TÁC ĐỘNG CỦA NỢ VAY ĐẾN HIỆU QUẢ TÀI CHÍNH DOANH NGHIỆP NIÊM YẾT TRÊN HOSE: NGHIÊN CỨU KẾT HỢP GIỮA MÔ HÌNH HỒI QUY VÀ PHƯƠNG PHÁP HỌC MÁY

ThS. Nguyễn Roãn Dũng[1] - PGS.TS. Nguyễn Anh Hiền[1]*

Bài viết nghiên cứu sự tác động của tỷ lệ nợ vay trên tài sản đến hiệu quả tài chính (ROA) của các doanh nghiệp niêm yết trên Sở giao dịch chứng khoán Thành phố Hồ Chí Minh (HOSE) giai đoạn từ năm 2015 đến năm 2024 với tập dữ liệu gồm 274 doanh nghiệp thuộc các ngành nghề khác nhau. Nghiên cứu sử dụng kết hợp giữa mô hình hồi quy và phương pháp học máy thông qua mô hình kinh tế lượng truyền thống FGLS và sự hỗ trợ của phương pháp học máy LightGBM cùng với giá trị SHAP áp dụng cho việc giải thích mô hình học máy. Kết quả nghiên cứu cho thấy tỷ lệ nợ vay trên tổng tài sản ảnh hưởng trái chiều đến hiệu quả tài chính của các công ty niêm yết trên HOSE.

• Từ khóa: *hiệu quả tài chính, mô hình hồi quy, phương pháp học máy, tỷ lệ nợ vay.*

This article studies the impact of debt to asset ratio on the return on assets (ROA) of listed companies on the Ho Chi Minh Stock Exchange (HOSE) from 2015 to 2024, using a dataset of 274 companies from various industries. The study combines regression and machine learning methods, employing the traditional FGLS econometric model and the LightGBM machine learning approach, along with SHAP values applied to explain the machine learning model. The research results show that the debt-to-total assets ratio negatively affects the financial performance of companies listed on HOSE.

• Key words: *financial performance, regression model, machine learning methods, debt ratio.*

Ngày nhận bài: 31/3/2026
Ngày gửi phản biện: 19/5/2026
Ngày duyệt đăng: 10/6/2026

DOI: https://doi.org/10.71374/jfarv.v26.i315.26

1. Giới thiệu

Doanh nghiệp (DN) hoạt động sản xuất kinh doanh trên thị trường ngoài việc phục vụ cho nhu cầu của xã hội là việc tìm kiếm lợi nhuận. Việc DN có lợi nhuận vừa là cơ sở vừa là động lực để DN phục vụ tốt hơn cho cộng đồng nói chung và cho chính bản thân của DN nói riêng. Tuy nhiên, sự phát triển và mở rộng của DN đòi hỏi phải cần vốn nhiều hơn. Vốn đó có thể bắt nguồn từ trước hết là khoản vốn tự có của bản thân DN, kế đến là khoản lợi nhuận giữ lại và tiếp theo chính là khoản vốn vay. Nguồn vốn vay này có thể là vay từ DN khác, vay từ các cá nhân có liên quan đến DN, vay từ các tổ chức tín dụng và các công ty tài chính, vay từ thị trường vốn bằng cách phát hành trái phiếu. Việc vay nợ trước mắt là giải toả được phần nào nhu cầu về vốn của DN nhưng nó sẽ ảnh hưởng như thế nào đến lợi nhuận của DN? Do đó việc nghiên cứu ảnh hưởng của nợ vay đến hiệu quả tài chính các DN niêm yết trên HOSE là vấn đề cần thiết.

Mặt khác, cùng với sự phát triển bứt phá của khoa học máy tính, các kỹ thuật học máy (Machine Learning - ML) đã trở thành công cụ hỗ trợ đắc lực cho việc xử lý các tập dữ liệu lớn và có cấu trúc phức tạp. Theo James và cộng sự (2013), ưu điểm vượt trội của học máy so với các mô hình kinh tế lượng truyền thống nằm ở khả năng tối ưu hóa độ chính xác trong dự báo. Trong giai đoạn gần đây, các nghiên cứu kinh tế đã bắt đầu kết hợp hoặc thay thế hoàn toàn các mô hình truyền thống bằng các phương pháp học máy.

