín dụng.

H5: Chỉ số vòng quay hàng tồn kho Inventory Turnover có ảnh hưởng đến đánh giá tín dụng.

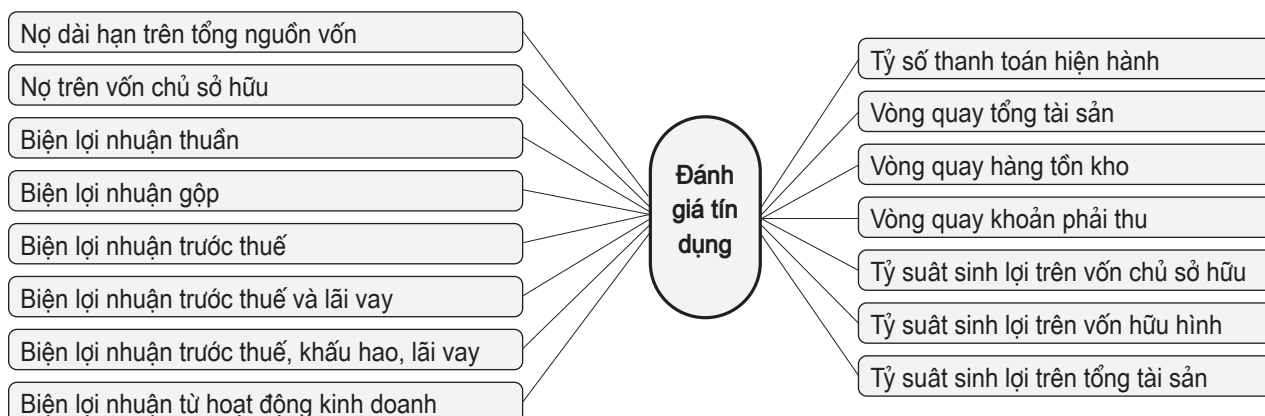
H6: Chỉ số biên lợi nhuận từ hoạt động kinh doanh Operating Margin có ảnh hưởng đến đánh giá tín dụng.

H7: Chỉ số biên lợi nhuận trước thuế và lãi vay EBIT Margin có ảnh hưởng đến đánh giá tín dụng.

H8: Chỉ số vòng quay khoản phải thu Receivable Turnover có ảnh hưởng đến đánh giá tín dụng.

H9: Chỉ số biên lợi nhuận trước thuế, lãi vay và

Hình 1: Mô hình nghiên cứu



khấu hao EBITDA Margin có ảnh hưởng đến đánh giá tín dụng.

H10: Chỉ số biên lợi nhuận trước thuế Pre-Tax Profit Margin có ảnh hưởng đến đánh giá tín dụng.

H11: Chỉ số biên lợi nhuận thuần Net Profit Margin có ảnh hưởng đến đánh giá tín dụng.

H12: Chỉ số vòng quay tổng tài sản Assets Turnover có ảnh hưởng đến đánh giá tín dụng.

H13: Chỉ số tỷ suất sinh lợi trên vốn chủ Return on Equity có ảnh hưởng đến đánh giá tín dụng.

H14: Chỉ số tỷ suất sinh lợi trên vốn hữu hình Return on Tangible Equity có ảnh hưởng đến đánh giá tín dụng.

H15: Chỉ số tỷ suất sinh lợi trên tổng tài sản Return on Assets có ảnh hưởng đến đánh giá tín dụng.

4. Phương pháp nghiên cứu

Tác giả đã kiểm tra dữ liệu và mã hóa để đưa lên phần mềm Orange Data Mining 3.34.0 và SPSS.26 để tiếp tục xử lý và phân tích dữ liệu hơn. Dữ liệu nghiên cứu được chọn từ 256 công ty phù hợp với nghiên cứu này. Đây là các DN được đánh giá tín dụng công khai trên website moodys.com.

5. Kết quả nghiên cứu

5.1. Kiểm định sự phù hợp của mô hình (Bảng 1)

Kết quả nghiên cứu cho thấy, chỉ số AUC: các mô hình có chỉ số tương đối giống nhau, sự chênh lệch là không đáng kể. Điều đó cho thấy với bộ dữ liệu, các phương pháp phân lớp đều cho ra kết quả tương đối khả quan với bộ dữ liệu.

