

SỬ DỤNG CÁC MÔ HÌNH CÂY PHÂN LỚP DỰ BÁO KIẾT QUỆ TÀI CHÍNH CHO DOANH NGHIỆP VIỆT NAM

HUỖNH THỊ CẨM HÀ

Trường Đại học Kinh tế Thành phố Hồ Chí Minh - hatcdn@ueh.edu.vn

NGUYỄN THỊ UYÊN UYÊN

Trường Đại học Kinh tế Thành phố Hồ Chí Minh - uyentcdn@ueh.edu.vn

LÊ ĐÀO TUYẾT MAI

Trường Đại học Kinh tế Thành phố Hồ Chí Minh - tuyet25mai06@gmail.com

(Ngày nhận: 12/05/2017; Ngày nhận lại: 28/05/2017; Ngày duyệt đăng: 04/08/2017)

TÓM TẮT

Bài viết thực hiện nhằm dự báo kiệt quệ tài chính (KQTC) cho các doanh nghiệp Việt Nam bằng việc sử dụng mô hình cây phân lớp dựa trên thuật toán C4.5 và thuật toán AdaBoost cho mẫu dữ liệu nghiên cứu gồm 664 công ty trên thị trường chứng khoán Việt Nam từ năm 2009-2015. Kết quả nghiên cứu cho thấy các mô hình cây phân lớp hoàn toàn phù hợp để dự báo KQTC cho doanh nghiệp, đặc biệt là mô hình cây dựa trên thuật toán Adaboost, với kết quả dự báo chính xác trên 90%. Chúng tôi tìm thấy ba thuộc tính đóng vai trò quan trọng nhất trong dự báo KQTC từ mô hình cây cho các doanh nghiệp Việt Nam. Thứ nhất, nếu chỉ số tăng trưởng trong vốn chủ sở hữu của doanh nghiệp không đạt được giá trị tối thiểu là -0,3757; doanh nghiệp sẽ đối diện với KQTC. Thứ hai, khi tăng trưởng trong vốn chủ sở hữu của doanh nghiệp lớn hơn -0,3757, doanh nghiệp vẫn sẽ rơi vào KQTC khi đồng thời có tỷ lệ nợ trên vốn chủ sở hữu lớn hơn 3,2136 và tăng trưởng trong vốn chủ sở hữu nhỏ hơn 0,1805. Thứ ba, nếu tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu không thể lớn hơn 0,2 và tỷ lệ nợ trên vốn chủ sở hữu vượt quá 4,3591; doanh nghiệp sẽ có khả năng KQTC.

Từ khóa: kiệt quệ tài chính; mô hình cây phân lớp; thuật toán Adaboost ; thuật toán C4.5.

Predicting financial distress using decision tree models: Evidence from Vietnam

ABSTRACT

This study aims to predict financial distress of Vietnamese firms using Decision Tree models from C4.5 algorithm and Adaboost algorithm to analyze the data collected from a sample of 664 firms listed on Vietnam stock market in the 2009-2015 period. The results show that Decision Trees models, especially the Decision Tree model based on Adaboost algorithm, can be properly used to measure firm financial distress with the forecast accuracy of over 90%. The study also detects three most important ratios for predicting financial distress of Vietnamese firms. First, a firm is considered to be in financial distress if its owners' equity ratio does not meet the minimum growth of -0.3757. Second, even if its owners' equity ratio growth is higher than -0.3757, a firm still faces financial distress when concurrently its debt-to-equity ratio is above 3.2136 and owners' equity ratio growth stays below 0.1805. Third, if a firm's return on equity falls below 0.2 and its debt-to-equity ratio exceeds 4.3591; it will be classified as being in financial distress.

Keywords: Adaboost algorithm; C4.5 algorithm; Decision Tree model; financial distress.

1. Giới thiệu

Trong bối cảnh hội nhập nền kinh tế toàn cầu ngày càng sâu rộng, doanh nghiệp Việt Nam luôn đối diện với nhiều rủi ro tiềm ẩn khó lường. KQTC có thể xảy ra đối với bất kỳ

doanh nghiệp nào cũng như xảy ra vào bất kỳ giai đoạn nào trong chu kỳ kinh doanh. Việc dự báo khi nào doanh nghiệp sẽ rơi vào tình trạng KQTC giúp cho các nhà quản trị doanh nghiệp có thể đưa ra các quyết định phù hợp

nhằm ngăn ngừa các nguy cơ gây KQTC, duy trì hoạt động và thúc đẩy doanh nghiệp tiếp tục tăng trưởng. Nhận thấy được vai trò quan trọng từ việc dự báo sớm KQTC, nhiều phương pháp từ đơn giản đến phức tạp nhằm dự báo KQTC đã được nghiên cứu trên thế giới, như phương pháp phân tích đơn biến - đa biến, phân tích logit, phân tích probit. Tuy nhiên, các phương pháp này lại đi kèm một số nhược điểm về giả định trong việc sử dụng các mô hình phân tích khiến việc thực hiện trở nên khó khăn. Thời gian gần đây, những mô hình cây phân lớp dựa trên các thuật toán và mô hình mạng thần kinh được mở rộng sử dụng trong các nghiên cứu dự báo KQTC, đã tạo được rất nhiều sự chú ý trong giới học thuật trên thế giới khi đã khắc phục được những hạn chế của các mô hình truyền thống, đồng thời lại trực quan và dễ sử dụng. Tuy vậy, hướng nghiên cứu sử dụng các mô hình cây phân lớp bên cạnh các mô hình truyền thống trong việc dự báo KQTC tại Việt Nam vẫn chưa được khai thác nhiều. Xuất phát từ ý tưởng này, bài viết muốn tìm hiểu khả năng sử dụng mô hình cây phân lớp trong việc dự báo KQTC cho các doanh nghiệp Việt Nam, thể hiện qua các mục tiêu cụ thể sau: (i) kiểm định sự phù hợp của mô hình cây phân lớp trong dự báo KQTC cho các doanh nghiệp Việt Nam, (ii) xác định các thuộc tính quan trọng để dự báo KQTC trong mô hình cây và (iii) đo lường mức độ dự báo chính xác và mức độ phù hợp của mô hình cây phân lớp dựa trên thuật toán C4.5 và thuật toán Adaboost. Kết quả nghiên cứu đóng góp việc ủng hộ mô hình cây phân lớp hoàn toàn phù hợp để dự báo KQTC cho doanh nghiệp Việt Nam, đặc biệt là mô hình cây dựa trên thuật toán Adaboost, khả năng dự báo chính xác trên 90%. Các thuộc tính tăng trưởng trong vốn chủ sở hữu, tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu, tỷ lệ nợ trên vốn chủ sở hữu đóng vai trò quan trọng nhất để dự báo KQTC trong mô hình cây phân lớp. Kết quả còn cho thấy nguyên nhân chủ yếu gây ra KQTC là do mức

độ sử dụng đòn bẩy tài chính quá cao và hoạt động kinh doanh không đạt hiệu quả.