Do đó, mục tiêu của nghiên cứu này là xác định mức độ ảnh hưởng của nợ vay trên tổng tài sản đến hiệu quả tài chính của các DN niêm yết trên HOSE trong giai đoạn 2015-2024 thông qua việc áp dụng kết hợp giữa mô hình hồi quy truyền thống và phương pháp học máy để củng cố độ tin cậy của kết quả nghiên cứu.

2. Tổng quan nghiên cứu và cơ sở lý thuyết

2.1. Tổng quan nghiên cứu

[1] Trường Đại học Sài Gòn
** Email liên hệ: nguyenanhhien@sgu.edu.vn*

Tổng quan nghiên cứu tác động của nợ vay đến hiệu quả tài chính

Theo các kết quả nghiên cứu nước ngoài và tại Việt Nam cho thấy sự không thống nhất về chiều hướng tác động của nợ vay đến hiệu quả tài chính DN, kết quả nghiên cứu phụ thuộc vào đặc điểm của thị trường và phương pháp nghiên cứu. Ở khu vực Châu Á, Hasan, M.B., Ahsan, A.F.M.M., Rahaman, M.A., & Alam, M.N. (2014) khi nghiên cứu 36 DN tại Bangladesh (giai đoạn 2007-2012) bằng mô hình POLS, FEM và REM đã tìm thấy tác động ngược chiều của nợ vay đến ROA. Tương tự, tại Trung Quốc, Ma, J.H. (2015) sử dụng mô hình hệ phương trình đồng thời (Simultaneous Equations) và Ye, Z. (2021) với kỹ thuật GMM đều xác nhận tác động ngược chiều của đòn bẩy tài chính lên hiệu quả hoạt động (ROE, ROA). Ngược lại, Olusola, B. E., Mengze, H., Chimezie, M. E., & Chinedum, A. P. (2022) tại Hong Kong lại tìm thấy tổng nợ có tác động cùng chiều đến ROA, cho thấy sự khác biệt giữa các thị trường tài chính phát triển và đang phát triển. Tại Anh, Abeywardhana, D.K.Y. (2015) nghiên cứu các DN phi tài chính nhỏ và vừa giai đoạn 1998-2008 cho thấy nợ ngắn hạn có tác động cùng chiều trong khi nợ vay và nợ dài hạn có tác động ngược chiều đến lợi nhuận DN. Tuy nhiên, nghiên cứu của Vuong, N.B., Vu, T.T.Q, & Mitra, P. (2017) trên các DN lớn tại Anh lại chỉ ra nợ dài hạn có quan hệ trái chiều với ROA, ROE, Tobin's Q trong hầu hết các ngành. Tại Romania, Nenu, E.A., Vintilă, G., & Gherghina, S.C. (2018) tìm thấy tác động trái chiều của cả nợ ngắn hạn và dài hạn đến ROA.

Ở trong nước, theo chiều hướng tác động ngược chiều của nợ vay đến hiệu quả tài chính, Đinh Văn Sơn và Trần Thị Phương Thảo (2017) nghiên cứu 346 DN giai đoạn 2011-2015 tìm thấy mối quan hệ ngược chiều giữa tỷ lệ nợ vay và ROA, ROE, Tobin's Q. Ngoài ra, các tác giả cũng phát hiện mối quan hệ phi tuyến (hình chữ U ngược) giữa nợ vay và ROE. Lê Hoàng Vinh và Nguyễn Ngọc Sơn (2018), Lê Minh Tường và Hứa Trung Phúc (2020), Bùi Vĩnh Thanh (2022) đều thống nhất rằng nợ trên tài sản có tác động ngược chiều đến hiệu quả tài chính. Ngược lại, Nguyễn Thị Nguyệt Dung và Hà Thành Công (2022) sử dụng mô hình GMM trên 481 DN (2012-2020) lại cho thấy cấu trúc tài chính có ảnh hưởng cùng chiều đến ROA, ROE.

