

Trang chủ/Diễn đàn khoa học/Nghiên cứu - Trao đổi

Nghiên cứu ứng dụng phân tích dữ liệu trong quản trị rủi ro tài chính tại các công ty niêm yết trên sàn chứng khoán Việt Nam

15:18 | 31/05/2024

EFR Bài viết xác định các yếu tố có tác động lớn tới rủi ro tài chính tại các doanh nghiệp sản xuất niêm yết trên sàn chứng khoán Việt Nam. Từ đó, đề xuất một số giải pháp và chính sách cho các doanh nghiệp.

TS. Nguyễn Thị Thanh Loan, Nguyễn Ngọc Hân, Nguyễn Thị Khánh Chúc,

Nguyễn Đình Đạt, Nguyễn Như Quỳnh, Trương Văn Trọng

Trường Đại học Công nghiệp Hà Nội

Email: nguyenthithanhloan@hau.edu.vn

Tóm tắt

Bài viết xác định các yếu tố có tác động lớn tới rủi ro tài chính tại các doanh nghiệp sản xuất niêm yết trên sàn chứng khoán Việt Nam dựa trên phương pháp chọn biến LASSO và xây dựng mô hình học máy dự báo rủi ro tài chính. Kết quả nghiên cứu cho thấy, mô hình LASSO đã xác định 5 chỉ số tài chính ảnh hưởng đến rủi ro tài chính mà các doanh nghiệp rủi ro gặp phải trong suốt 5 năm (2017-2021) gồm: Hệ số khả năng thanh toán ngắn hạn (Tài sản ngắn hạn/Nợ ngắn hạn); Số vòng quay tổng tài sản (Doanh thu thuần/Tổng tài sản); ROA (Lợi nhuận sau thuế/Tổng tài sản); Tỷ lệ Nợ dài hạn trên Tổng tài sản; Tỷ lệ Nợ ngắn hạn trên Tổng nợ phải trả. Mạng nơon nhân tạo (Artificial Neural Network - ANN) cho thấy, Hiệu quả cao nhất trong việc dự đoán tình hình rủi ro tài chính. Từ đó, đề xuất một số giải pháp và chính sách cho các doanh nghiệp.

Từ khóa: quản trị rủi ro tài chính, học máy, phân tích dữ liệu, phân tích dự đoán

Summary

By LASSO variable selection method and machine learning to predict financial risk, the paper aims to discover factors that have a major impact on financial risk in manufacturing enterprises listed on the Vietnamese stock exchange. Research results indicates that the LASSO model identifies 5 financial indicators affecting financial risks which risky businesses faced during 5 years (2017-2021), they are Short-term solvency ratio (Short-term assets/Short-term liabilities); Total asset turnover (Net revenue/Total assets); ROA (Profit after tax/Total assets); Long-term debt to total asset ratio; Short-term debt to total debt ratio. Artificial Neural Network (ANN) shows the highest efficiency in predicting financial risk. From those finding, the paper proposes some solutions and policies for businesses.

Keywords: financial risk management, machine learning, data analysis, predictive analysis

GIỚI THIỆU

Trong nền kinh tế hiện đại, các doanh nghiệp sản xuất đóng vai trò quan trọng trong tiến trình công nghiệp hóa, hiện đại hóa của Việt Nam. Hoạt động kinh doanh trong môi trường cạnh tranh gay gắt và môi trường quốc tế đầy biến động, các doanh nghiệp sản xuất gặp không ít rủi ro, đặc biệt là rủi ro tài chính. Công tác quản trị rủi ro tài chính trong các doanh nghiệp Việt Nam có những chuyển biến tích cực, tuy nhiên mới thể hiện ở bề rộng, chưa đi vào chiều sâu. Công tác quản trị rủi ro tài chính ở các doanh nghiệp còn yếu ở khâu đo lường và đánh giá tác động của rủi ro tài chính. Khi xem xét rủi ro tài chính, để có được nhiều thông tin hữu ích hơn, chúng ta cũng cần sử dụng phương pháp tiếp cận nghiệp vụ thông minh.

Trong thời đại công nghiệp 4.0, thông tin và dữ liệu được coi là tài nguyên quý giá nên phân tích thông tin và dữ liệu đóng vai trò quan trọng trong việc quản lý rủi ro hiệu quả (Vuong, 2023). Bằng cách cung cấp thông tin, dữ liệu một cách chính xác và đầy đủ, doanh nghiệp có thể sử dụng công nghệ thông tin, cũng như các mô hình quản trị rủi ro hiện đại nhằm nhận diện, kiểm soát rủi ro tốt hơn.

TỔNG QUAN NGHIÊN CỨU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Tổng quan nghiên cứu

Huang và cộng sự (2021) đã sử dụng 3 thuật toán Random forest (RF), Support vector machine (SVM) và AdaBoost để đánh giá rủi ro doanh nghiệp. Kết quả nghiên cứu đã cho thấy, 3 loại thuật toán học máy đều đạt hiệu quả đánh giá rủi ro và có độ tin cậy cao trong việc đánh giá rủi ro tài chính của doanh nghiệp. Murugan (2023) đã tiến hành phân tích và sử dụng

dữ liệu trên quy mô lớn và sử dụng các chiến lược học máy: K-nearest neighbor (KNN), logistic regression (LR) và XGBoost để dự đoán khả năng xảy ra mất nợ và các sự cố liên quan trong việc cho vay. Nghiên cứu này cũng đã đề cập đến sự kết hợp các công nghệ mới (IOT) và machine learning để triển khai các giải pháp quản lý rủi ro. Điều này cho thấy tiềm năng của các công nghệ mới trong quản lý và dự báo các rủi ro tài chính.