2. Tổng quan các nghiên cứu trước

2.1. Thước đo nhận diện kiệt quệ tài chính

Hầu hết các nghiên cứu trước đây về dự báo KQTC tập trung nhiều vào dự báo phá sản (Altman, 1968; Ball và Foster, 1982; Moses và Liao, 1995). Tuy nhiên, các nghiên cứu gần đây chỉ ra rằng KQTC không đồng nhất với phá sản và cho rằng không phải tất cả doanh nghiệp trải qua KQTC cuối cùng đều sẽ đệ đơn phá sản (Ward và Foster, 1997; He và cộng sự, 2010). Việc xuất hiện nhiều quan điểm khác nhau về KQTC trong các nghiên cứu về dự báo KQTC là do sự khác nhau trong việc lựa chọn mẫu nghiên cứu giữa các nghiên cứu thực nghiệm cũng như sự đa dạng và phức tạp của các trạng thái KQTC (Wruck, 1990), gồm có: trạng thái thất bại, mất khả năng thanh khoản, vỡ nợ và phá sản (Altman và Hotchkiss, 2005). Chính vì vậy, có nhiều thước đo để nhận diện tình trạng KQTC của doanh nghiệp. Một số nghiên cứu nhận diện tình trạng KQTC dựa vào các dữ liệu sổ sách kế toán và dữ liệu thị trường (Denis và Denis, 1995; Andrade và Kaplan, 1998; Whitaker, 1999). Nhiều nghiên cứu khác lại dựa vào các động thái của doanh nghiệp như cắt giảm hay ngưng chi trả cổ tức, hủy niêm yết, nộp đơn xin phá sản hoặc thực hiện mua bán sáp nhập với doanh nghiệp khác (Turetsky và McEwn, 2001; Altman và Hotchkiss, 2005). Thời gian gần đây, nhiều nghiên cứu đã khẳng định rằng các chỉ số xác định KQTC như chỉ số Z của Altman (Altman, 1968); chỉ số O của Ohlson (Ohlson, 1980); hoặc chỉ số Zmijewski (Zmijewski, 1984) có thể được sử dụng như một thước đo để xác định doanh nghiệp có đang trong tình trạng KQTC hay không (Grice, 2000; Altman và cộng sự, 2010). Trong đó, chỉ số Zmijewski được sử dụng phổ biến nhất vì không nhạy cảm với các trạng thái khác nhau của KQTC và cũng không nhạy cảm với ngành (Munsif và cộng sự, 2011; Kim và Upneja, 2014).

2.2. Mô hình cây phân lớp dùng trong dự báo kiệt quệ tài chính

Các bằng chứng thực nghiệm dự báo KQTC thời gian qua đã cho thấy các mô hình này đã và đang được cải thiện về khả năng dự báo lẫn độ chính xác qua từng thời kỳ khác nhau, từ phương pháp so sánh các tỷ số tài chính giữa các doanh nghiệp (Fitzpatrick, 1931) đến phương pháp phân tích đơn biến (Beaver, 1960), phân tích phân biệt đa biến (Altman, 1968) và phân tích thống kê xác suất có điều kiện Logistic (Ohlson, 1980). Phân tích phân biệt đa biến và phân tích Logistic là hai phương pháp phổ biến vì có độ chính xác cao. Tuy nhiên, cả hai mô hình này đều có những hạn chế về giả định khiến việc sử dụng trở nên khó khăn. Phân tích phân biệt đa biến giả định các biến độc lập có phân phối chuẩn và có ma trận phương sai - hiệp phương sai phải giống nhau giữa các doanh nghiệp KQTC và không KQTC, phân tích Logistic (Ohlson, 1980) lại có giả định biến động đồng nhất của dữ liệu và sự nhạy cảm với đa cộng tuyến. Từ những năm đầu thập niên 90, với sự phát triển của khoa học công nghệ, mô hình mạng thần kinh nhân tạo (ANN - Artificial Neural networks) và mô hình cây phân lớp (Decision tree classification) trở thành hai phương pháp tiếp cận phi tham số thường được sử dụng trong nghiên cứu dự báo KQTC. Tuy thường xuyên gặp vấn đề quá khớp dữ liệu¹ và rất khó để khái quát một cấu trúc cây quy chuẩn cho những trường hợp khác nhau vì sự nhạy cảm với thay đổi mẫu, song mô hình cây phân lớp lại có ưu thế vượt trội vì có thể khắc phục hoàn toàn các nhược điểm của những mô hình dự báo khác khi không đòi hỏi bất kỳ giả định về mẫu nghiên cứu cũng như không quá phức tạp khi sử dụng và không gặp vấn đề hộp đen² liên quan đến tầm quan trọng của mỗi biến như mô hình ANN (Kim và Upneja, 2014). Bên cạnh đó, những hạn chế của mô hình cây phân lớp có thể dễ dàng được khắc phục, cụ thể là vấn đề quá khớp dữ liệu được khắc phục bằng kỹ thuật cắt tỉa nhánh và

vấn đề nhạy cảm với thay đổi mẫu được cải thiện bởi kỹ thuật boosting (Bastos, 2008; Kotsiantis và cộng sự, 2006).

Với những ưu thế vượt trội của mình, mô hình cây phân lớp được đánh giá là một công cụ mạnh, phổ biến và đặc biệt thích hợp cho phân lớp dữ liệu nói chung và dự báo KQTC nói riêng. Mô hình cây phân lớp là một loại kỹ thuật khai phá dữ liệu dùng trong thống kê thể hiện dưới dạng biểu đồ phát triển hình cây gồm: nút trên cùng của cây là gốc; mỗi nút trong của cây biểu diễn một kiểm tra trên một thuộc tính đơn; nhánh cây biểu diễn các kết quả của kiểm tra trên nút trong; nút lá của cây biểu diễn sự phân phối của các lớp giá trị, là đích đến cuối cùng của sự phân lớp. Quá trình tạo cây xuất phát từ nút gốc với tất cả mẫu huấn luyện, sau đó phân chia một cách đệ quy dựa trên thuộc tính tốt nhất để có được sự phân phối của các lớp giá trị ở nút lá. Mô hình cây phân lớp có năm ưu điểm lớn so với các kỹ thuật khai phá dữ liệu khác: (1) việc giải thích cho bất kỳ một sự phân lớp hay dự báo nào của mô hình này cũng đều tương đối minh bạch do quá trình xây dựng cây rất rõ ràng, thậm chí với cả những tập dữ liệu lớn khiến cho hình dáng cây trở nên phức tạp; (2) các thuật toán được sử dụng thường là những kiểm tra đơn giản, dễ dàng tính toán tại từng nút, do vậy, dễ dàng tạo ra những cây phân lớp với số phân nhánh thấp; (3) mô hình này rút trích ra các quy luật nhất định từ tập dữ liệu, từ đó, dự báo những xu hướng trong tương lai của dữ liệu nên không yêu cầu bất kỳ giả định thống kê liên quan đến các dữ liệu trong mẫu; (4) mô hình này có khả năng xử lý với cả thuộc tính liên tục lẫn rời rạc (5) kết quả của mô hình này có thể cho thấy được những thuộc tính nào là quan trọng nhất trong quá trình phân lớp

Trong các thuật toán xây dựng mô hình cây phân lớp, C4.5 là thuật toán được sử dụng phổ biến nhất (Quinlan, 1996; Shirata, 1998) do thuật toán C4.5 có khả năng làm việc với thuộc tính liên tục, thuộc tính có nhiều giá trị

và dữ liệu bị thiếu hoặc bị nhiễu. Thuật toán C4.5 thực hiện phân ngưỡng thuộc tính liên tục bằng phép tách nhị phân và dựa vào độ đo Gain Ratio để lựa chọn thuộc tính tốt nhất để phát triển. Bản thân thuật toán C4.5 đã bao hàm các kỹ thuật nhằm khắc phục các nhược điểm quá khớp dữ liệu cũng như cải thiện sự ổn định và mức độ dự báo chính xác của mô hình cây phân lớp, bao gồm cắt tia³ và thử nghiệm chéo⁴, có chức năng làm giảm mức độ phân chia của mô hình trong trường hợp các nút có chứa những quan sát giống hệt nhau về giá trị hoặc biến phụ thuộc không có tiêu chí dừng. Chính vì vậy, mô hình cây phân lớp được xây dựng từ thuật toán C4.5 thường được sử dụng để dự báo KQTC (Huang và cộng sự, 2005; Bastos, 2008; Kim và Upneja, 2014).

