



APPROACH TO BEHAVIOR IDENTIFICATION TRADITIONAL AND MODERN EARNINGS MANAGEMENT

Tran Thi Tuyet Van^{1*}

¹Ho Chi Minh University of Banking, Vietnam

ARTICLE INFO	ABSTRACT
<p>DOI: 10.52932/jfm.v16i1.541</p> <p><i>Received:</i> June 07, 2024</p> <p><i>Accepted:</i> August 01, 2024</p> <p><i>Published:</i> February 25, 2025</p> <p>Keywords: Machine learning; Forecast Profit management</p> <p>JEL codes: F42; F53; F55</p>	<p>Earnings Management (EM) is a strategy intentionally used by board of director to align a company's earnings targets with predetermined goals. While some view this as a useful tool in financial reporting, others view this as a fraudulent act that distorts the company's true financial status. Therefore, detecting EM is very important for financial report users. With the method of analyzing and synthesizing related research, the study focuses on methods to determine EM through different approaches, helping the author go deeper and understand more specifically the nature of the research, from traditional methods to today's modern data mining models. Each method has different advantages and disadvantages when applied in practical research. Simultaneously combining research methods can provide a comprehensive understanding of their practices and implications in current EM research.</p>

*Corresponding author:

Email: vanttt@hub.edu.vn



TIẾP CẬN PHƯƠNG PHÁP XÁC ĐỊNH HÀNH VI QUẢN TRỊ LỢI NHUẬN THEO HƯỚNG TRUYỀN THỐNG VÀ HIỆN ĐẠI

Trần Thị Tuyết Vân^{1*}

¹Trường Đại học Ngân hàng Thành phố Hồ Chí Minh

THÔNG TIN	TÓM TẮT
<p>DOI: 10.52932/jfm.v16i1.541</p> <p>Ngày nhận: 07/06/2024</p> <p>Ngày nhận lại: 01/08/2024</p> <p>Ngày đăng: 25/02/2025</p> <p>Từ khóa: Dự đoán; Học máy; Quản trị lợi nhuận</p> <p>Mã JEL: F42; F53; F55</p>	<p>Quản trị lợi nhuận (Earning Management – EM) là một chiến lược được ban quản trị cố tình sử dụng để điều chỉnh chỉ tiêu thu nhập của công ty với các mục tiêu đã xác định trước. Trong khi một số ý kiến coi đây là một công cụ hữu ích trong báo cáo tài chính thì luồng ý kiến khác lại xem đây là một hành vi lừa đảo làm sai lệch tình trạng tài chính thực sự của công ty. Do đó, việc nghiên cứu EM có ý nghĩa rất quan trọng đối với các đối tượng sử dụng Báo cáo tài chính. Với phương pháp phân tích, tổng hợp các nghiên cứu liên quan, nghiên cứu tập trung vào các phương pháp xác định EM thông qua các cách tiếp cận khác nhau, giúp tác giả đi sâu, hiểu cụ thể hơn về bản chất nghiên cứu, từ phương pháp truyền thống tới mô hình khai phá dữ liệu hiện đại. Mỗi một phương pháp đều có ưu, nhược điểm khác nhau khi ứng dụng trong nghiên cứu thực tiễn. Việc kết hợp đồng thời các phương pháp trong nghiên cứu có thể cung cấp sự hiểu biết toàn diện về thực tiễn và ý nghĩa của chúng trong nghiên cứu EM hiện nay.</p>

1. Đặt vấn đề

Các công ty niêm yết trên sàn chứng khoán cố gắng thu hút vốn từ các nhà đầu tư và chủ nợ bằng cách thường xuyên công bố kết quả hoạt động tài chính thông qua báo cáo thu nhập và đáp ứng dự báo thu nhập của các nhà phân tích tài chính (Degeorge và cộng sự, 1999). Để đáp ứng mong đợi của các bên liên quan và không muốn bỏ lỡ những kỳ vọng này, quản lý một

tổ chức có thể cố ý gây ảnh hưởng đến thu nhập được ghi nhận trong báo cáo tài chính (Rodriguez-Ariza và cộng sự, 2016). Hiện nay, bên cạnh các phương pháp truyền thống phổ biến được sử dụng trong việc phát hiện hành vi quản trị lợi nhuận như mô hình Jones (1991), mô hình của Dechow và cộng sự (1995) (mô hình Jones điều chỉnh), mô hình Kothari và cộng sự (2005), mô hình Raman và Shahrur (2008) còn có sự tích hợp các thuật toán học máy mang lại một giải pháp đầy hứa hẹn để nâng cao hiệu quả hoạt động kiểm toán, đặc biệt là trong việc phát hiện sự bất thường. EM được xác định qua các phương pháp này như thế nào

*Tác giả liên hệ:

Email: vanttt@hub.edu.vn

và liệu phương pháp học máy có thực sự ưu việt hơn phương pháp đo lường truyền thống hay không trong khi các phương pháp đều tồn tại song song ưu và nhược điểm? Đó cũng chính là mục đích hướng đến của bài nghiên cứu tổng quan này.