Đào Trọng Thịnh và Doãn Văn Toàn (2016) đã ứng dụng mạng nơ ron thần kinh để phân loại khách hàng có thể cung cấp tín dụng, điều này giúp tăng khả năng quản lý tín dụng trong hoạt động cung cấp tín dụng cho khách hàng của các ngân hàng. Đoàn Khánh Hưng và Trần Thị Hiền (2019) chỉ ra rằng, cấu trúc tài chính có ảnh hưởng đến rủi ro tài chính, thông qua sử dụng nguồn tiền một cách hiệu quả giúp giảm thiểu rủi ro cho doanh nghiệp.

Nghiên cứu của Lê Hải Trung và Trương Thị Thùy Dương (2023) áp dụng mô hình LASSO để chọn ra các chỉ số tài chính có ảnh hưởng trực tiếp tới khả năng phá sản của doanh nghiệp đó. Kết quả nghiên cứu cho thấy, mô hình LASSO giúp nâng cao hiệu suất dự báo phá sản của các doanh nghiệp niêm yết trên sàn chứng khoán.

Phương pháp nghiên cứu

Nghiên cứu này sử dụng mô hình điểm số Z-score để tính toán khả năng phá sản của các doanh nghiệp, từ đó xây dựng biến phụ thuộc phân loại doanh nghiệp với 2 thuộc tính rủi ro và không rủi ro. Tiếp đến là sử dụng mô hình LASSO để chọn ra được 5 biến độc lập quan trọng, có ảnh hưởng lớn tới biến phụ thuộc trong tổng số 29 biến độc lập mà nhóm đã đưa ra từ bộ dữ liệu đã thu thập và mô hình mạng Nơ-ron thần kinh để dự báo phân loại doanh nghiệp thuộc loại rủi ro hay không rủi ro và sau đó dùng ma trận nhầm lẫn để đánh giá hiệu suất của mô hình dự báo.

Mô hình Z-score xây dựng biến phụ thuộc

Z-score thể hiện khả năng phá sản của một công ty trong tài chính, gọi là hệ số nguy cơ phá sản. Giá trị Z-score là một giá trị bằng số được sử dụng trong thống kê để xác định mối quan hệ của giá trị một điểm với giá trị trung bình bộ dữ liệu, hay độ lệch chuẩn so với giá trị trung bình. Các chỉ tiêu sử dụng trong công thức tính toán đều dễ dàng thu thập được trên báo cáo tài chính hàng năm của doanh nghiệp. Mô hình điểm số Z-score được Edward Altman đưa ra lần đầu tiên vào năm 1968, sau đó, các biến thể của công thức được Altman phát triển cho các trường hợp ngành khác nhau.

Mô hình Z-score trong ngành sản xuất được biểu diễn như sau:

$$Z = 1,2X1 + 1,4X2 + 3,3X3 + 0,6X4 + 1,0X5$$

Trong đó: X1: Tỷ số vốn lưu động trên tổng tài sản (Working Capitals/Total Assets); X2: Tỷ số lợi nhuận giữ lại trên tổng tài sản (Retained Earnings/Total Assets); X3: Tỷ số lợi nhuận trước lãi vay và thuế trên tổng tài sản (EBIT/Total Assets); X4: Giá trị thị trường của vốn chủ sở hữu trên giá trị sổ sách của tổng nợ (Market Value of Total Equity/Book Values of Total Liabilities); X5: Tỷ số doanh số trên tổng tài sản (Sales/Total Assets).

Cách phân loại doanh nghiệp:

- + Nếu Z >= 1,81: Doanh nghiệp nằm trong vùng an toàn, chưa có nguy cơ phá sản
- + Nếu Z < 1,81: Doanh nghiệp nằm trong vùng nguy hiểm, nguy cơ phá sản cao

Mô hình LASSO lựa chọn các biến độc lập

Tibshirani (1996) đã giới thiệu và phát triển mô hình hàm phạt LASSO để lựa chọn các biến giải thích có độ tương quan cao tới biến được giải thích trong mô hình dự báo.

Mô hình LASSO có dạng:

$$\sum_{i=1}^n \left(\left(-Y_{i,t} \left(\beta_0 + \sum_{k=1}^p x_{i,t-1,k} \beta_k \right) \right) + \log(1 + e^{\beta_0 + \sum_{k=1}^p x_{i,t-1,k} \beta_k}) \right)$$

Với điều kiện:

$$\sum_{k=1}^p \omega_k |\beta_k| < \lambda$$

Trong đó: $y_{i,t}$ là biến nhị phân biểu diễn tình trạng của doanh nghiệp i tại thời điểm t ; $x_{i,t-1,k}$ là các chỉ tiêu tài chính thứ k của doanh nghiệp i tại thời điểm $t-1$; n là số doanh nghiệp trong mẫu thống kê.