Ngoài ra, để cải thiện sự nhạy cảm với thay đổi mẫu và nâng cao mức độ dự báo chính xác của mô hình cây phân lớp, nhiều nghiên cứu gần đây sử dụng các phương pháp tập hợp mô hình.⁵ Trong đó, thuật toán Adaboost được đề xuất bởi Freund và Schapire (1996) là một trong những phương pháp tập hợp mô hình quan trọng nhất vì có nhiều bằng chứng vững chắc về mặt lý thuyết và thực nghiệm cho thấy Adaboost có mức độ dự báo chính xác cao, đơn giản, ứng dụng rộng rãi và thành công (Alfaro và cộng sự, 2008a; Quinlan, 1996). Thêm vào đó, các nghiên cứu khác cho thấy Adaboost không thường gặp vấn đề quá khớp dữ liệu. Tương tự C4.5, Adaboost cải thiện cho mô hình cây phân lớp cũng xây dựng các điều kiện phân lớp thuộc tính bằng phép tách nhị phân. Tuy nhiên, thuật toán Adaboost sử dụng thêm các trọng số thích ứng để thiết lập các phân lớp trên tập huấn luyện. Sau khi quá trình phân lớp đầu tiên được xây dựng, quá trình phân lớp tiếp theo sẽ được thiết lập dựa trên một trọng số khác với trọng số của phân lớp đầu tiên đối với phân lớp không thỏa điều kiện. Quá trình này sẽ lặp lại cho đến khi đạt được các điều kiện dừng và các phân lớp đơn được

kết hợp thành bộ phân lớp cuối cùng với mức độ chính xác cao. Dựa vào các tổng quan nghiên cứu trên thế giới, nhóm tác giả bài viết kế thừa ứng dụng mô hình cây dựa trên thuật toán C4.5 và Adaboost để kiểm định mức độ phù hợp của mô hình khi dự báo KQTC cho các doanh nghiệp Việt Nam.

3. Dữ liệu và phương pháp nghiên cứu

3.1. Dữ liệu nghiên cứu

Mẫu dữ liệu của bài viết gồm 664 công ty phi tài chính được niêm yết trên HSX và HNX từ 2009 đến 2015, chúng tôi loại ra khỏi mẫu quan sát các công ty thuộc các lĩnh vực tài chính, ngân hàng, bảo hiểm vì các công ty này còn chịu điều tiết bởi một số luật điều chỉnh riêng biệt khác khiến các báo cáo tài chính sẽ khác đáng kể so với các công ty nhóm ngành khác. Các dữ liệu tài chính và quản trị của các doanh nghiệp được thu thập từ báo cáo tài chính đã kiểm toán và báo cáo thường niên; dữ liệu giá chứng khoán được thu thập từ cophieu68.vn; dữ liệu GDP được lấy từ Tổng cục thống kê Việt Nam. Ngoài ra, đối với các doanh nghiệp đã hủy niêm yết bắt buộc, bài viết thu thập dữ liệu từ một năm trước khi hủy bỏ niêm yết trở về trước trong giai đoạn từ năm 2009 đến năm 2015 để đảm bảo khả năng dự báo sớm KQTC trước một năm. Chúng tôi nhận diện doanh nghiệp đang trong tình trạng KQTC khi dòng tiền tạo ra không đủ chi trả các khoản nợ và giá trị thị trường của doanh nghiệp bị sụt giảm dựa trên phương pháp của Whitaker (1999), đồng thời, kết hợp thêm mô hình chỉ số Zmijewski⁶ để xác định tình trạng KQTC của các doanh nghiệp Việt Nam. Ngoài ra, nhằm tăng quy mô của mẫu nghiên cứu và sức mạnh dự báo của mô hình nghiên cứu, nhóm tác giả đề tài bổ sung các doanh nghiệp bị hủy niêm yết bắt buộc trên HSX và HNX. Như vậy, doanh nghiệp được nhận diện là KQTC trong năm quan sát khi thỏa mãn ít nhất một trong những điều kiện sau: *Thứ nhất*, dòng tiền hoạt động của doanh nghiệp không đủ chi trả các khoản nợ ngắn hạn và giá trị thị trường của doanh nghiệp bị sụt giảm. *Thứ hai*,

chỉ số Zmijewski của doanh nghiệp vượt quá giá trị không. *Thứ ba*, doanh nghiệp bị hủy niêm yết bắt buộc hoặc tuyên bố phá sản. Cuối cùng, tác giả thu được mẫu dữ liệu gồm

268 doanh nghiệp trên HSX, 325 doanh nghiệp trên HNX và 71 doanh nghiệp đã hủy niêm yết bắt buộc. Thống kê số quan sát của mẫu dữ liệu được thể hiện ở Bảng 1.

Bảng 1

Thống kê số quan sát của mẫu dữ liệu doanh nghiệp

Năm	KQTC		Không KQTC		Tổng	
	Số quan sát	Tỷ trọng	Số quan sát	Tỷ trọng	Số quan sát	Tỷ trọng
2009	85	1,89%	568	12,61%	653	14,50%
2010	81	1,80%	583	12,94%	664	14,74%
2011	101	2,24%	564	12,52%	665	14,76%
2012	121	2,69%	531	11,79%	652	14,47%
2013	115	2,55%	530	11,76%	645	14,32%
2014	93	2,06%	530	11,76%	623	13,83%
2015	88	1,95%	515	11,43%	603	13,39%
Tổng	684	15,18%	3821	84,82%	4505	100%

Nguồn: Nhóm tác giả tự tổng hợp

3.2. Phương pháp nghiên cứu

Kế thừa nghiên cứu của Kim và Upneja (2014), bài viết sử dụng mô hình cây phân lớp dựa trên thuật toán C4.5 và thuật toán Adaboost. Hai thuật toán này đều xem xét tất cả thuộc tính để phân chia tập dữ liệu đã cho và chọn ra thuộc tính có giá trị Gain Ratio lớn nhất. Tuy nhiên, thuật toán Adaboost bổ sung thêm trọng số để giúp mô hình cây phân lớp tăng tính thích ứng với sự thay đổi của mẫu nghiên cứu. Gain Ratio của một thuộc tính X trong tập dữ liệu D được tính theo công thức:

$$\text{GainRatio}(X) = \frac{\text{InformationGain}(X)}{\text{SplitInfo}_X(D)} \quad (1)$$

Với:

$$\text{InformationGain}(X) = \text{Info}(D) - \text{Info}_X(D)$$

$$= \left[\sum_{i=1}^n -p_i \log_2(p_i) \right] - \left[\sum_{j=1}^v \frac{|D_j|}{|D|} \sum_{i=1}^n -p_i \log_2(p_i) \right] \quad (2)$$

$$\text{SplitInfo}_X(D) = \sum_{j=1}^v -\frac{|D_j|}{|D|} \log_2 \frac{|D_j|}{|D|} \quad (3)$$

Trong đó,

D là tập huấn luyện. Các phân lớp của D có giá trị $d_i = \{d_1, d_2, \dots, d_n\}$;

Thuộc tính X có các giá trị $x_j = \{x_1, x_2, \dots, x_v\}$, dùng thuộc tính X để phân chia tập huấn luyện D thành v tập con $D_j = \{D_1, D_2, \dots, D_v\}$;

$|D_j|, |D|$: số quan sát thuộc tập con D_j và tập huấn luyện D;

p_i là tỷ lệ các quan sát thuộc phân lớp i chia cho tất cả các quan sát của tập D.

Information Gain của thuộc tính X được tính theo công thức (2) cho biết lượng thông tin thu được sau khi dùng thuộc tính X phân lớp. Vì độ đo Information Gain có xu hướng thiên vị cho các thuộc tính có nhiều giá trị nên độ đo Gain Ratio được tính theo công thức (1) như một đại lượng Information Gain đã được chuẩn hóa nhằm khắc phục hạn chế của độ đo

Information Gain.