Về những chỉ số còn lại: ngoại trừ SVM, 2 mô hình Neural Networks và Logistic Regression có hiệu suất phân lớp chính xác tương đối gần nhau. Cụ thể: (1) Tính chính xác sử dụng Classification Accuracy - CA: 2 mô hình cho ra kết quả tương đối gần nhau ở khoảng 65 - 70% là khá cao, tuy nhiên cần xem xét thêm những chỉ số khác; (2) Độ chuẩn xác Precision: 2 mô hình cho kết quả tương tự như CA khi nằm trong khoảng 65 - 70%, có thể

xem như một kết quả khả quan cho cả 2 mô hình; (3) Độ phủ Recall: cả 2 mô hình cho kết quả phân lớp trong khoảng 65 - 70%, cho thấy việc mô hình phân loại đúng khi so sánh với bộ dữ liệu huấn luyện gốc đạt hiệu suất khá cao; (4) Chỉ số F1: là một chỉ số có tính điều hòa giữa 2 chỉ số Precision và Recall, chỉ số F1 cao cho thấy, mô hình có sự phù hợp cao với bộ dữ liệu. Với mức F1 trong khoảng 65 - 70%, cả 2 mô hình Neural Networks và Logistic Regression cho thấy sự phù hợp tốt và có thể ứng dụng vào thực tế với những tham số giống với mô hình.

Ngoài việc xem xét các chỉ số, tác giả còn xem xét ma trận nhầm lẫn để tìm kiếm những hạn chế của mô hình. (Bảng 2)

Có thể thấy, nhóm A có tỷ lệ dự báo đúng là 65,31%, nhóm B là 76,38%, chỉ có nhóm C là 32,26%. Như vậy, việc nhóm C có tỷ lệ dự báo thấp hơn 2 nhóm còn lại rất có thể đến từ nguyên nhân gom nhóm ban đầu không tốt, có sự tương đồng lớn giữa những đối tượng thuộc B và C khiến mô hình bị nhầm lẫn khi thực hiện phân lớp. Những đối tượng gây nhầm lẫn rất có thể đến từ những công ty có xếp hạng cận dưới nhóm B và cận trên nhóm C.

Bảng 1. Kết quả chỉ số đánh giá mô hình phân lớp

Mô hình	AUC	CA	F1-score	Precision	Recall
SVM	0.789	0.609	0.567	0.548	0.609
Neural Networks	0.767	0.656	0.647	0.649	0.656
Logistic Regression	0.772	0.668	0.663	0.665	0.668

Nguồn: Kết quả phân tích từ SPSS

Bảng 2. Ma trận nhầm lẫn Logistic Regression

Mô hình phân lớp dự báo					
Dữ liệu gốc		A	B	C	Tổng
	A	64	30	4	98
	B	20	97	10	127
	C	2	19	10	31
	Tổng	86	146	24	256

Nguồn: Kết quả phân tích từ SPSS

Bảng 3. Kiểm định Likelihood Ratio

Biến độc lập	Giả thuyết	Chi bình phương	Sig.	Chấp nhận/ Bác bỏ
Hệ số chặn		40.032	0	
Chỉ số thanh toán	H1	32.344	0	Chấp nhận
Nợ dài hạn/ Vốn chủ sở hữu	H2	34.202	0	Chấp nhận
Chỉ số nợ/ Vốn chủ sở hữu	H3	89.738	0	Chấp nhận
Biên lợi nhuận gộp	H4	5.263	72	Chấp nhận
Vòng quay hàng tồn kho	H5	303	859	Bác bỏ
Biên lợi nhuận hoạt động	H6	827	661	Bác bỏ
EBIT	H7	6.598	37	Chấp nhận
Vòng quay khoản phải thu	H8	2.589	274	Bác bỏ
EBITDA	H9	21.058	0	Bác bỏ
Biên lợi nhuận trước thuế	H10	2.349	309	Bác bỏ
Biên lợi nhuận thuần	H11	1.182	554	Bác bỏ
Vòng quay tổng tài sản	H12	2.797	247	Bác bỏ
Lợi nhuận trên vốn chủ sở hữu (ROE) %	H13	39	981	Bác bỏ
Lợi nhuận trên vốn chủ sở hữu hữu hình %	H14	2.232	328	Bác bỏ
Lợi nhuận trên tổng tài sản (ROA) %	H15	87.994	0	Chấp nhận

Nguồn: Kết quả phân tích từ SPSS

Trong một mô hình phân lớp tốt, hiện tượng 2 nhóm cách xa nhau như A-C bị phân loại nhầm sẽ xấp xỉ dưới 5% và mô hình đã đáp ứng yêu cầu với chỉ 2,34% là phân loại nhầm. Như vậy, dựa trên những nhận xét về chỉ số đánh giá cũng như từ mô hình, có thể thấy 2 mô hình Neural Networks và Logistic Regression sẽ là 2 mô hình phù hợp cho dữ liệu. Việc sử dụng mô hình nào cũng sẽ đáp ứng tốt với bộ dữ liệu.