Hàm điều kiện LASSO có dạng:

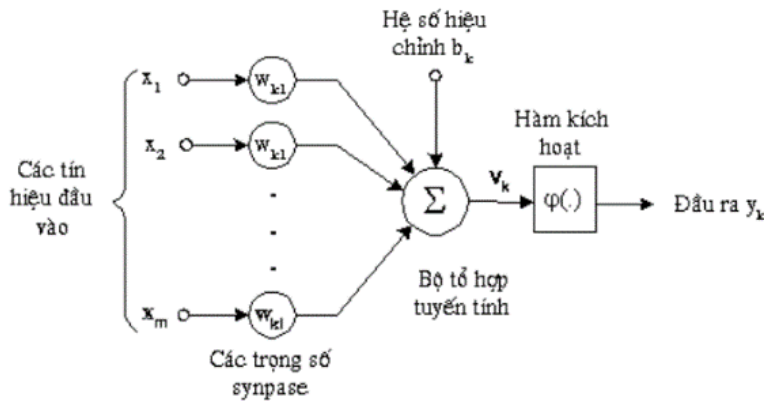
$$\sum_{k=1}^p \omega_k |\beta_k| < \lambda$$

thể hiện hàm phạt (penalty function) với điều kiện về giá trị ước lượng của các tham số bị giới hạn bởi hệ số chuyển λ . Giá trị λ càng nhỏ thì số biến giải thích được giữ lại mô hình dự báo càng ít (Lê Hải Trung và cộng sự, 2023). Hàm phạt này còn được gọi là hàm phạt "L1". LASSO sẽ tự động đưa ước lượng tham số của biến giải thích không có ý nghĩa về 0 và giảm giá trị ước lượng của các biến giải thích ít có nghĩa về giá trị nhỏ. Một ưu điểm nữa của phương pháp LASSO là xử lý vấn đề đa cộng tuyến giữa các biến giải thích với nhau. Điều này có giúp ích rất lớn đến mô hình dự báo, bởi lẽ việc dự báo rủi ro tài chính thường sử dụng nhiều biến số tài chính có tỷ lệ tương quan cao (Tian và cộng sự, 2015).

Mô hình mạng nơ ron nhân tạo dự báo doanh nghiệp gặp rủi ro tài chính

Mạng nơ ron nhân tạo (ANN) là mô hình tính toán mô phỏng các chức năng mạng sinh học thần kinh của con người. Mỗi nơ ron là một đơn vị xử lý thông tin và là thành phần cơ bản của một mạng nơ ron. Cấu trúc của một nơ ron được mô tả như Hình 1.

Hình 1: Mạng nơ ron nhân tạo



Nguồn: Nguyễn Chính Kiên (2017)

Các thành phần cơ bản của một nơ ron nhân tạo bao gồm:

- Tập các đầu vào: Là các tín hiệu vào (input signals) của nơ ron, các tín hiệu này thường được đưa vào dưới dạng một vec-tơ m chiều.
- Tập các liên kết: Mỗi liên kết được thể hiện bởi một trọng số (gọi là trọng số liên kết – Synaptic weight). Trọng số liên kết giữa tín hiệu vào thứ j với nơ ron k thường được kí hiệu là w_{jk} . Thông thường, các trọng số này được khởi tạo một cách ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học mạng.
- Bộ tổng (Summing function): Thường dùng để tính tổng của tích các đầu vào với trọng số liên kết của nó.
- Ngưỡng (còn gọi là một độ lệch - bias): Ngưỡng này thường được đưa vào như một thành phần của hàm truyền.
- Hàm truyền (Transfer function) – còn gọi là Hàm kích hoạt (Activation function): Hàm này được dùng để giới hạn phạm vi đầu ra của mỗi nơ ron. Nó nhận đầu vào là kết quả của hàm tổng và ngưỡng đã cho. Thông thường, phạm vi đầu ra của mỗi nơ ron được giới hạn trong đoạn [0,1] hoặc [-1, 1]. Các hàm truyền rất đa dạng được liệt kê trong bảng 1.1, có thể là các hàm tuyến tính hoặc phi tuyến. Việc lựa chọn hàm truyền nào là tùy thuộc vào từng bài toán và kinh nghiệm của người thiết kế mạng.
- Đầu ra: Là tín hiệu đầu ra của một nơ ron, với mỗi nơ ron sẽ có tối đa là một đầu ra.

Nơ ron nhân tạo cũng giống như các nơ ron thần kinh sinh học, nhận các tín hiệu đầu vào, xử lý bằng cách nhân các tín hiệu này với trọng số liên kết, tính tổng các tích thu được rồi gửi kết quả tới hàm truyền, kết quả của hàm truyền chính là tín hiệu đầu ra.

$$Y = (X_1w_1 + b_1) + (X_2w_2 + b_2) + \dots + (X_nw_n + b_n)$$

Mỗi nơon đơn lẻ có chức năng xử lý thông tin nhất định, nhưng khả năng tính toán nơon chủ yếu có được nhờ sự liên kết có cấu trúc thống nhất các nơon có thể xử lý những tính toán phức tạp và cho ra một kết quả chính xác.

Ma trận nhầm lẫn đánh giá hiệu suất dự báo

Trong nghiên cứu này, chúng tôi sử dụng phương pháp so sánh giữa các mô hình dự báo khả năng gặp rủi ro tài chính dựa trên ma trận nhầm lẫn (Confusion matrix). Đây là phương pháp đánh giá hiệu suất phân loại các quan sát vào hai lớp rủi ro hay không rủi ro bởi độ chính xác và các mức độ bao quát của phép phân loại. Lớp rủi ro nhận giá trị 1, không rủi ro nhận giá trị 0.

Ma trận nhầm lẫn gồm các chỉ số sau (Hình 2): TP (true positive) là số dự đoán tích cực, nghĩa là số lượng doanh nghiệp gặp rủi ro tài chính được dự báo đúng là gặp rủi ro tài chính; TN (true negative) là số lượng doanh nghiệp không gặp rủi ro tài chính được dự báo không gặp rủi ro tài chính, FP (false positive) (dương tính giả) là số lượng các doanh nghiệp không gặp rủi ro tài chính nhưng dự báo gặp rủi ro tài chính, FN (false negative) (âm tính giả) là số lượng doanh nghiệp gặp rủi ro tài chính nhưng được dự báo không gặp rủi ro tài chính.