Tổng quan các nghiên cứu sử dụng phương pháp học máy

Vukovic và cộng sự (2023) đã thực hiện nghiên cứu so sánh khi sử dụng mô hình tác động ngẫu nhiên (REM) để đánh giá các nhân tố ảnh hưởng đến ROA, đồng thời áp dụng các mô hình học máy (Deep Neural Network - DNN, Long Short-Term Memory - LSTM, Random Forest) để dự báo hiệu quả tài chính của các DN bán lẻ từ năm 2017 đến năm 2020. Kết quả cho thấy các thuật toán học máy nâng cao mang lại khả năng dự báo chính xác hơn đáng kể. Thay vì tiếp cận theo hướng giải thích, một số nghiên cứu tập trung thẳng vào mục tiêu dự báo. Chẳng hạn, Saha và cộng sự (2023) đã áp dụng phương pháp Regression Trees, Boosted Trees, Bayesian Additive Regression Trees để dự báo doanh thu và lợi nhuận của các DN sản xuất tại Ấn Độ.

Tiếp bước các nghiên cứu nước ngoài, Tạ Văn Thắng và cộng sự (2025) đã thực hiện nghiên cứu so sánh giữa mô hình bình phương tối thiểu thông thường (OLS) cùng với các kỹ thuật học máy như Random Forest, ElasticNet và XGBoost để dự báo thu ngân sách nhà nước giai đoạn 1991-2004, khẳng định tính hiệu quả của các mô hình học máy trong quản lý vĩ mô. Kế đến là trong quản trị rủi ro DN, Giang Thị Thu Huyền và cộng sự (2024) đã sử dụng các mô hình máy học Multiple Regression Model (MRL), LASSO, GAM (Generalized Additive Model), Random Forests, Gradient Boosting Regression Trees, Neural Network cùng với mô hình OLS truyền thống để dự báo mức độ căng thẳng tài chính cho 646 DN niêm yết tại Việt Nam trong giai đoạn 2012-2022. Ngoài ra, Đỗ Thị Vân Trang và cộng sự (2024) đã mở rộng phạm vi phân tích sức mạnh của máy học để nghiên cứu ảnh hưởng của các quyết định tài chính đến giá trị DN niêm yết trên hai sàn HOSE và HNX.

Tổng kết các nghiên cứu trên, có thể thấy một sự phát triển từ việc sử dụng các mô hình kinh tế lượng dựa trên các giả định chặt chẽ (như OLS, FEM, REM, FGLS) sang các kỹ thuật học máy linh hoạt, có khả năng học hỏi từ dữ liệu. Tuy nhiên, tính giải thích vẫn chưa rõ ràng do mô hình học máy thường bị coi là "hộp đen", điều này sẽ gây khó khăn trong việc giải thích cơ chế kinh tế đằng sau các biến số.

Mặc dù đã có nhiều nghiên cứu xem xét tác động của nợ vay đến hiệu quả tài chính của DN

nhưng chủ yếu sử dụng phương pháp hồi quy truyền thống với các kết quả không đồng nhất. Do đó, điểm khác biệt trong nghiên cứu này là sử dụng kết hợp giữa mô hình hồi quy truyền thống và phương pháp học máy với giá trị SHAP để củng cố kết quả hồi quy từ đó gia tăng độ tin cậy của kết quả nghiên cứu.

2.2. Cơ sở lý thuyết

Lý thuyết Modigliani-Miller (MM) được Modigliani, F., Miller, M. H. (1958) đưa ra để giải thích về cơ cấu vốn trong DN sẽ không ảnh hưởng đến giá trị của DN, có nghĩa là dù DN có vay nợ nhiều hay ít hay thậm chí không vay thì giá trị DN cũng không bị ảnh hưởng. Tuy nhiên, kết luận này phải đi kèm với một số điều kiện khắt khe là DN hoạt động trong một môi trường không có thuế, không có chi phí giao dịch và các cá nhân cùng với DN có khả năng vay nợ với cùng một lãi suất. Tiếp tục nghiên cứu trong môi trường có thuế thu nhập DN, Modigliani, F., Miller, M. H. (1963) đã đưa ra một khám mới về giá trị của DN là nó sẽ tăng tuyến tính cùng với việc vay nợ do DN sẽ được giảm trừ phần thuế bởi chi phí lãi vay. Chính điều này đã khuyến khích các DN sử dụng nợ vay càng nhiều thì hiệu quả tài chính của DN càng được nâng cao.