2. Cơ sở lý thuyết

2.1. Quản trị lợi nhuận (EM)

Thu nhập là yếu tố giải thích chính cho lợi nhuận và được các nhà phân tích, nhà đầu tư cũng như ban giám đốc coi là chỉ mục quan trọng cung cấp thông tin trong báo cáo tài chính (Bhojraj và cộng sự, 2009; Degeorge và cộng sự, 1999; Hazarika và cộng sự, 2012; Kothari & Sloan, 1992; Leuz và cộng sự, 2003). Các nhà điều hành, quản lý có thể cố ý gây ảnh hưởng đến thu nhập được ghi nhận trong báo cáo tài chính theo hướng hợp pháp hoặc bất hợp pháp (Rodriguez-Ariza và cộng sự, 2016) và hành động đó gọi là hành vi quản trị thu nhập hay quản trị lợi nhuận (EM). EM có thể tác động 2 mặt tới chất lượng báo cáo tài chính. Thứ nhất, EM che giấu tình hình tài chính thực sự của công ty mà các bên liên quan có quyền được biết (Bajra & Cadez, 2018; Gunny, 2010; Lassoued và cộng sự, 2017). EM có thể dẫn đến mất danh tiếng (Rodriguez-Ariza và cộng sự, 2016) hoặc thậm chí là các vụ kiện pháp lý thậm chí đây được xem là dấu hiệu của tham nhũng trong quản lý hoặc “nỗ lực” đánh lừa các nhà đầu tư (Gunny, 2010). Thứ hai, một số nhà nghiên cứu cũng đã chỉ ra những mặt tích cực của EM (Bajra & Cadez, 2018; Jiraporn và cộng sự, 2008). Quản trị lợi nhuận được xem là tác động tiêu cực bởi các vụ bê bối công ty gần đây và do đó, đây được xem là hoạt động gây bất lợi cho công ty. Tuy nhiên, một số nghiên cứu cho rằng quản trị lợi nhuận có thể có lợi vì nó cải thiện, nâng cao giá trị thông tin của thu nhập. Bằng chứng thực nghiệm cho thấy có mối quan hệ nghịch đảo giữa chi phí đại diện và quản lý thu nhập. Điều này có nghĩa các nhà quản lý có thể tùy ý quyết định về thu nhập để truyền

đạt thông tin tới các cổ đông cũng như công chúng và quản trị lợi nhuận dường như không mang lại lợi ích riêng cho ban quản lý (Jiraporn và cộng sự, 2008) hay việc ban quản lý cố gắng tránh giảm thu nhập và thua lỗ thông qua hành vi quản trị lợi nhuận (Burgstahler và Dichev 1997; Beatty và cộng sự, 2002)

Quản trị lợi nhuận được chia thành 2 loại: quản trị lợi nhuận thực tế (Real Earnings Management - REM) và quản trị lợi nhuận dựa trên cơ sở dồn tích (Accruals Earnings Management - AEM). Cả hai hình thức quản lý lợi nhuận đều làm tổn hại đến giá trị thông tin của báo cáo tài chính. Chính vì vậy, việc phát hiện hành vi quản trị lợi nhuận là một câu hỏi quan trọng trong nghiên cứu kế toán (Bhojraj và cộng sự, 2009; Dechow & Skinner, 2000; Efendi và cộng sự, 2007). Đa phần hoạt động quản trị lợi nhuận đều được thực hiện thông qua việc thao túng các khoản dồn tích vì các khoản dồn tích mang tính chủ quan do đó dễ dàng điều chỉnh. Do đó, nội dung bài viết chủ yếu tập trung vào các nghiên cứu xác định hành vi quản trị lợi nhuận thông qua các khoản dồn tích (AEM).