Hình 2: Ma trận nhầm lẫn

		Giá trị thật	
		1	0
Kết quả dự báo	1	TP	FP
	0	FN	TN

Nguồn: Shrivastava và các cộng sự (2020)

Độ chính xác của mô hình là tỷ lệ dự báo đúng, được tính theo công thức sau:

$$Accuracy = (TP + TN) / (TP + FP + TN + FN)$$

Vì vậy để nâng cao hiệu quả dự báo của mô hình, ngoài độ chính xác của mô hình, hai tiêu chí sau được sử dụng để đánh giá hiệu quả dự báo, Precision và Recall

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

Precision cho biết tỷ lệ dự đoán doanh nghiệp gặp rủi ro tài chính thực sự là gặp rủi ro tài chính, Recall cho biết tỷ lệ dự báo đúng doanh nghiệp gặp rủi ro tài chính trên tổng doanh nghiệp gặp rủi ro tài chính.

Dữ liệu và biến số

Dữ liệu: Các doanh nghiệp trong nghiên cứu bao gồm các doanh nghiệp sản xuất được niêm yết trên thị trường chứng khoán của Việt Nam. Các chỉ tiêu tài chính sử dụng để dự báo nguy cơ rủi ro tài chính của các doanh nghiệp này được tính toán từ các chỉ số, chỉ tiêu trong các báo cáo tài chính công khai đã được kiểm toán (bảng cân đối kế toán, báo cáo lưu chuyển tiền tệ, báo cáo kết quả kinh doanh) tại thời điểm cuối năm của các doanh nghiệp sản xuất được niêm yết trên thị trường chứng khoán Việt Nam từ năm 2017 đến hết năm 2021 tổng cộng gồm có 2.152 quan sát.

Biến phụ thuộc: Nghiên cứu sử dụng điểm z-score (Altman, 1968) để phân loại doanh nghiệp thuộc loại rủi ro hay không rủi ro. Biến phụ thuộc (y) gán giá trị là 1 nếu điểm Z-score < 1,81 (doanh nghiệp gặp rủi ro tài chính), biến y được gán giá trị là 0 nếu điểm Z-score ≥ 1,81 (doanh nghiệp an toàn tài chính). Việc phân loại như vậy cũng phù hợp với những kết quả nghiên cứu trên. Kết quả phân nhóm 2.152 quan sát có 564 quan sát thuộc nhóm có nguy cơ rủi ro tài chính và 1.588 quan sát không có nguy cơ rủi ro tài chính.

Biến độc lập: Dựa trên nguồn số liệu hiện có và những gợi ý từ các công trình nghiên cứu của Zeytinoglu và cộng sự (2013), Valaskova và cộng sự (2018)...., cũng như từ thực tế hoạt động của các doanh nghiệp, nghiên cứu đã xây dựng, lựa chọn 29 biến số (Bảng 1).

Bảng 1: Các biến có ảnh hưởng tới rủi ro tài chính

Biến	Diễn giải	Biến	Diễn giải
X1	Tài sản ngắn hạn/Nợ ngắn hạn	X16	Doanh thu/Vốn cổ phần
X2	(Tài sản ngắn – hàng tồn kho)/Nợ ngắn hạn	X17	Nợ ngắn hạn/Tổng tài sản
X3	Doanh thu/Tồn kho	X18	Nợ dài hạn/Tổng tài sản
X4	Tài sản phải thu/(Doanh thu/365)	X19	Lợi nhuận trước thuế/Vốn cổ phần
X5	Doanh thu/Tài sản cố định	X20	Vốn hóa thị trường
X6	Doanh thu/Tổng tài sản	X21	Hàng tồn kho/Bán hàng

X7	Tổng nợ/Tổng tài sản	X22	Hệ số vòng quay tài sản hàng tồn kho
X8	Vốn chủ sở hữu/Tổng tài sản	X23	LN sau thuế/Doanh thu thuần
X9	Tổng nợ/Vốn cổ phần	X24	Giá vốn hàng bán/Khoản phải thu
X10	ROS	X25	Tài sản cố định/tổng tài sản
X11	ROA	X26	Nợ ngắn hạn/Nợ phải trả
X12	ROE	X27	Nợ phải trả/Tổng tài sản
X13	Tiền và tương đương tiền/Nợ ngắn hạn	X28	Log (Tổng tài sản)
X14	(Tài sản ngắn hạn – Nợ ngắn hạn)/Tổng tài sản	X29	Vốn lưu động/Tổng tài sản
X15	Phải thu ngắn hạn/Tài sản ngắn hạn		

Nguồn: Tổng hợp và đề xuất của nhóm tác giả

KẾT QUẢ NGHIÊN CỨU

Trong dự báo rủi ro tài chính đối với các doanh nghiệp sản xuất niêm yết trên sàn chứng khoán tại Việt Nam, chúng tôi đã chia tập dữ liệu thành 2 tập: tập huấn luyện (train) gồm 4 năm từ 2017-2020 và tập kiểm tra (test) là dữ liệu năm 2021, trong đó tập huấn luyện dùng để máy học và tập kiểm tra để kiểm tra khả năng học của máy. Trong một số trường hợp để tránh cho dữ liệu xảy ra hiện tượng quá khớp (Overfitting), chúng tôi đã sử dụng phương pháp chọn lọc dựa trên mô hình LASSO để lựa chọn ra những biến có ảnh hưởng đến tài chính doanh nghiệp và có khả năng gây ra rủi ro cao đến tài chính doanh nghiệp. Cụ thể, chúng tôi thực hiện chia quy trình nghiên cứu thành ba phần:

Áp dụng mô hình LASSO

Nhóm sử dụng hồi quy LASSO tìm các biến giải thích quan trọng. Sau khi sử dụng hồi quy LASSO, nhóm đã thu được kết quả các biến chọn được là X1, X6, X11, X18, X26. Các biến đặc trưng cho Tài sản ngắn hạn/Nợ ngắn hạn (X1), Doanh thu/Tổng tài sản (X6), ROA (Lợi nhuận ròng/Tổng tài sản) (X11), Nợ dài hạn/Tổng tài sản (X18), Nợ ngắn hạn/Tổng tài sản (X26). Kết quả cho thấy, các biến thể hiện: Khả năng thanh toán, Hiệu quả tài chính, Khả năng tạo lợi nhuận, Tỷ lệ nợ và Đòn bẩy tài chính của doanh nghiệp giúp dự báo, cũng như phân biệt các doanh nghiệp có khả năng đối mặt với rủi ro tài chính cao hay thấp.

Lựa chọn mô hình học máy

Nhóm chúng tôi thực hiện lại việc dự báo các doanh nghiệp gặp rủi ro tài chính với các mô hình học máy sử dụng các biến tài chính đã được lọc ra nhờ kết quả của hồi quy LASSO. Hiệu suất của các mô hình này được thể hiện ở Bảng 2.

Bảng 2: Kết quả dự báo với các biến từ mô hình LASSO của các mô hình học máy

Model	accuracy_score	f1_score	precision_score	recall_score
Logistic Regression	94,43%	88,16%	88,16%	88,16%
Decision Tree	91,64%	82,80%	80,25%	85,53%
Random Forest	95,05%	89,33%	90,54%	88,16%
KNN	93,81%	86,11%	91,18%	81,58%
SVM	95,67%	90,91%	89,74%	92,11%
ANN	96,28%	92,00%	93,24%	90,79%

Nguồn: Nhóm tác giả tính toán từ dữ liệu thu thập

Kết quả (Bảng 2) cho thấy, dự báo phân loại doanh nghiệp trên tập kiểm tra cho thấy tất cả các mô hình đều cho khả năng dự báo đúng trên 92% với mô hình mạng nơ ron nhân tạo (ANN) cho kết quả dự báo chính xác cao nhất với xấp xỉ 96,3% cao hơn so với các mô hình học máy thông minh khác như Random Forest hay SVC. Và bất ngờ là mô hình hồi quy Logistics truyền thống, thì có độ chính xác là hơn 94% chỉ xếp thứ 4 sau các mô hình thông minh khác.

Chỉ tiêu Recall_score cao đồng nghĩa với việc tỷ lệ các doanh nghiệp trong diện rủi ro bị bỏ sót là thấp hơn và Precision_score cao phản ánh độ chính xác dự báo cao. Mô hình tốt nhất là mô hình ngoài độ chính xác cao thì còn có đồng thời cả 2 chỉ tiêu Recall_score và Precision_score đều cao (> 90%). Ở hai chỉ tiêu này thì mô hình mạng nơ ron nhân tạo (ANN) tiếp tục cho hiệu suất dự báo cao nhất. Mô hình Decision Tree các chỉ số thể hiện khả năng bỏ sót doanh nghiệp trong diện rủi ro cao hơn và độ chính xác dự báo thấp hơn (Recall_score và Precision_score đều < 85%), cho thấy tính vững của mô hình Decision Tree không lớn.

Từ kết quả so sánh đánh giá trên, nhóm chúng tôi quyết định lựa chọn mô hình nơ ron nhân tạo (ANN) để tiếp tục thực hiện đánh giá và dự báo, nhận diện rủi ro tài chính trong doanh nghiệp sản xuất niêm yết trên sàn chứng khoán.

Phần 3: Huấn luyện và thực nghiệm dữ liệu bằng mô hình nơ ron nhân tạo

Mô hình mạng nơ ron nhân tạo (ANN) là một dạng phổ biến của mạng nơ ron trong học máy, trong đó có ít nhất một lớp ẩn giữa lớp đầu vào và lớp đầu ra.

Huấn luyện lại trên bộ dữ liệu

Sau khi lựa chọn mô hình mạng nơ ron là mô hình chính, nhóm chúng tôi thực hiện việc huấn luyện lại mô hình mạng nơ ron dựa trên bộ dữ liệu của các doanh nghiệp thuộc ngành sản xuất niêm yết trên thị trường chứng khoán từ năm 2017 đến năm 2020 với các biến X1, X6, X11, X18, X26 (Bảng 1) được chọn lọc nhờ kết quả kiểm thử của mô hình hồi quy LASSO. Nhóm tiếp tục phân chia ngẫu nhiên tập dữ liệu huấn luyện từ năm 2017 đến 2020, mô hình mạng nơ ron với 70% dữ liệu 'train' và 30% dữ liệu 'test', sau đó cho chạy mô hình mạng nơ ron. Kết quả độ chính xác của mô hình trên tập 'train' là 95% và độ chính xác trên tập 'test' là 96%.

Độ chính xác của mô hình được sử dụng để đánh giá hiệu suất của mô hình mạng nơ ron, đối với khả năng phân loại thì độ chính xác thường là một chỉ số quan trọng để biết mức độ chính xác sau khi được huấn luyện của mô hình trên tập kiểm thử. Tuy nhiên, để thấy được rõ độ tin cậy của kết quả mà mô hình đạt được, cũng như phát hiện ra lỗi mà mô hình gặp phải, từ đó cải thiện và tinh chỉnh mô hình để nâng cao hiệu suất dự đoán và giảm thiểu các lỗi trong quá trình huấn luyện và kiểm thử mô hình, nhóm chúng tôi đã sử dụng ma trận nhầm lẫn (Confusion Matrix) giúp đánh giá và hiệu suất của một mô hình phân loại, đồng thời hỗ trợ quyết định trong việc cải thiện mô hình.