Bên cạnh đó, theo lý thuyết tín hiệu, DN có kế hoạch về lợi nhuận thấp sẽ vay nợ với mức thấp, ngược lại DN có kế hoạch về lợi nhuận cao hơn sẽ vay nợ với mức cao hơn. Ross, S.A. và cộng sự (2010) cho rằng DN khi vay nợ sẽ xem xét hợp lý tăng mức độ nợ khi lợi nhuận dự kiến tăng lên. Lúc này nhận định của thị trường là phát ra "tín hiệu" DN có vay nợ thì trạng thái hoạt động sản xuất kinh doanh là tốt. Do đó, trong điều kiện sản xuất kinh doanh bình thường, DN vẫn duy trì khả năng vay mượn để có thể tận dụng những cơ hội tốt sau này.

Theo lý thuyết đại diện của Jensen, M.C., Meckling, W.H. (1976) và mở rộng ra là lý thuyết dòng tiền tự do của Jensen, M.C. (1986) cho rằng các nhà quản lý luôn có động cơ vì lợi ích tiêu dùng cá nhân khi dòng tiền của DN là cao hơn so với mức bình thường. Khi đó nếu dòng tiền này được chi trả cho nợ và lãi vay thì sẽ hạn chế được phần nào hành vi của nhà quản lý. Ngoài ra, dựa vào lý thuyết này, theo Brigham, E.F., & Ehrhardt, M.C. (2009) ủng hộ việc DN vay nợ là để kiểm soát các nhà quản lý và khiến họ cẩn thận với số tiền của cổ đông, cố gắng điều hành công việc tốt hơn, làm tăng giá trị của DN trên thị trường. Tuy nhiên, việc sử dụng vốn vay quá cao có thể làm gia tăng rủi ro hành vi lạm dụng tài sản của DN bởi các nhà quản lý và gia tăng chi phí giám sát, điều này dẫn đến hiệu quả tài chính của DN suy giảm.

Các lý thuyết trên cho thấy nợ vay có ảnh hưởng đến hiệu quả tài chính của DN. Đây là căn cứ để xây dựng mô hình nghiên cứu ở phần sau.

3. Phương pháp nghiên cứu

3.1. Quy trình nghiên cứu

Nghiên cứu này sử dụng phương pháp định lượng để thực hiện các mục tiêu nghiên cứu. Bước đầu tác giả tổng hợp các nghiên cứu trong và ngoài nước để xác định khe hổng nghiên cứu và đề xuất mô hình. Tiếp theo, tác giả thu thập các số liệu liên quan để thực hiện thống kê mô tả và thực hiện hồi quy đa biến với các kỹ thuật POLS, FEM, REM, FGLS. Bên cạnh đó, các phương pháp học máy cũng được ứng dụng để hỗ trợ cho các mô hình kinh tế lượng truyền thống. Cuối cùng là phân tích kết quả từ phương pháp thống kê truyền thống và phương pháp học máy, từ đó đưa ra những kết luận và khuyến nghị phù hợp.

4.2. Mô hình nghiên cứu

Theo các lý thuyết và các nghiên cứu trước có liên quan, mô hình hồi quy tổng quát được đề xuất như sau:

$$ROA_{it} = \beta_0 + \beta_1 LEV_{it} + \beta_2 SIZE_{it} + \beta_3 TANG_{it} + \beta_4 ASSETG_{it} + \beta_5 SALESG_{it} + \varepsilon_{it}$$

Trong đó:

ROA: Biến phụ thuộc, hiệu quả tài chính là tỷ số giữa lợi nhuận sau thuế trên tổng tài sản.

LEV: Biến độc lập, tỷ lệ nợ vay là tỷ số giữa tổng nợ vay trên tổng tài sản của DN.

Các biến kiểm soát gồm:

+ Quy mô DN (SIZE) là kết quả logarit của tổng tài sản DN

+ Tỷ lệ tài sản cố định (TANG) là tỷ số tài sản cố định trên tổng tài sản của DN.

+ Tỷ lệ tăng trưởng tài sản (ASSETG) được tính dựa trên chênh lệch giữa tài sản năm hiện tại và tài sản của năm trước sau đó chia tài sản năm trước.

+ Tỷ lệ tăng trưởng doanh thu (SALESG) được tính dựa vào chênh lệch giữa doanh thu năm hiện tại và doanh thu năm trước sau đó chia cho doanh thu năm trước.