Các thuộc tính được sử dụng trong bài viết này gồm 25 các chỉ số tài chính, sắp xếp theo 5 nhóm: nhóm chỉ số khả năng thanh toán, nhóm chỉ số cấu trúc vốn và khả năng trả nợ, nhóm chỉ số khả năng sinh lợi, nhóm chỉ số hoạt động, nhóm chỉ số tăng trưởng và các thông tin phi tài chính như tỷ lệ sở hữu của Hội đồng quản trị, xu hướng giá cổ phiếu, thu nhập trên mỗi cổ phiếu và thay đổi trong GDP. Các thuộc tính này được tóm tắt ở Bảng 2. Nghiên cứu này được thực hiện theo trình tự sau:

Bước một: Thống kê mô tả các thuộc tính được sử dụng để dự báo KQTC.

Bước hai: Sử dụng cây phân lớp bởi thuật toán C4.5 và Adaboost cho toàn bộ các thuộc tính được xác định thông qua thẻ lệnh J48 và thẻ lệnh AdaboostM1 trong chương trình WEKA 3.6.9. Từ đó, bài viết kiểm định mức độ phù hợp, độ chính xác từ việc sử dụng mô hình cây phân lớp để dự báo KQTC. Các mức độ dự báo chính xác của mô hình cho biết độ khớp giữa giá trị dự báo và giá trị quan sát thực tế. Nếu mức độ dự báo chính xác càng lớn thì mô hình sẽ càng phù hợp.

Bước ba: Do có một số thuộc tính không tác động hoặc tác động rất ít đến quá trình phân lớp mô hình cây, bài viết lược bỏ các thuộc tính này ra khỏi mẫu nghiên cứu để giảm độ nhiễu bằng chức năng lược bỏ thuộc tính của WEKA dựa trên bảng xếp hạng Gain Ratio của tất cả thuộc tính. Sau đó, tiếp tục áp dụng thuật toán C4.5 và Adaboost trên cây phân lớp cho mẫu dữ liệu sau khi đã lược bỏ

một số thuộc tính. Việc lược bỏ được thực hiện lần lượt từng thuộc tính, từ thuộc tính có Gain ratio thấp nhất đến khi mô hình có mức độ dự báo chính xác tổng thể lớn nhất.

Bước bốn: Từ các kết quả thu được ở bước 3, nhóm tác giả thực hiện đánh giá và so sánh sức mạnh dự báo của các mô hình và xem xét mức độ phù hợp của 2 thuật toán trong việc dự báo KQTC, thông qua ba chỉ tiêu: mức độ dự báo chính xác tổng thể, hệ số Kappa và diện tích dưới đường cong ROC (hay còn gọi là AUC - Area under ROC curve). Hệ số Kappa là tỷ số giữa tỷ lệ phù hợp quan sát và tỷ lệ phù hợp lý thuyết, nhằm kiểm tra mức độ lặp lại của các dự báo khi áp dụng với một mẫu nghiên cứu khác. Hệ số này biến thiên từ -1 đến +1 tương ứng với hoàn toàn không lặp lại đến lặp lại hoàn toàn. Nếu mô hình có mức độ lặp lại cao khi áp dụng với một mẫu nghiên cứu khác, điều này hàm ý rằng mô hình có tính ứng dụng thực tiễn tốt và nên sử dụng để dự báo KQTC. AUC được dùng để đo lường tính chính xác của mô hình dự báo. Theo đó giá trị của phần diện tích nằm dưới đường ROC có thể được dùng để đo lường tính chính xác của mô hình dự báo, khả năng phân biệt của mô hình tốt hay xấu. Giá trị của AUC nhỏ hơn 0,6 cho thấy khả năng phân biệt kém của mô hình, AUC nằm trong khoảng 0,8 đến 0,9 là khá tốt; trên 0,9 cho là tốt.

Bảng 2

Mô tả các thuộc tính được sử dụng trong mô hình

Tên thuộc tính		Kí hiệu	Mô tả thuộc tính
Nhóm chỉ số khả năng thanh	Tỷ số thanh toán hiện hành	X1	<u>Tài sản ngắn hạn</u> Nợ ngắn hạn
	Tỷ số thanh toán nhanh	X2	<u>Tài sản ngắn hạn - Hàng tồn kho</u> Nợ ngắn hạn

Tên thuộc tính		Kí hiệu	Mô tả thuộc tính
toán	Vòng quay các khoản phải thu	X3	$\frac{\text{Doanh thu thuần}}{\text{Các khoản phải thu}}$
	Tỷ lệ dòng tiền hoạt động trên nợ ngắn hạn	X4	$\frac{\text{Dòng tiền hoạt động}}{\text{Nợ ngắn hạn}}$
Cấu trúc vốn và khả năng trả nợ	Tỷ lệ nợ trên vốn chủ sở hữu (VCSH)	X5	$\frac{\text{Tổng nợ phải trả}}{\text{VCSH}}$
	Tỷ lệ tài sản cố định (TSCĐ) trên vốn dài hạn	X6	$\frac{\text{TSCĐ}}{\text{Tổng nguồn vốn - Nợ ngắn hạn}}$
	Tỷ lệ dòng tiền hoạt động trên tổng nợ	X7	$\frac{\text{Dòng tiền hoạt động}}{\text{Tổng nợ}}$
Nhóm chỉ số khả năng sinh lợi	Biên lợi nhuận ròng	X8	$\frac{\text{Lợi nhuận ròng}}{\text{Doanh thu thuần}}$
	Biên thu nhập hoạt động	X9	$\frac{\text{Thu nhập hoạt động}}{\text{Doanh thu thuần}}$
	Tỷ lệ lợi nhuận ròng trên giá trị sổ sách vốn cổ phần	X10	$\frac{\text{Lợi nhuận ròng}}{\text{Giá trị sổ sách vốn cổ phần}}$
	Tỷ lệ lợi nhuận ròng trên VCSH	X11	$\frac{\text{Lợi nhuận ròng}}{\text{VCSH}}$
	Tỷ lệ thu nhập hoạt động trên giá trị sổ sách vốn cổ phần	X12	$\frac{\text{Thu nhập hoạt động}}{\text{Giá trị sổ sách vốn cổ phần}}$
Nhóm chỉ số hoạt động	Vòng quay tổng tài sản	X13	$\frac{\text{Doanh thu thuần}}{\text{Tổng tài sản}}$
	Kỳ luân chuyển hàng tồn kho	X14	$\frac{\text{Hàng tồn kho} \times 365}{\text{Giá vốn hàng bán}}$
	Vòng quay TSCĐ	X15	$\frac{\text{Doanh thu thuần}}{\text{TSCĐ}}$
Nhóm chỉ số tăng trưởng	Tăng trưởng trong doanh thu	X16	$\frac{\text{Doanh thu thuần năm } t}{\text{Doanh thu thuần năm } t-1} - 1$
	Tăng trưởng trong tài sản	X17	$\frac{\text{Tổng tài sản năm } t}{\text{Tổng tài sản năm } t-1} - 1$
	Tăng trưởng trong thu nhập hoạt động	X18	$\frac{\text{Thu nhập hoạt động năm } t}{\text{Thu nhập hoạt động năm } t-1} - 1$
	Tăng trưởng trong lợi nhuận ròng	X19	$\frac{\text{Lợi nhuận ròng năm } t}{\text{Lợi nhuận ròng năm } t-1} - 1$