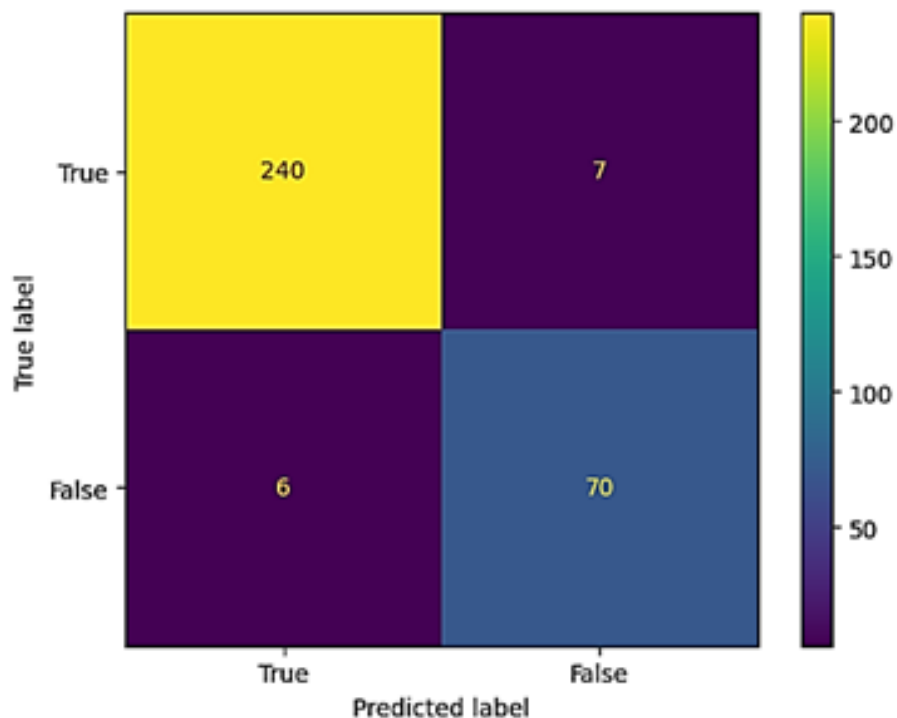
Kết quả thu được ma trận nhầm lẫn sau với thực y là kết quả thực tế và thực x là kết quả dự báo của mô hình.

- Actual: 0 là kết quả thực tế doanh nghiệp không có rủi ro.
- Actual: 1 là kết quả thực tế doanh nghiệp có rủi ro.
- Predicted: 0 là kết quả dự báo doanh nghiệp không có rủi ro.
- Predicted: 1 là kết quả dự báo doanh nghiệp có rủi ro.

Từ kết quả của ma trận nhầm lẫn (Hình 3) ta có thể thấy được:

- Tọa độ (Actual:0, Predicted: 0) - Thực tế doanh nghiệp không có rủi ro, mô hình dự báo doanh nghiệp không có rủi ro: 240 doanh nghiệp.
- Tọa độ (Actual:0, Predicted: 1) - Thực tế doanh nghiệp không có rủi ro, mô hình dự báo doanh nghiệp có rủi ro: 7 doanh nghiệp.
- Tọa độ (Actual: 1, Predicted: 0) - Thực tế doanh nghiệp có rủi ro, mô hình dự báo doanh nghiệp không có rủi ro: 6 doanh nghiệp.
- Tọa độ (Actual:1, Predicted: 1) - Thực tế doanh nghiệp có rủi ro, mô hình dự báo doanh nghiệp có rủi ro: 70 doanh nghiệp.

Hình 3: Kết quả Confusion Matrix trên tập dữ liệu giai đoạn 2017-2020



Nguồn: Nhóm tác giả tính toán từ dữ liệu thu thập

Như đã nói ở phần trước chỉ tiêu *Recall_score* cao đồng nghĩa với việc tỷ lệ các doanh nghiệp trong diện rủi ro bị bỏ sót là thấp hơn và *Precision_score* cao phản ánh độ chính xác dự báo cao. Sau khi huấn luyện lại trên bộ dữ liệu thì chỉ tiêu *Recall_score* và *Precision_score* đạt được lần lượt là 92,1% và 90,9%.