4.3. Phương pháp chọn mẫu, thu thập dữ liệu và công cụ phân tích

Trong nghiên cứu, tác giả sử dụng phương pháp phi xác suất bằng kỹ thuật chọn mẫu thuận tiện vì một số DN không đủ số liệu báo cáo tài chính (BCTC) cần thiết cho việc nghiên cứu. Dữ liệu được lấy từ các BCTC đã được kiểm toán của các DN niêm yết trên HOSE giai đoạn từ năm 2015 đến năm 2024. Kết quả có 274 DN đáp ứng các điều kiện đưa ra như trên, khi đó có 2.740 quan sát được sử dụng phục vụ cho việc phân tích.

Tác giả sử dụng phần mềm Stata để tiến hành việc phân tích hồi quy truyền thống và ngôn ngữ lập trình Python cho việc phân tích hồi quy bằng phương pháp học máy. Trước tiên, tác giả sử dụng phương pháp phân tích tương quan giữa biến phụ thuộc với các biến độc lập và giữa các biến độc lập với nhau. Kế đến là tác giả sử dụng phương pháp phân tích hồi quy đa biến giữa biến độc lập và biến phụ thuộc, lựa chọn cách thức phương pháp hồi quy phù hợp (POLS, FEM, REM, FGLS) và thực hiện việc kiểm định các khiếm khuyết của mô hình hồi quy bằng phần mềm Stata.

Bên cạnh đó, ngôn ngữ lập trình Python và các gói thư viện hỗ trợ được sử dụng để phân tích kết quả hồi quy bằng các phương pháp học máy như: Phương pháp hồi quy tuyến tính (Linear Regression - LN), phương pháp cây quyết định (Decision Trees - DT), phương pháp rừng ngẫu nhiên (Random Forests - RF), phương pháp tăng cường độ dốc (Gradient Boost - GB), phương pháp tăng cường độ dốc cực đại (eXtreme Gradient Boosting - XGBoost), phương pháp máy tăng cường độ dốc mức độ nhẹ (Light Gradient Boosting Machine - LightGBM).

Để đánh giá hiệu quả của các mô hình học máy, các chỉ số Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-Squared được sử dụng để đo lường các sai số trong việc dự báo cũng như đưa ra mức độ giải thích của mô hình đối với biến phụ thuộc, từ đó cung cấp cơ sở để giúp so sánh giữa các phương pháp học máy với nhau. Phương pháp học máy thích hợp nhất là phương pháp có các chỉ số MAE, MSE, RMSE nhỏ nhất và R-Squared lớn nhất.

Các phương pháp học máy thường là các "hộp đen" và khó khăn trong việc diễn giải các kết quả. Vì vậy, theo Lundberg, S. M., Lee, S. I. (2017) giá trị SHAP (SHapley Additive exPlanation) được sử dụng để giải thích kết quả đầu ra của một mô hình học máy. Với sự trợ giúp của giá trị SHAP, kết quả của các mô hình học máy sẽ được giải thích thông qua các đồ thị cùng với các biến giải thích một cách trực quan.

4. Kết quả nghiên cứu và bàn luận

4.1. Kết quả theo các phương pháp hồi quy truyền thống

Trước tiên là phân tích tương quan để xác định có hay không mối quan hệ tuyến tính giữa biến phụ thuộc (ROA) với biến độc lập (tỷ lệ nợ vay của DN) và các biến kiểm soát (quy mô DN, tỷ lệ tài sản cố định, tỷ lệ tăng trưởng tài sản, tỷ lệ tăng trưởng doanh thu). Kết quả được mô tả chi tiết theo bảng 1.

Bảng 1: Ma trận hệ số tương quan

	ROA	LEV	SIZE	TANG	ASSETG	SALESG
ROA	1,0000					
LEV	-0,3974	1,0000				
SIZE	-0,0886	0,3187	1,0000			
TANG	0,0324	-0,0610	0,0665	1,0000		
ASSETG	-0,0002	0,0505	0,0510	0,0016	1,0000	
SALESG	-0,0121	0,0345	0,0004	-0,0135	0,0084	1,0000

Nguồn: kết quả phân tích từ Stata

Theo kết quả phân tích tương quan, hiệu quả tài chính DN (ROA) có tương quan ngược chiều với tỷ lệ nợ vay trên tổng tài sản (LEV), quy mô DN (SIZE), tỷ lệ tăng trưởng doanh thu (SALESG) nhưng lại có tương quan cùng chiều với tỷ lệ tài sản cố định (TANG). Ngoài ra, kết quả còn xác định được mối tương quan giữa các biến độc lập với nhau và cho thấy các giá trị tuyệt đối của chúng đều nhỏ hơn 0,8. Theo Hair và cộng sự (2006), nếu có hệ số tương quan nào lớn hơn 0,8 thì sẽ tồn tại vấn đề đa cộng tuyến trong mô hình.