2.2. Yếu tố đo lường quản trị lợi nhuận thông qua các khoản dồn tích (AEM)

Kế toán dồn tích nhằm mục đích ghi lại các tác động kinh tế đối với một công ty có liên quan đến tiền mặt diễn ra trong các giai đoạn thay vì chỉ đơn thuần là thu chi (Dechow & Skinner, 2000). Hệ thống kế toán này cung cấp quyền kiểm soát quản lý đối với việc lựa chọn và cho phép ban quản lý sử dụng quyền quyết định vốn có trong hệ thống để xác định thu nhập được báo cáo nhằm đạt được mục đích. Nghĩa là, các nhà quản lý sử dụng phán đoán của mình dựa trên thông tin có sẵn để ước tính các khoản mục khác nhau được công bố trong báo cáo tài chính (Al-Sraheen, 2019; Fang và cộng sự, 2016).

Việc đánh giá quyền quyết định của nhà quản lý đối với thu nhập là rất quan trọng để

khám phá EM. Các khoản dồn tích thay đổi thời điểm báo cáo thu nhập và cho phép ban quản lý chuyển thu nhập giữa các kỳ báo cáo. Các khoản dồn tích bao gồm các khoản dồn tích không tùy ý (Nondiscretionary accrual – NDA) và các khoản dồn tích tùy ý (discretionary accrual – DA). Trong đó, các khoản dồn tích không tùy ý nằm ngoài tầm kiểm soát của các nhà quản lý vì đây là kết quả của các hoạt động thường ngày bình thường của công ty, trong khi các khoản dồn tích tùy ý phát sinh từ các lựa chọn của ban quản lý (Dechow và cộng sự, 1995). Theo các nghiên cứu, một lượng lớn các khoản dồn tích tùy ý cho thấy rằng một công ty đang tham gia vào AEM. Do đó, các khoản dồn tích tùy ý được xem là yếu tố để nắm bắt phạm vi của AEM.

3. Các hướng tiếp cận nghiên cứu

3.1. Tiếp cận theo hướng truyền thống – Mô hình Jones điều chỉnh

Trong các nghiên cứu trước đây cho thấy phương pháp AEM không ảnh hưởng đến hoạt động hiệu quả tài chính dài hạn của công ty (Cohen & Zarowin, 2010; Dechow và cộng sự, 2012). Dechow và cộng sự (1995) đã đánh giá hiệu quả hoạt động (dựa trên sức mạnh của các mô hình) của 05 kỹ thuật đo AEM theo các mô hình mô hình Jones (Jones, 1991), mô hình Jones điều chỉnh (Dechow và cộng sự, 1995), mô hình DeAngelo (DeAngelo, 1986), mô hình Dechow và Sloan (Dechow & Sloan, 1991) và mô hình Healy (Healy, 1985). Những phát hiện chứng minh rằng mô hình Jones điều chỉnh xác định được mức độ của AEM hiệu quả nhất. Dechow và cộng sự cho rằng thông số kỹ thuật tiêu chuẩn của Jones không nắm bắt được hết tác động của việc thao túng dựa trên doanh số bán hàng do thực tế cho thấy những thay đổi về doanh số bán hàng được coi là mang lại lợi ích, tăng lên các khoản dồn tích không tùy ý. Trong phiên bản mới hơn này, mô hình bổ sung thêm biến nợ phải thu vào mô hình gốc. Thay đổi duy nhất so với mô hình đầu tiên của Jones là những thay đổi về doanh thu được điều chỉnh phù hợp với sự thay đổi của các khoản phải thu trong thời gian diễn ra sự kiện. Phiên bản sửa

đổi đưa ra giả định rằng tất cả các thay đổi về doanh thu trả chậm trong thời gian diễn ra sự kiện là kết quả của việc quản lý thu nhập do thực tế là thu nhập được quản lý dễ dàng hơn thông qua việc thực hiện quyền quyết định đối với việc xác nhận doanh thu bán hàng trả chậm hơn là thông qua việc thực hiện quyền quyết định đối với việc bán hàng bằng tiền mặt.