Thử lại trên dữ liệu thực tế năm 2021

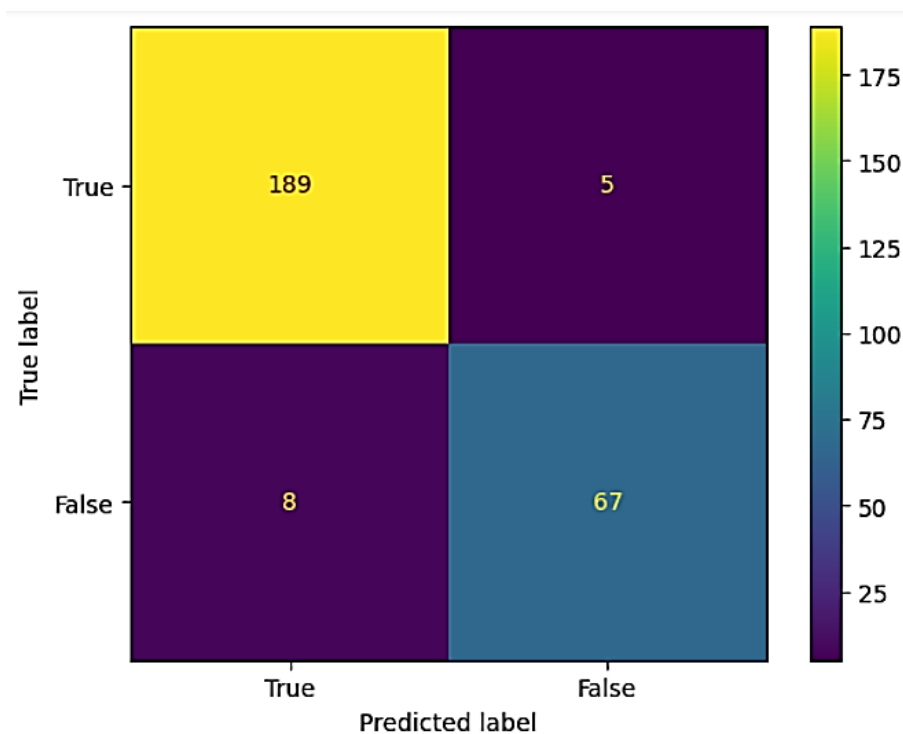
Để xác định mô hình vẫn hoạt động tốt và không gặp hiện tượng overfitting, nhóm tác giả thực hiện kiểm tra mô hình trên tập dữ liệu của năm 2021. Độ chính xác của mô hình trên tập dữ liệu của các doanh nghiệp thuộc ngành sản xuất niêm yết trên thị trường chứng khoán từ năm 2017 đến 2020 là 96% vậy đối với dữ liệu mà mô hình chưa được học và kiểm thử thì sẽ cho ra kết quả chính xác đến bao nhiêu phần trăm?

Để kiểm chứng hiệu suất cả mô hình, nhóm tác giả cho mô hình tiếp tục chạy trên tập dữ liệu thực tế của các doanh nghiệp thuộc ngành sản xuất niêm yết trên thị trường chứng khoán năm 2021 với các biến X1, X6, X11, X18, X26. Kết quả thu được độ chính xác lên tới 95%.

Và kết quả khi sử dụng ma trận nhầm lẫn (Hình 4) cho thấy:

- Tọa độ (Actual: 0, Predicted: 0) - Thực tế doanh nghiệp không có rủi ro, mô hình dự báo doanh nghiệp không có rủi ro: 189 doanh nghiệp.
- Tọa độ (Actual: 0, Predicted: 1) - Thực tế doanh nghiệp không có rủi ro, mô hình dự báo doanh nghiệp có rủi ro: 5 doanh nghiệp.
- Tọa độ (Actual: 1, Predicted: 0) - Thực tế doanh nghiệp có rủi ro, mô hình dự báo doanh nghiệp không có rủi ro: 8 doanh nghiệp.
- Tọa độ (Actual: 1, Predicted: 1) - Thực tế doanh nghiệp có rủi ro, mô hình dự báo doanh nghiệp có rủi ro: 67 doanh nghiệp

Hình 4: Kết quả Confusion Matrix trên tập dữ liệu năm 2021



Nguồn: Nhóm tác giả tính toán từ dữ liệu thu thập

Kết luận: Mô hình học máy mà nhóm xây dựng sử dụng mô hình mạng Neuron thần kinh kết hợp các chỉ tiêu tài chính từ mô hình hồi quy LASSO đưa ra độ chính xác lên đến 95% đối với dữ liệu thực, và tỷ lệ sai sót là 5%. Trong đó, có 2,03% dự đoán doanh nghiệp gặp rủi ro tài chính nhưng thực tế doanh nghiệp không gặp rủi ro tài chính, điều này không gây nguy hiểm cho doanh nghiệp và chỉ có 2,97% dự đoán doanh nghiệp không gặp rủi ro tài chính, nhưng thực tế doanh nghiệp đang gặp rủi ro tài chính gây nguy hiểm cho doanh nghiệp thuộc nhóm này. Vì thế, mô hình mà nhóm xây dựng có thể được xem là một mô hình có độ chính xác cao và hữu ích cho công tác quản lý rủi ro của doanh nghiệp.

KẾT LUẬN VÀ KHUYẾN NGHỊ

Kết luận

Dựa trên bộ dữ liệu của 269 doanh nghiệp sản xuất niêm yết trên sàn chứng khoán Việt Nam từ năm 2017 đến 2021, với 29

Chỉ số ảnh hưởng tới rủi ro tài chính, kết quả nghiên cứu cho thấy từ 29 chỉ số này thông qua mô hình LASSO chọn ra được 5 chỉ số về: Khả năng thanh toán, Hiệu suất hoạt động, Khả năng sinh lời, Cơ cấu nợ và Tỷ lệ tài sản là nợ dài hạn của doanh nghiệp có ảnh hưởng nhất tới Rủi ro tài chính của doanh nghiệp trong 5 năm (2017-2021). Điều này phù hợp với thực tế khi các chỉ số này đều được các doanh nghiệp dùng để xem xét để đánh giá tài chính của doanh nghiệp.