Kế đến là thực hiện phân tích hồi quy bằng các kỹ thuật truyền thống POLS, FEM, REM. Sau đó là thực hiện việc kiểm định để lựa chọn mô hình phù hợp, kiểm định các khiếm khuyết của mô hình và kết quả được tổng hợp theo như bảng 2.

Bảng 2 - Kết quả các mô hình hồi quy theo phương pháp truyền thống

	POLS	FEM	REM	FGLS
LEV	-0,1686	-0,1136	-0,1376	-0,1360
	0,0000	***0,0000***	***0,0000***	***0,0000***
SIZE	0,0024	-0,0140	-0,0042	0,0009

	POLS	FEM	REM	FGLS
	0,0270	***0,0000***	***0,0290***	0,1470
TANG	0,0018	-0,0319	-0,0142	-0,0044
	0,7960	***0,0170***	0,1830	0,2330
ASSETG	0,0010	0,0013	0,0011	0,0000
	0,2950	0,0790	0,1190	0,9840
SALESG	0,0000	0,0000	0,0000	0,0000
	0,9120	0,6520	0,6810	0,5360
_cons	0,0786	0,5262	0,2541	0,1011
	0,0080	***0,0000***	***0,0000***	***0,0000***
R2(Adj/Overall)	15,84%	9,12%	14,47%	
F test	Prob > F = 0,0000			
Hausman test		Prob > chi2 = 0,0000		
B-P test	Prob > chibar2 = 0,000			
Chọn	FEM			

Nguồn: kết quả phân tích từ Stata

Theo bảng 2, kết quả mô hình là FEM được lựa chọn. Cùng với kết quả kiểm định sự đa cộng tuyến bằng hệ số phóng đại phương sai từ bảng 3, kết quả kiểm phương sai sai số thay đổi và tính chất tự tương quan từ bảng 4, mô hình FGLS được lựa chọn để khắc phục các tính chất tự tương quan và phương sai sai số thay đổi trong mô hình FEM.

Bảng 3 - Hệ số phóng đại phương sai VIF

	VIF	1/VIF
LEV	1,12	0,8894
SIZE	1,12	0,8897
TANG	1,01	0,9879
ASSETG	1,00	0,9960
SALESG	1,00	0,9985
Mean VIF	1,05	

Bảng 4 - Kiểm định phương sai và tự tương quan

*** Kiểm định phương sai thay đổi (Wald)**	
Prob > chi2	
Phương sai sai số thay đổi?	Có
*** Kiểm định tính tự tương quan (Wooldridge)**	
Prob > F	
Tự tương quan?	Có

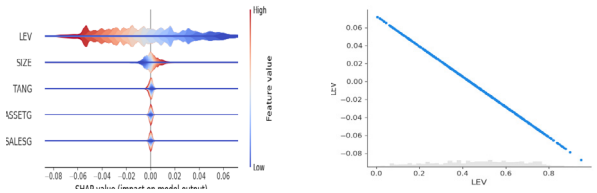
Nguồn: Kết quả phân tích từ Stata

Theo kết quả từ bảng 2, mô hình FGLS cho kết luận là tỷ lệ nợ vay trên tổng tài sản có tác động ngược chiều đến hiệu quả tài chính ROA.

4.2. Kết quả theo các phương pháp học máy:

Với sự hỗ trợ của ngôn ngữ lập trình Python, các phương pháp học máy được sử dụng trên tập dữ liệu huấn luyện là 80% và tập dữ liệu kiểm thử là 20%. Khi đó các kết quả của các phương pháp học máy thông qua công cụ giải thích SHAP lần lượt là các hình sau:

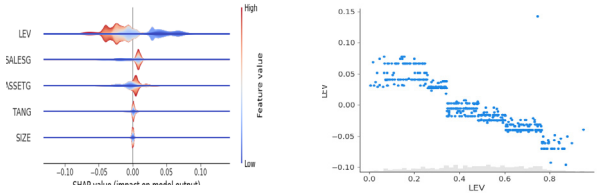
Hình 1 - Mô hình Linear Regression (LR)