Tên thuộc tính		Kí hiệu	Mô tả thuộc tính
Tăng trưởng trong VCSH		X20	$\frac{\text{VCSH năm } t}{\text{VCSH năm } t-1} - 1$
Tỷ lệ sở hữu của Hội đồng quản trị		X21	$\frac{\text{Lượng cổ phần sở hữu bởi Hội đồng quản trị}}{\text{Giá trị sổ sách vốn cổ phần}}$
Các thuộc tính phi tài chính	Xu hướng giá cổ phiếu	X22	$\frac{\text{Giá cổ phiếu năm } t}{\text{Giá cổ phiếu năm } t-1}$
	Thu nhập trên mỗi cổ phiếu	X23	$\frac{\text{Lợi nhuận ròng}}{\text{Số lượng cổ phiếu đang lưu hành}}$
	Thực hành quản trị	X24	Chỉ số quản trị doanh nghiệp CGI ⁷
	Thay đổi trong GDP	Xs	$\text{GDP}_t - \text{GDP}_{t-1}$

Nguồn: Nhóm tác giả tự tổng hợp

4. Kết quả nghiên cứu và thảo luận
Thống kê mô tả

Bảng 3 cung cấp thông tin về giá trị trung bình các thuộc tính giữa các doanh nghiệp KQTC và không KQTC. Số liệu thống kê cho thấy có sự khác biệt rõ rệt giữa giá trị trung bình giữa các doanh nghiệp KQTC và không KQTC. Đầu tiên, giá trị trung bình của các chỉ số thuộc nhóm khả năng thanh toán ở các doanh nghiệp không KQTC cao hơn ở các doanh nghiệp KQTC, cho thấy doanh nghiệp KQTC sẽ gặp khó khăn trong việc thanh toán các khoản nợ vay, nhất là các khoản nợ ngắn hạn. Thứ hai, việc sử dụng đòn bẩy tài chính của các doanh nghiệp KQTC là khá cao và chỉ số thể hiện khả năng trả nợ của các doanh

nh nghiệp KQTC thể hiện không tốt bằng các công ty không KQTC. Thứ ba, giá trị trung bình của các chỉ số thuộc *nhóm khả năng sinh lợi* ở doanh nghiệp KQTC thấp hơn rất nhiều, cho thấy các doanh nghiệp có hoạt động kinh doanh kém hiệu quả dễ xảy ra KQTC hơn. Thứ tư, *nhóm chỉ số tăng trưởng* của doanh nghiệp KQTC thấp hơn doanh nghiệp không KQTC. Ở thuộc tính phi tài chính, giá cổ phiếu của doanh nghiệp KQTC có xu hướng sụt giảm nhiều hơn so với doanh nghiệp không KQTC, *thu nhập trên mỗi cổ phiếu* của các doanh nghiệp không KQTC có giá trị trung bình thấp hơn các doanh nghiệp KQTC và năng lực quản trị của doanh nghiệp KQTC là kém hơn doanh nghiệp không KQTC.

Bảng 3

Thống kê giá trị trung bình các thuộc tính giữa doanh nghiệp KQTC và doanh nghiệp không KQTC

Thuộc tính		Doanh nghiệp không KQTC	Doanh nghiệp KQTC	Tổng thể doanh nghiệp
X1	Tỷ số thanh toán hiện hành	2,751	1,193	2,515
X2	Tỷ số thanh toán nhanh	2,098	0,716	1,888
X3	Vòng quay các khoản phải thu	10,974	7,502	10,447
X4	Tỷ lệ dòng tiền hoạt động trên nợ ngắn hạn	0,254	0,033	0,220

Thuộc tính		Doanh nghiệp không KQTC	Doanh nghiệp KQTC	Tổng thể doanh nghiệp
X5	Tỷ lệ nợ trên VCSH	1,158	7,863	2,176
X6	Tỷ lệ TSCĐ trên vốn dài hạn	0,412	0,732	0,460
X7	Tỷ lệ dòng tiền hoạt động trên tổng nợ	0,195	-0,007	0,164
X8	Biên lợi nhuận ròng	0,099	-0,275	0,042
X9	Biên thu nhập hoạt động	0,120	-0,265	0,061
X10	Tỷ lệ lợi nhuận ròng trên giá trị sổ sách vốn cổ phần	0,261	0,011	0,223
X11	Tỷ lệ lợi nhuận ròng trên VCSH	0,126	-0,329	0,057
X12	Tỷ lệ thu nhập hoạt động trên giá trị sổ sách vốn cổ phần	0,297	0,036	0,258
X13	Vòng quay tổng tài sản	1,159	0,836	1,110
X14	Kỳ luân chuyển hàng tồn kho	2.581	12.093	4.025
X15	Vòng quay TSCĐ	26,941	38,114	28,637
X16	Tăng trưởng trong doanh thu	0,370	0,138	0,335
X17	Tăng trưởng trong tài sản	0,237	0,102	0,216
X18	Tăng trưởng trong thu nhập hoạt động	1,775	0,548	1,589
X19	Tăng trưởng trong lợi nhuận ròng	1,449	0,236	1,265
X20	Tăng trưởng trong VCSH	0,335	-0,056	0,276
X21	Tỷ lệ sở hữu của Hội đồng quản trị	0,295	0,300	0,296
X22	Xu hướng giá cổ phiếu	0,917	0,806	0,900
X23	Thu nhập trên mỗi cổ phiếu	2.608	111	2.229
X24	Thực hành quản trị	40,008	34,916	39,235
Xs	Thay đổi trong GDP	3,69E+14	3,77E+14	3,70E+14

Nguồn: Nhóm tác giả tự tổng hợp

Kiểm định sự phù hợp của mô hình cây phân lớp khi dự báo kiệt quệ tài chính

Thông qua thẻ lệnh J48 trong chương trình WEKA 3.6.9 và thực hiện 10 thử nghiệm chéo với độ tin cậy cho quá trình cắt tĩa nhánh là 0,25; mô hình cây phân lớp dựa trên thuật toán C4.5 có kích thước là 27 nút với 14 nút

lá. Thuật toán Adaboost đã bổ sung thêm trọng số 0,89 để thiết lập cây phân lớp với kích thước là 35 nút với 18 nút lá. Sau khi lược bỏ lần lượt từng thuộc tính, bài viết thu được kết quả gồm 10 thuộc tính có ảnh hưởng đến quá trình dự báo mô hình, được thể hiện ở Bảng 4.

Bảng 4

Các thuộc tính có ảnh hưởng đến quá trình dự báo KQTC

Xếp hạng	Thuộc tính	Tên thuộc tính	Giá trị thông tin
1	X5	Tỷ lệ nợ trên VCSH	0,37023
2	X21	Tỷ lệ sở hữu của Hội đồng quản trị	0,14011
3	X20	Tăng trưởng trong VCSH	0,10267
4	X9	Biên thu nhập hoạt động	0,83900
5	X12	Tỷ lệ thu nhập hoạt động trên giá trị sổ sách vốn cổ phần	0,07812
6	X1	Tỷ số thanh toán hiện hành	0,07445
7	X6	Tỷ lệ TSCĐ trên vốn dài hạn	0,07072
8	X8	Biên lợi nhuận ròng	0,06425
9	X11	Tỷ lệ lợi nhuận ròng trên VCSH	0,05992
10	X2	Tỷ số thanh toán nhanh	0,57720

Nguồn: Nhóm tác giả tự tổng hợp

Sau khi lược bỏ các thuộc tính không tác động hoặc tác động rất ít đến quá trình phân lớp, kích thước mô hình cây phân lớp từ thuật toán C4.5 giảm còn 23 nút với 12 nút lá, trong khi đó, mô hình từ thuật toán Adaboost lại tăng kích thước lên 47 nút với 24 nút lá. Mức độ dự báo chính xác của cả hai mô hình đều cải thiện (Bảng 5). Ở thuật toán C4.5 sau khi lược bỏ một số thuộc tính đã dự báo chính xác 93,4% doanh nghiệp KQTC (ban đầu là 92,8%); 99,4% là mức dự báo chính xác cho doanh nghiệp không KQTC (ban đầu là 99,3 %). Tương tự, ở mô hình cây phân lớp từ thuật toán Adaboost

sau khi lược bỏ một số thuộc tính mức độ dự báo chính xác là cao hơn so với ban đầu, cụ thể mức độ dự báo chính xác 99,5% cho các doanh nghiệp không KQTC (ban đầu 99,4%) và dự báo chính xác 94,2% cho doanh nghiệp KQTC. Do vậy, bài nghiên cứu sử dụng kết quả của mô hình cây phân lớp sau khi loại bỏ một số thuộc tính để thực hiện các phân tích. Ngoài ra, kết quả cho thấy các mô hình cây dự báo KQTC đều có mức độ dự báo chính xác là trên 90%, hàm ý rằng việc sử dụng mô hình cây phân lớp để dự báo KQTC cho các doanh nghiệp Việt Nam là hoàn toàn phù hợp.