Trong bài nghiên cứu, nhóm tác giả cũng so sánh các mô hình dự báo và chỉ ra rằng, mạng nơ ron nhân tạo là mô hình tốt nhất dùng để dự báo xem doanh nghiệp có gặp rủi ro tài chính không. Điều này phù hợp với kết quả nghiên cứu của Đào Trọng Thịnh và Doãn Văn Toàn (2016) và cũng phù hợp với kết quả nghiên cứu của Qin (2022). Mô hình mạng nơ ron nhân tạo có khả năng thích ứng mạnh mẽ khi đối mặt với các loại dữ liệu khác nhau, khả năng hoạt động mạnh mẽ. Mô hình chúng tôi xây dựng có độ chính xác 96% trên bộ dữ liệu huấn luyện và 95% dự đoán chính xác doanh nghiệp có gặp rủi ro về tài chính không dựa trên tập dữ liệu thực tế, mô hình cũng hoạt động tốt với các chỉ số đánh giá mô hình học máy, như: recall 92,1%, precision là 90,9%.

Kết quả nghiên cứu có thể hỗ trợ các phòng ban trong doanh nghiệp phát hiện rủi ro tài chính, giúp nhà đầu tư cũng như chủ doanh nghiệp trong việc đưa ra quyết định của mình. Bên cạnh đó, kết quả này là tài liệu để hỗ trợ các nhóm nghiên cứu tiếp theo.

Khuyến nghị

Quản trị rủi ro tài chính là một yếu tố quan trọng của quản trị rủi ro trong doanh nghiệp đảm bảo quá trình vận hành thành công, bao gồm các hoạt động như nhận diện các rủi ro có thể xảy ra, đánh giá tác động đối với hoạt động kinh doanh và chuẩn bị kế hoạch đối phó với những sự kiện bất lợi. Nhưng việc quản trị rủi ro tài chính cho các doanh nghiệp đòi hỏi tốn rất nhiều nguồn lực và thời gian. Doanh nghiệp cần thành lập bộ phận quản trị rủi ro tài chính, tuân thủ tiến trình quản trị rủi ro tài chính, sử dụng nhiều kỹ thuật về nhận dạng rủi ro, tìm kiếm các nguồn tài trợ cho các hoạt động quản trị rủi ro tài chính... Việc ứng dụng phân tích dữ liệu trong quản trị rủi ro tài chính có thể giúp doanh nghiệp đánh giá chi tiết hơn, toàn diện hơn bản thân doanh nghiệp đang trong tình trạng như thế nào. Điều này giúp phát huy tối đa hiệu quả nguồn lực, tiết kiệm thời gian mà vẫn cho ra kết quả chính xác, giúp ích rất nhiều cho việc phát triển và tồn tại của doanh nghiệp. Doanh nghiệp cần chú trọng hơn về nhân lực và nguồn lực trong việc phát triển phân tích dữ liệu, nên kết hợp giữa kinh nghiệm và sự hiểu biết dữ liệu để đưa ra quyết định đúng đắn. Bởi dữ liệu là yếu tố phản ánh chân thực nhất tất cả mọi vấn đề của doanh nghiệp./.

Tài liệu tham khảo

1. Altman, E. (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, 23, 589-609.
2. Đào Trọng Thịnh, Doãn Văn Toàn (2016), Ứng dụng mô hình mạng thần kinh nhân tạo trong quản trị rủi ro tín dụng tại một số ngân hàng thương mại trên địa bàn TP. Biên Hòa, *Tạp chí Khoa học Lạc Hồng*, số 5, 31-35
3. Đoàn Khánh Hưng và Trần Thị Hiền (2019), Tác động của cấu trúc tài chính đến rủi ro tài chính của các doanh nghiệp dịch vụ du lịch niêm yết tại Việt Nam, *Tạp chí Khoa học Đại học Huế: Kinh tế và Phát triển*, 128(5A), 93-104.
4. Huang, B., Wei, J., Tang, Y., and Liu, C. (2021), *Enterprise risk assessment based on machine learning*, Computational Intelligence and Neuroscience.
5. Murugan, M. S. (2023), Large-scale data-driven financial risk management & analysis using machine learning strategies, *Measurement: Sensors*, 27.
6. Lê Hải Trung, Trương Thị Thùy Dương (2023), Ứng dụng phương pháp LASSO trong dự báo rủi ro phá sản của các doanh nghiệp Việt Nam, *Tạp chí Khoa học & Đào tạo Ngân hàng*, số 250.
7. Qin, W. (2022), Research on financial risk forecast model of listed companies based on convolutional neural network, *Scientific Programming*, 1-10, DOI 10.1155/2022/3652931.
8. Shrivastava, S., Jeyanthi, P. M., and Singh, S. (2020), Failure prediction of Indian Banks using SMOTE, LASSO regression, bagging and boosting, *Cogent Economics & Finance*, 8(1).
9. Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society*, B, 58, 267-288.
10. Valaskova, K., Kliestik, T., and Kovacova, M. (2018), Management of financial risks in Slovak enterprises using regression analysis, *Oeconomia copernicana*, 9(1), 105-121.
11. Valaskova, K., Kliestik, T., Svabova, L., and Adamko, P. (2018), Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis, *Sustainability*, 10(7).
12. Vuong, Q. H. (2023). *Mindsponge Theory*. Walter de Gruyter GmbH.
13. Zeytinoglu, E., and Akarim, Y. D. (2013), Financial failure prediction using financial ratios: An empirical application on Istanbul Stock Exchange, *Journal of Applied Finance and Banking*, 3(3).

Ngày nhận bài: 08/5/2024; Ngày phản biện: 15/5/2024; Ngày duyệt đăng: 31/5/2024

URL: <https://kinhtevadubao.vn/nghien-cuu-ung-dung-phan-tich-du-lieu-trong-quan-tri-rui-ro-tai-chinh-tai-cac-cong-ty-niem-yet-tren-san-chung-khoan-viet-nam-28875.html>

© **Kinh tế và Dự báo - Bộ Kế hoạch và Đầu tư**