Nguồn: kết quả hình ảnh từ Python

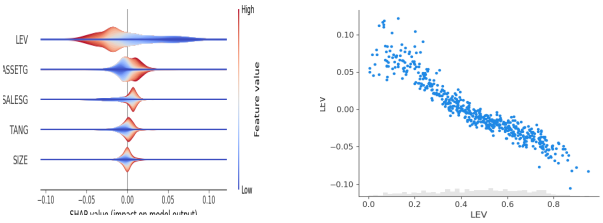
Theo hình 1, tỷ lệ nợ vay trên tài sản (LEV) là biến có đóng góp lớn nhất trong mô hình LR (các biến kiểm soát được sắp xếp theo thứ tự quan trọng từ cao xuống thấp theo trục đứng bên trái), có tác động trái chiều một cách rõ nét đến hiệu quả tài chính ROA (màu sắc của trục đứng bên phải từ thấp đến cao).

Hình 2 - Mô hình Decison Trees



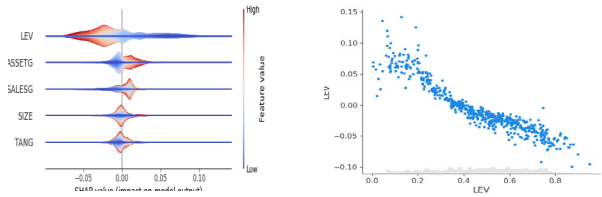
Nguồn: kết quả hình ảnh từ Python

Hình 3 - Mô hình Random Forests



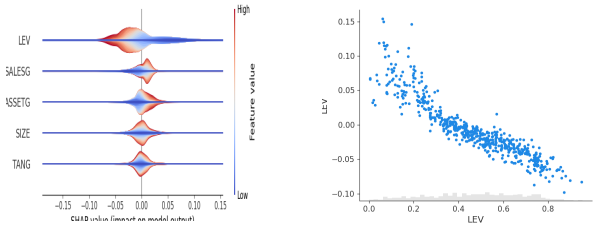
Nguồn: kết quả hình ảnh từ Python

Hình 4 - Mô hình Gradient Boost



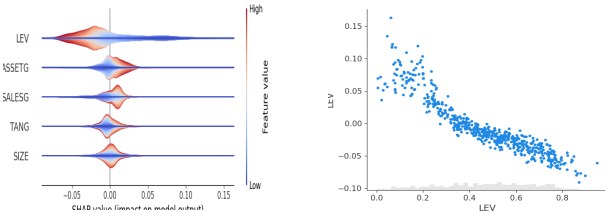
Nguồn: kết quả hình ảnh từ Python

Hình 5 - Mô hình XGBoost



Nguồn: kết quả hình ảnh từ Python

Hình 6 - Mô hình LightGBM



Nguồn: Kết quả hình ảnh từ Python

Tương tự với các hình 2 (DT), hình 3 (RF), hình 4 (GB), hình 5 (XGBoost), hình 6 (LightGBM) cũng cho kết quả là tỷ lệ nợ vay trên tổng tài sản (LEV) tác động trái chiều đến hiệu quả tài chính DN ROA là tốt nhất trong các biến trong mô hình.

Tiếp theo là việc đánh giá hiệu quả của các phương pháp học máy (LR, DT, RF, GB, XGBoost, LightGBM) với các chỉ số MAE, MSE, RMSE, R-Squared.

Bảng 5: Các chỉ số đánh giá hiệu quả của các phương pháp học máy

Mô hình	MAE	MSE	RMSE	R-Squared
Linear Regression	0,0443	0,0044	0,0665	0,2217
Decision Trees	0,0435	0,0042	0,0649	0,2586
Random Forests	0,0410	***0,0035***	0,0594	0,3789
Gradient Boost	0,0417	0,0036	0,0602	0,3621
XGBoost	0,0438	0,0039	0,0625	0,3131
LightGBM	***0,0408***	***0,0035***	***0,0589***	***0,3889***

Nguồn: kết quả phân tích từ Python

Theo kết quả từ bảng 5, phương pháp học máy LightGBM với các chỉ số MAE, MSE, RMSE là nhỏ nhất và chỉ số R-Squared lớn nhất trong các phương pháp học máy. Do vậy, mô hình LightGBM là mô hình tốt nhất thể hiện tỷ lệ nợ có tác động trái chiều đến hiệu quả tài chính ROA của DN.