Bảng 5

Mức độ dự báo chính xác của các mô hình cây phân lớp

Mô hình		Mức độ dự báo chính xác		
		KQTC	Không KQTC	Tổng thể
Toàn bộ thuộc tính	C4.5	92,80%	99,30%	98,35%
	Adaboost	94,20%	99,30%	98,49%
Sau khi lược bỏ một số thuộc tính	C4.5	93,40%	99,40%	98,51%
	Adaboost	94,20%	99,50%	98,67%

Ghi chú: C4.5, Adaboost là các thuật toán được sử dụng trong mô hình cây phân lớp.

Nguồn: Nhóm tác giả tự tổng hợp

Các kết quả nghiên cứu của mô hình cây sau khi lược bỏ một số thuộc tính được thể hiện ở Bảng 6. Ở mô hình cây từ thuật toán C4.5 chọn *chỉ số X5 (nợ trên vốn chủ sở hữu)* là thuộc tính quan trọng nhất để dự báo KQTC cho doanh nghiệp, điều này cho thấy khi mức độ sử dụng đòn bẩy tài chính của công ty lớn hơn giá trị 3,3531 khi đó doanh nghiệp được dự báo là KQTC. Chỉ số X12 nhỏ hơn -0,1642 và X20 nhỏ hơn -0,3757 hàm ý khi *tỷ lệ thu nhập hoạt động trên giá trị sổ sách vốn cổ phần không vượt qua giá trị -0,1642* và mức *tăng trưởng trong vốn chủ sở hữu thấp hơn giá trị -0,3757* doanh nghiệp sẽ rơi vào tình trạng KQTC. Ngoài ra, doanh nghiệp cũng được xem là có khả năng KQTC cao hơn các doanh nghiệp khác nếu khả năng sinh lợi của công ty thấp, cụ thể là *tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu (X11)* nhỏ hơn giá trị 0,0586 mặc dù doanh nghiệp có *tỷ lệ thu nhập hoạt động trên giá trị sổ sách vốn cổ phần (X12)* lớn hơn -0,1642 và *tỷ lệ nợ trên vốn chủ sở hữu (X5)* lớn hơn 3,1224.

Ở thuật toán Adaboost tìm thấy *chỉ số thể hiện sự tăng trưởng vốn chủ sở hữu qua các năm (X20)* là thuộc tính quan trọng nhất để dự báo KQTC, tiếp theo là *thuộc tính X5 và X11*. Kết quả cho thấy, với mức *tăng trưởng trong vốn chủ sở hữu (X20)* của doanh nghiệp không đạt được giá trị tối thiểu là -0,3757, doanh nghiệp sẽ được dự báo là KQTC, điều này giống với dự báo của thuật toán C4.5. Ngoài ra, mặc dù *tăng trưởng trong vốn chủ sở hữu (X20)* của doanh nghiệp lớn hơn -0,3757, doanh nghiệp vẫn sẽ rơi vào KQTC khi đồng thời có *tỷ lệ nợ trên vốn chủ sở hữu*

(X5) lớn hơn 3,2136 và *tăng trưởng trong vốn chủ sở hữu (X20)* nhỏ hơn 0,1805. Tiếp sau đó, nếu *tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu (X11)* không thể lớn hơn 0,2 và *tỷ lệ nợ trên vốn chủ sở hữu (X5)* vượt quá 4,3591; doanh nghiệp sẽ có khả năng KQTC cao hơn. Bên cạnh đó, doanh nghiệp có khả năng KQTC khi *tỷ lệ tài sản cố định trên vốn dài hạn (X6)* và *tỷ số thanh toán nhanh (X2)* lần lượt vượt quá 1,32 và 0,6169.

Như vậy, mô hình cây dựa trên 2 thuật toán cho thấy quy mô vốn chủ sở hữu, đòn bẩy tài chính, khả năng sinh lợi, khả năng thanh toán lần lượt là các chỉ số đóng vai trò quan trọng trong việc dự báo KQTC của các công ty Việt Nam. Với vốn chủ sở hữu của một doanh nghiệp sụt giảm qua các năm cho thấy sự bất ổn trong quá trình hoạt động, báo hiệu khả năng rơi vào KQTC của doanh nghiệp đang tăng cao. Thứ hai, doanh nghiệp có mức độ nợ sử dụng trong cấu trúc vốn càng cao càng làm tăng rủi ro tài chính khi doanh nghiệp không đủ thanh toán các khoản lãi vay và do vậy dễ xảy ra KQTC. Thứ ba, doanh nghiệp KQTC thường có tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu thấp, điều này cho thấy doanh nghiệp hoạt động kém hiệu quả hơn các doanh nghiệp khác. Thứ tư, tài sản cố định của doanh nghiệp được tài trợ bằng các khoản vay ngắn hạn cho thấy sự không ổn định trong cấu trúc vốn của doanh nghiệp nên dễ đối diện với KQTC. Cuối cùng, khả năng thanh toán nhanh quá cao cũng làm tăng khả năng xảy ra KQTC, biểu hiện của việc doanh nghiệp nắm giữ quá nhiều tài sản có tính thanh khoản cao và làm giảm hiệu quả sử dụng vốn của doanh nghiệp.

Bảng 6

Kết quả dự báo của mô hình cây phân lớp

Thuật toán	Cấp độ	Thuộc tính	Tên thuộc tính	Ngưỡng phân lớp	Kết quả
C4.5	1	X5	Tỷ lệ nợ trên vốn chủ sở hữu	> 3,3531	KQ
	2	X12	Tỷ lệ thu nhập hoạt động trên giá trị sổ sách vốn cổ phần	≤ -0,1642	KQ

Thuật toán	Cấp độ	Thuộc tính	Tên thuộc tính	Ngưỡng phân lớp	Kiệt quệ
Adaboost	3	X20	Tăng trưởng trong vốn chủ sở hữu	$\leq -0,3757$	KQ
	3	X5	Tỷ lệ nợ trên vốn chủ sở hữu	$> 3,1224$	
	4	X11	Tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu	$\leq 0,0586$	KQ
	1	X20	Tăng trưởng trong vốn chủ sở hữu	$\leq -0,3757$	KQ
	2	X5	Tỷ lệ nợ trên vốn chủ sở hữu	$> 3,2136$	
	3	X20	Tăng trưởng trong vốn chủ sở hữu	$\leq 0,1805$	KQ
	4	X11	Tỷ lệ lợi nhuận ròng trên vốn chủ sở hữu	$\leq 0,2000$	KQ
	5	X5	Tỷ lệ nợ trên vốn chủ sở hữu	$> 4,3591$	KQ
	3	X6	Tỷ lệ tài sản cố định trên vốn dài hạn	$> 1,3200$	KQ
	4	X2	Tỷ số thanh toán nhanh	$> 0,6169$	KQ