Trong mô hình Jones điều chỉnh, các khoản phải thu được tính vào doanh thu vì chúng có thể liên quan đến việc điều chỉnh quản trị lợi nhuận, sử dụng biến $(\Delta REV_{it} - \Delta REC_{it})$ thay cho biến ΔREV_{it} , ngầm giả định rằng tất cả những thay đổi về doanh thu từ các khoản phải thu là do quản trị lợi nhuận

$$\frac{TA_{it}}{A_{it-1}} = \alpha_1 \frac{1}{A_{it-1}} + \alpha_2 \frac{\Delta REV_{it} - \Delta REC_{it}}{A_{it-1}} + \alpha_3 \frac{PPE_{it}}{A_{it-1}} + \varepsilon_{it}$$

$$\begin{matrix} \text{Tổng các} \\ \text{khoản dồn tích} \end{matrix} = \begin{matrix} \text{Các khoản} \\ \text{dồn tích} \\ \text{không tùy ý} \end{matrix} + \begin{matrix} \text{Các khoản} \\ \text{dồn tích} \\ \text{tùy ý} \end{matrix}$$

Trong đó:

ΔREV_{it} : chênh lệch doanh thu của công ty i trong năm t và t-1

ΔREC_{it} : chênh lệch các khoản phải thu thuần của công ty i trong năm t và t-1

PPE_{it} : Tổng tài sản, nhà xưởng và thiết bị của công ty i năm t

ε_{it} : sai số trong năm t đối với công ty i

3.2. Tiếp cận theo hướng hiện đại – Mô hình Cây quyết định

Nhóm mô hình học máy cây quyết định bao gồm các thuật toán dựa trên cấu trúc cây để thực hiện các nhiệm vụ học có giám sát như phân loại và hồi quy. Các thuật toán học có giám sát được sử dụng phổ biến nhất trong việc phát hiện hành vi thao túng quản lý thu nhập và phát hiện gian lận tài chính như phương pháp rừng ngẫu nhiên (Random forest), phương pháp cây quyết định (Decision tree), đây cũng chính là những mô hình nổi bật của hướng nghiên cứu này. Mô hình này dựa trên việc chia tập dữ liệu thành các nhóm nhỏ hơn bằng cách sử dụng các quyết định dựa trên giá trị của các thuộc

tính (features). Các thuật toán này có ưu điểm là dễ hiểu, dễ giải thích và hiệu quả trong việc xử lý các loại dữ liệu khác nhau.

Việc xây dựng mô hình học máy để phát hiện hành vi thao túng quản lý thu nhập, phát hiện gian lận tài chính thường trải qua các bước sau:

Bước 1: Thu thập dữ liệu

Đây là bước đầu tiên trong tất cả các nhiệm vụ học máy, nhằm lấy dữ liệu, cả về số lượng lẫn chất lượng. Dữ liệu này có thể là báo cáo tình hình tài chính của công ty có sẵn trên trang website chính thức của công ty hoặc hệ thống dữ liệu Fiinpro. Dữ liệu này được sử dụng để đào tạo và thử nghiệm thuật toán học máy.

Bước 2: Chuẩn bị dữ liệu

Bước này nhằm xử lý các giá trị bị thiếu, loại bỏ các phần tử nhiễu, thống nhất thang đo trong dữ liệu. Đây được coi là một bước rất quan trọng trong việc chuẩn bị sẵn sàng dữ liệu để xây dựng mô hình học máy (Machine Learning - ML) thành công và tăng cường độ chính xác của mô hình. Sau đó, chia dữ liệu thành hai phần, tức là đào tạo và kiểm tra. Phần đào tạo sẽ được sử dụng để dạy thuật toán ML, trong khi phần kiểm tra sẽ được sử dụng để đánh giá hiệu suất của mô hình đã tạo.

Các phương pháp nghiên cứu sử dụng trong nghiên cứu như:

Phương pháp Rừng ngẫu nhiên (Random forest – RF): Bản chất của thuật toán rừng ngẫu nhiên đó là không hoàn toàn dựa trên quyết định của một cây để ra quyết định mà có thể kết hợp được nhiều cây để có kết quả hiệu quả phân loại rõ ràng hơn. Do đó, RF có thể khắc phục nhược điểm là giảm lỗi sai sót trong dự đoán và đem đến hiệu suất dự đoán cao hơn. Các mô hình RF phổ biến do độ chính xác cao và chi phí tính toán tương đối thấp (Tuan Le và cộng sự, 2021)