4.3. Bàn luận

Với hai phương pháp tiếp cận khác nhau là mô hình hồi quy truyền thống và phương pháp học máy đều cho cùng kết quả nghiên cứu là tỷ lệ nợ trên tài sản có tác động ngược chiều đến hiệu quả tài chính của các DN niêm yết trên HOSE. Kết quả này cũng phù hợp với các nghiên cứu của Đinh Văn Sơn, Trần Thị Phương Thảo (2017); Chi, N.T.D và cộng sự (2019); Bùi Vĩnh Thanh (2022). Ngoài ra, ở các quốc gia như Bangladesh, Vương quốc Anh, Romani, Trung Quốc, Đài Loan lần lượt các nghiên cứu của Hasan, M.B. và cộng sự (2014); Abeywardhana, D.K.Y. (2015); Nenu, E.A., Vintilă, G.,& Gherghina, S.C. (2018); Ye, Z. (2021); Vu Thi, A. H., & Phung, T. D. (2021) đều cùng có kết quả là tỷ lệ nợ trên tài sản tác động ngược chiều đến hiệu quả tài chính ROA.

5. Kết luận và khuyến nghị

Với các phương pháp mô hình hồi quy (OLS, FEM, REM, FGLS) và sự hỗ trợ của các phương pháp học máy (LR, DT, RF, GB, XGBoost, LightGBM), cùng cho kết quả là tỷ lệ nợ vay trên tài sản có tác động ngược chiều đến hiệu quả tài chính của DN niêm yết trên HOSE. Nói cách khác, khi tỷ lệ nợ trong cơ cấu vốn của DN càng cao thì hiệu quả tài chính càng thấp và ngược lại. Điều này thể hiện rằng DN hoạt động sản xuất kinh doanh thông qua vốn vay quá nhiều sẽ phát sinh nhiều chi phí lãi và giảm hiệu quả hoạt động của DN. Kết quả nghiên cứu này được củng cố thông qua sự kết giữa phương pháp hồi quy truyền thống và phương pháp học máy. Từ kết quả nghiên cứu, tác giả đề xuất một số khuyến nghị như sau:

Một là, các DN niêm yết cần chú ý đến các hình thức huy động vốn khác ngoài việc đi vay để bù đắp khoản vốn thiếu hụt trong hoạt động sản xuất kinh doanh.

Hai là, khi thực hiện việc đầu tư trên thị trường chứng khoán, các nhà đầu tư cần lưu ý đến khả năng sinh lời ROA, cơ cấu nợ vay của DN để có dự báo chính xác xu hướng và triển vọng phát triển của DN. Ngoài ra, nhà đầu tư cần kết hợp với việc các phân tích môi trường vĩ mô trong nước cũng như trên thế giới và phân tích kỹ thuật để có những quyết định hợp lý.

Tài liệu tham khảo:

Bùi Thị Ngọc, Nguyễn Thị Thanh Phương (2023). Tác động của cơ cấu vốn đến hiệu quả kinh doanh của các công ty niêm yết. Tạp chí Kinh tế & Phát triển, 309.

Chakri, P., Pratap, S., Lakshay, & Gouda, S.K. (2023). An exploratory data analysis approach for analyzing financial accounting data using machine learning. Decisions Analytics Journal, 7(2023)

Đinh Văn Sơn, Trần Thị Phương Thảo (2017). Ảnh hưởng của cơ cấu vốn đến hiệu quả kinh doanh của các công ty cổ phần niêm yết trên thị trường chứng khoán Việt Nam. Tạp chí Khoa học Thương mại, 114, 2-14

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems (NIPS), 30, 3146-3154

Lundberg, S. M., Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NIPS), 30

Saha, D., Young, T.M., & Thacker, J. (2023). Predicting firm performance and size using machine learning with a Bayesian perspective. Machine Learning with Applications, 11(2023)

Vuong, N.B., Vu, T.T.Q. & Mitra, P. (2017). Impact of Capital Structure on Firm's Financial Performance: Evidence from United Kingdom. Journal of Finance & Economics Research, 2(1), 18-31. https://doi.org/10.20547/jfer1702102

Ye, Z. (2021). Capital Structure and Corporate Performance: A Case of China. BCP Business & Management, 15(2021), 403-418