Nguồn: Nhóm tác giả tự tổng hợp. Ghi chú: KQ: kiệt quệ

Đánh giá khả năng dự báo kiệt quệ tài chính giữa các mô hình cây

Để so sánh khả năng dự báo KQTC giữa 2 thuật toán từ mô hình cây, bài viết sử dụng lần lượt 3 tiêu chí. Kết quả Bảng 7 cho thấy, diện tích dưới đường cong ROC (AUC) của mô hình cây từ thuật toán C4.5 và AdaBoost đều lớn hơn 0,9 (cột 3); hệ số Kappa đều dương và gần bằng 1; mức độ dự báo chính xác tổng thể đều gần bằng 1. Mặc dù giá trị

AUC của thuật toán C4.5 lớn hơn một chút so với thuật toán Adaboost song, cả mức độ dự báo chính xác tổng thể lẫn hệ số Kappa của mô hình cây từ thuật toán Adaboost là cao hơn C4.5, hàm ý thuật toán Adaboost có khả năng dự báo chính xác hơn và tính ứng dụng thực tiễn tốt hơn C4.5. Chính vì vậy, thuật toán Adaboost nên được ưu tiên sử dụng trong mô hình cây phân lớp nhằm dự báo KQTC cho các doanh nghiệp Việt Nam.

Bảng 7

Sức mạnh dự báo của các mô hình cây phân lớp

Mô hình	Mức độ dự báo chính xác tổng thể	Hệ số Kappa	AUC
C4.5	0,9851	0,9414	0,994
Adaboost	0,9867	0,9477	0,991

Nguồn: Nhóm tác giả tự tổng hợp

Bằng việc sử dụng mô hình cây phân lớp được xây dựng từ thuật toán Adaboost, bài viết thu được các kết quả phù hợp với các tiêu chí nhận diện KQTC. Cụ thể, thuộc tính dự báo KQTC cao nhất là *tăng trưởng trong vốn chủ sở hữu (X20)* (khi X20 nhỏ hơn $-0,3757$ hoặc nhỏ hơn $0,1805^8$ doanh nghiệp được xác

định là rơi vào KQTC), hàm ý vốn chủ sở hữu của doanh nghiệp đang giảm mạnh, hoạt động sản xuất kinh doanh bị thua lỗ và tổng số lỗ lũy kế đã khiến vốn chủ sở hữu bị sụt giảm. Điều này phù hợp với tiêu chí doanh nghiệp bị hủy niêm yết bắt buộc hoặc tuyên bố phá sản dựa trên Quy chế niêm yết. Tiếp theo, nếu tỷ

lợi nhuận ròng trên vốn chủ sở hữu (X11) nhỏ hơn hoặc bằng 0,2 hoặc *tỷ lệ nợ trên vốn chủ sở hữu (X5)* lớn hơn 4,3591; các doanh nghiệp sẽ có khả năng KQTC, phù hợp với tiêu chí nhận diện doanh nghiệp rơi vào KQTC khi chỉ số Zmijewski⁸ của doanh nghiệp đó vượt quá giá trị không bởi chỉ số này tỷ lệ nghịch với lợi nhuận ròng, khả năng thanh toán và tỷ lệ thuận với mức độ nợ của doanh nghiệp. Ngoài ra, doanh nghiệp có khả năng KQTC khi *Tỷ lệ tài sản cố định trên vốn dài hạn (X6)* lớn hơn 1,32 thể hiện sự không ổn định trong cấu trúc vốn của doanh nghiệp vì đã tài trợ cho các tài sản cố định bằng những khoản vay ngắn hạn, và *tỷ số thanh toán nhanh (X2)* vượt quá 0,6169 thể hiện sự thiếu hiệu quả trong việc sử dụng vốn, đây là những phát hiện mới của mô hình cây phân lớp dùng để dự báo KQTC tại Việt Nam.

Kết quả này đồng tình với Kim và Upneja (2014) khi vốn chủ sở hữu có xu hướng giảm, mức độ nợ tăng cao, khả năng thanh toán quá cao và khả năng sinh lời kém, doanh nghiệp sẽ dễ rơi vào KQTC. Bên cạnh đó, bài viết còn cho thấy các doanh nghiệp Việt Nam có xu hướng tài trợ các tài sản cố định bằng những khoản nợ ngắn hạn sẽ có nguy cơ rơi vào KQTC. Nhóm tác giả cho rằng điều này bắt nguồn từ hai nguyên nhân: (1) mức độ tài trợ bằng nợ trong cấu trúc vốn của doanh nghiệp quá cao, thậm chí là sử dụng khá nhiều nợ ngắn hạn để tài trợ cho tài sản cố định làm tăng rủi ro tài chính và làm kém khả năng thanh toán các khoản nợ của doanh nghiệp, đặc biệt khi nhiều khoản nợ ngắn hạn đáo hạn cùng một lúc; (2) doanh nghiệp sử dụng vốn kém hiệu quả khi đã giữ quá nhiều tài sản có tính thanh khoản cao và hoạt động sản xuất kinh doanh của các doanh nghiệp không hiệu quả khiến doanh nghiệp bị thua lỗ.

5. Kết luận

Bài viết cho thấy mô hình cây phân lớp hoàn toàn phù hợp để dự báo KQTC cho các doanh nghiệp Việt Nam. Kết quả là bằng

chứng ủng hộ việc sử dụng mô hình cây phân lớp trong dự báo KQTC, bên cạnh các mô hình phân tích nghiên cứu trước đây. Theo đó, mô hình cây phân lớp dựa trên thuật toán Adaboost là mô hình phù hợp để dự báo KQTC cho kết quả dự báo chính xác trên 90%. Các chỉ số tăng trưởng vốn chủ sở hữu, nợ trên vốn chủ sở hữu, lợi nhuận ròng trên vốn chủ sở hữu là ba thuộc tính đóng vai trò quan trọng nhất trong việc dự báo KQTC của doanh nghiệp. Kết quả bài viết còn cho thấy các doanh nghiệp Việt Nam trong thời gian nghiên cứu bị KQTC là do thực hiện các quyết định tài trợ và quyết định đầu tư không đúng đắn xuất phát từ việc sử dụng đòn bẩy tài chính quá cao trong cấu trúc vốn của doanh nghiệp kèm theo là hoạt động sản xuất kinh doanh không hiệu quả. Những bằng chứng của nghiên cứu này là cơ sở khoa học quan trọng đối với nhà quản trị tài chính khi xây dựng các chiến lược kế hoạch tài chính trong tương lai của doanh nghiệp. Từ kết quả đạt được cho thấy các nhà quản trị tài chính cần: (1) thận trọng đối với quyết định huy động nguồn tài trợ, khai thác triệt để các nguồn vốn nội bộ, đặc biệt là từ nguồn lợi nhuận giữ lại vì bên cạnh giảm bớt các chi phí sử dụng vốn doanh nghiệp mà còn hạn chế việc sử dụng nguồn từ nợ vay, đặc biệt là nợ ngắn hạn; (2) gia tăng việc khai thác các tài sản có tính thanh khoản cao để nâng cao hiệu quả đầu tư; (3) thường xuyên kiểm tra lại các quyết định đầu tư để có chiến lược điều chỉnh kế hoạch kinh doanh kịp thời; (4) đưa mô hình cây phân lớp dựa trên thuật toán Adaboost vào phần mềm quản trị doanh nghiệp để ứng dụng trong việc dự báo sớm tình trạng KQTC. Trong tương lai, để đạt được kết quả tốt hơn, bài viết có thể tiếp tục các thử nghiệm bằng cách thay các biến phi tài chính khác hoặc thử nghiệm các điều kiện khác nhau, trong đó bao gồm các tình huống KQTC dẫn đến hoặc là nộp đơn phá sản hoặc là quay trở lại với điều kiện tài chính khỏe mạnh■