5.2. Kiểm định giả thuyết nghiên cứu

Để thực hiện đánh giá kiểm định giả thuyết, nhóm sử dụng mô hình Logistic Regression và dùng kiểm định Likelihood Test cũng như chỉ số Sig để có tương quan đánh giá với mức ý nghĩa thống kê 10%. (Bảng 3)

Như vậy, với mức ý nghĩa 10%, nhóm bác bỏ giả thuyết H5, H6, H8, H9, H10, H11, H12, H13, H14. Như vậy, những chỉ số không có tác động đáng kể đến đánh giá tín dụng DN là: Vòng quay hàng tồn

kho; Biên lợi nhuận hoạt động; Vòng quay các khoản phải thu; EBITDA; Tỷ suất lợi nhuận trước thuế; Tỷ suất lợi nhuận ròng; Vòng quay tài sản; Lợi nhuận trên vốn chủ sở hữu; Lợi nhuận trên Vốn chủ sở hữu hữu hình.

Đồng thời những giả thuyết được chấp nhận là H1, H2, H3, H4, H7, H15. Những chỉ số có ảnh hưởng đáng kể đến kết quả đánh giá tín dụng là chỉ số thanh khoản; Nợ dài hạn/Vốn; Tỷ lệ nợ trên vốn chủ sở hữu; Biên lợi nhuận gộp; EBIT; Lợi nhuận trên tài sản.

6. Kiến nghị

Mô hình đánh giá khả năng chi trả nợ dựa trên những chỉ số tài chính công ty tại thời điểm công bố báo cáo tài chính. Do đó khi áp dụng mô hình, nhóm tác giả khuyến nghị các nhà đầu tư nên kết hợp thêm các yếu tố phân tích của chuyên gia, phân tích vĩ mô, phân tích ngành,... để hỗ trợ cho quyết định đầu tư thêm phần chính xác.

- Mô hình của nhóm không thể thay thế hoặc ứng dụng thực tiễn cho công việc thẩm định tín dụng vì tính chính xác cũng như tính toàn diện còn hạn chế.

- Những chỉ số nhà đầu tư nên đặc biệt quan tâm

khi xem xét đầu tư trái phiếu các kỳ hạn là những chỉ số về nợ DN, chỉ số về tính thanh khoản cũng như tình hình kinh doanh. Bên cạnh đó, những chỉ số về dòng tiền cũng như cách DN sử dụng dòng tiền cũng có tính cần thiết khi xem xét đầu tư ■

TÀI LIỆU THAM KHẢO:

1. Chen, S., Ma, H., & Wu, Q. (2019). Bank credit and trade credit: Evidence from natural experiments. *Journal of Banking & Finance*, 108, 105616.
2. Thủy, N. T. T., & Thuận, N. V. (2021). Tác động của quản trị công ty đến hiệu quả hoạt động DN: nghiên cứu thực nghiệm từ các công ty niêm yết trên thị trường chứng khoán Việt Nam. *Tạp chí Nghiên cứu Tài chính-Marketing*, 1-17.
3. Trung, B. P. (2019). Ứng dụng phương pháp thống kê xây dựng mô hình xếp hạng tín dụng cho DN niêm yết trên thị trường chứng khoán Việt Nam. *Tạp chí Phát triển kinh tế*, 43-51.
4. Veeramani, V., & Jeyakarthic, M. (2019, November). A futuristic framework for financial credit score prediction system using pso based feature selection with random tree data classification model. In *2019 International conference on smart systems and inventive technology (ICSSIT)* (pp. 826-831). IEEE.

Ngày nhận bài: 20/2/2023

Ngày phản biện đánh giá và sửa chữa: 16/3/2023

Ngày chấp nhận đăng bài: 15/4/2023

Thông tin tác giả:

TS. NGUYỄN QUỐC ANH

Khoa Ngân hàng, Trường Đại học Kinh tế Thành phố Hồ Chí Minh

USING THE CLASSIFICATION MODEL FOR THE CORPORATE CREDIT RATING

● Ph.D **NGUYEN QUOC ANH**

Faculty of Banking, University of Economics Ho Chi Minh City

ABSTRACT:

This study uses popular financial ratios to find a model of corporate credit rating. The study also explores which financial ratios significantly affect the credit rating of a company. The study uses a logistic regression analysis model and machine learning methods (SVM and Neural Networks) to analyze a dataset of 256 companies rated by Moody's credit and their financial indicators. The study's results show that classification models have accuracy scores ranged from 60 to 70 percent. The study also tests and finds out which indicators have significant impacts as well as those that have no significant impact on the credit rating of a company.

Keywords: enterprise, classification model, credit.