Chú thích:

- ¹ **Vấn đề quá khớp dữ liệu** - Overfitting: xảy ra khi mô hình quá tập trung vào việc đoán đúng hết tất cả các điểm dữ liệu của tập huấn luyện hiện tại, mặc dù có thể là một số đặc trưng riêng của tập dữ liệu hiện tại. Việc đó lại làm giảm khả năng dự đoán chính của mô hình trên một tập dữ liệu mới khác.
- ² **Vấn đề hộp đen** - Black box problem: hệ thống không thể cung cấp lời giải thích thích hợp cho câu trả lời mà mô hình đưa ra.
- ³ **Cắt tỉa** – pruned: tối ưu hóa quá trình phân chia tạo nút con nhằm giảm thiểu các tổn thất.
- ⁴ **Thử nghiệm chéo** – cross validation: một kỹ thuật lấy lại mẫu sử dụng nhiều mẫu nhỏ huấn luyện và nhiều mẫu thử nghiệm ngẫu nhiên từ tập dữ liệu ban đầu. Thử nghiệm chéo được ưu tiên sử dụng hơn các thủ thuật lấy mẫu hold-out vì mức độ dự báo chính xác được ước tính chính xác hơn khi cỡ mẫu nhỏ.
- ⁵ **Phương pháp Tập hợp mô hình** – Ensemble: là loại mô hình được tổng hợp từ nhiều mô hình con (weaker model) được huấn luyện độc lập. Các thuật toán của phương pháp này thường dùng để cải thiện cho nhiều loại mô hình cơ sở khác nhau như cây phân lớp, SVM, Naive Bayes,...
- ⁶ **Mô hình chỉ số Zmijewski** (Zmijewski, 1984) mô tả xác suất kiệt quệ tài chính:

$$P(B=1) = P(B^* > 0); B^* = -4,3 - 4,5ROA + 5,7FINL - 0,004LIQ;$$

Trong đó P^* : xác suất; B : 1 nếu doanh nghiệp KQTC, 0 nếu doanh nghiệp không KQTC;
 B^* : chỉ số Zmijewski;
 ROA : thu nhập ròng/tổng tài sản;
 $FINL$: tổng nợ/tổng tài sản;
 LIQ : Tài sản ngắn hạn/nợ ngắn hạn;

Nếu chỉ số Zmijewski của doanh nghiệp cụ thể B^* vượt quá 0, doanh nghiệp sẽ được nhận diện là KQTC. Chỉ số Zmijewski càng cao, khả năng KQTC càng lớn.
- ⁷ Chỉ số CGI căn cứ trên pháp luật Việt Nam, các nguyên tắc quản trị doanh nghiệp của Tổ chức Hợp tác và Phát triển Kinh tế và những thông lệ tốt trên Thế giới, bao gồm năm nội dung chính: Quyền cổ đông, đối xử bình đẳng với cổ đông, vai trò của các bên liên quan, minh bạch và công bố thông tin, trách nhiệm của Hội đồng quản trị Tổ chức Hợp tác và Nguồn: *Tổ chức Tài chính Quốc tế và Ủy ban Chứng khoán Nhà nước hợp tác thực hiện nhằm cải thiện công tác quản trị doanh nghiệp ở Việt Nam*.
- ⁸ Với điều kiện Tỷ lệ nợ trên vốn chủ sở hữu (X5) lớn hơn 3, 2136.

Tài liệu tham khảo

- Alfaro, E., Gamez, M., & Garcia, N. (2008b). "Linear discriminant analysis versus adaboost for failure forecasting". *Spanish Journal of Finance and Accounting*, 37(137), 13-32.
- Alfaro, E., Garcia, N., Gamez, M., & Elzondo, D. (2008a). "Bankruptcy forecasting: an empirical comparison of Adaboost and neural networks". *Decision support system*, 45(1), 110-122.
- Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *The Journal of Finance*, 23(4), 589-609.
- Altman, E. I., & Hotchkiss, E. (2005). *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt* (3rd ed). NJ: John Wiley & Sons.
- Altman, E., Sabato, G., & Wilson, N. (2010). "The value of non-financial information in small and medium - sized enterprise risk management". *Journal of Credit Risk*, 6(2), 95-127.
- Andrade, G., & Kaplan, S. (1998). "How costly is financial (not economic) distress? Evidence from highly leveraged transactions that became distressed". *The Journal of Finance*, 53(5), 1143-1493.
- Ball, R., & Foster, G. (1982). "Corporate financial reporting: a methodological review of empirical research". *Journal of Accounting Research*, 20, 161-248.
- Bastos, J. (2008). "Credit scoring with boosted decision trees". *Munich Personal RePEc Archive*, 8156.

- Beaver, W. H. (1966). "Financial ratios as predictors of failure, empirical research in accounting: Selected studies". *Journal of Accounting Research*, 5, 71-111.
- Denis, D., & Denis D. (1995). "Causes of financial distress following leveraged recapitalizations". *Journal of Financial Economics*, 37, 129-157
- Fitzpatrick, P. (1931). *Symptoms of Industrial Failure*. Catholic University of America Press.
- Freund, Y., & Schapire, R. (1996). *Experiments with a new boosting algorithm*. Proceedings of the Thirteenth International Conference on Machine Learning, Morgan Kauffman, San Francisco.
- He, Q., Chong, T., Li, L., & Zhang, J. (2010). "A competing risks analysis of corporate survival". *Financial Management*, 39(4), 1697-1718.
- Huang, K., Yu, H., & Chen, C. (2005). *The application of decision trees to forecast financial distress companies*. Proceedings of International Conference on Intelligent Technologies and Applied Statistics, Taipei, Taiwan.
- Kim, S., & Upneja, A. (2014). "Predicting restaurant financial distress using decision tree and adaboosted decision tree models". *Economic Modelling*, 36, 354-362.
- Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). "Predicting fraudulent financial statements with machine learning techniques". *Lecture Notes in Artificial Intelligence*, 3955, 538-542.
- Moses, D., & Liao, S. (1995). "On developing models for failure prediction". *The Journal of Commercial Bank Lending*, 69 (7), 27-38.
- Munsif, V., Raghunandan, D., Rama, V., & Singhvi, M. (2011). "Audit fees after remediation of internal control weaknesses". *Accounting Horizons*, 25(1), 87-105.
- Ohlson, J. A. (1980). "Financial ratios and the probabilistic prediction of bankruptcy". *Journal of Accounting Research*, 18(1), 109-131.
- Quinlan, J. (1996). *Bagging, boosting, and C4.5*. Proceedings, Fourteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference. AAAI Press/MIT Press, Menlo Park, 725-730.
- Shirata, C. (1998). "Financial ratios as predictors of bankruptcy in Japan: an empirical research". Proceedings of the Second Asian Pacific Interdisciplinary Research in Accounting Conference, Osaka, Japan.
- Turetsky, H., & McEwn, R. A. (2001). "An empirical investigation of firm longevity: A model of the ex ante predictors of financial distress". *Review of Quantitative Finance and Accounting*, 16, 323-343
- Whitaker, R. (1999). "The early stages of financial distress". *Journal of Economics and Finance*, 23(2), 123-133.
- Wruck, K. H. (1990). "Financial distress, reorganization and organizational efficiency". *Journal of Financial Economics*, 27, 419-444.
- Zmijewski, M. (1984). "Methodological issues related to the estimation of financial distress prediction models". *Journal of Accounting Research*, 22, 59-86.