Phương pháp cây quyết định (Decision tree): Cây quyết định là một công cụ phân loại trong học máy, nơi các quyết định được thực hiện dựa trên thuộc tính của dữ liệu. Trong bối cảnh phát

hiện hành vi gian lận, cây quyết định thường sử dụng Entropy hoặc chỉ số Gini để tối ưu hóa quá trình phân loại, nhằm phân biệt giao dịch gian lận với giao dịch hợp pháp. Cây quyết định có thể được sử dụng cho cả vấn đề phân loại và hồi quy. Thuật toán này được sử dụng để phân tích các tập dữ liệu lớn dựa trên quy tắc chia, mang lại dự đoán tốt nhất. Ngoài việc không bị ảnh hưởng bởi bất kỳ giả thuyết thống kê nào về dữ liệu mẫu, đặc điểm chính của cây quyết định là khả năng xử lý một phần dữ liệu và kiểm tra mối quan hệ tiềm năng giữa các biến đầu vào-đầu ra

Bước 3: Xây dựng mô hình

Bước này là nơi phần đào tạo dữ liệu kết nối với thuật toán ML. Thuật toán này tận dụng mô hình toán học phức tạp để tìm hiểu dữ liệu và đưa ra dự đoán.

Bước 4: Thử nghiệm mô hình

Bước này được sử dụng để xác thực mô hình đã xây dựng từ bước trước đó trong phần kiểm tra tập dữ liệu và kiểm tra hiệu suất của nó theo bất kỳ số liệu đo lường nào thông qua Ma trận nhầm lẫn (Confusion matrix), các chỉ số đánh giá từ Ma trận như độ chính xác (Accuracy), độ nhạy (Recall), Precision và điểm F1 (F1 Score) là các thước đo toán học rộng rãi nhất được sử dụng để đánh giá hiệu suất của các phương pháp ML, v.v. Nếu kết quả thu được không đạt yêu cầu, quy trình nên quay lại quá trình đào tạo hoặc thậm chí quay lại bước chuẩn bị dữ liệu

Bước 5: Hiển thị kết quả

Đây là bước cuối cùng nhằm hiển thị kết quả thu được từ các bước trước đó (bước huấn luyện và kiểm tra) trong dạng bảng hoặc biểu đồ như đường, thanh, hình tròn, v.v. bằng cách sử dụng một trong các công cụ trực quan hóa phân tích như Excel

Về biến nghiên cứu, các mô hình học máy cũng sử dụng bộ dữ liệu như bộ dữ liệu của phương pháp truyền thống. Các khoản dồn tích tùy ý được sử dụng làm biến đại diện cho quản trị lợi nhuận trong các nghiên cứu học máy này.

Khoản dồn tích không tùy ý được trừ khỏi tổng số dồn tích để đạt được số dồn tích tùy ý. Bảng cân đối kế toán và Báo cáo lưu chuyển tiền tệ là hai báo cáo cơ bản để tính tổng các khoản dồn tích. Vì dùng chung một bộ số liệu nên đây là căn cứ quan trọng để chúng ta xem xét đến tính hiệu quả của các mô hình với nhau. Liệu rằng mô hình học máy có hiệu quả hơn phương pháp truyền thống hay không, có nâng cao khả năng nhận biết kịp thời hoặc đánh giá được rủi ro phù hợp không?

4. Kết quả nghiên cứu và thảo luận

4.1. Kết quả so sánh tính hiệu quả

Học máy hiện tại đang được sử dụng trong ngân hàng và kế toán để phát hiện gian lận, chẳng hạn như chuẩn đoán khủng hoảng tài chính (Bernardo và cộng sự 2013, Hsu & Pai 2013; Verikas và cộng sự 2010); hiệu quả hoạt động của ngân hàng (Fethi & Pasiouras 2010; Shen & Tzeng 2014), dự đoán Kim và cộng sự, 2022; Tam và cộng sự, 2023). Khả năng của máy tính làm cho các kỹ thuật học máy hiệu quả hơn trong việc xử lý các vấn đề tài chính lớn. Các nghiên cứu trên cho thấy khoa học dữ liệu có khả năng dự đoán tốt hơn so với phương pháp thống kê thông thường. Các phương pháp kiểm toán truyền thống, kiểm định mô hình truyền thống có thể không còn phù hợp trong bối cảnh tích hợp liên tục giữa tiến bộ công nghệ và quá trình hoạt động của công ty.

Nghiên cứu của Dbouk và Zaarour (2017) cho thấy các mô hình toán học hoạt động tốt hơn các kiểm toán viên. Nghiên cứu chỉ ra rằng các phương pháp của kiểm toán viên thủ công rất khó phát hiện EM trên báo cáo tài chính cụ thể, tỷ lệ phân loại là (86,84%) khi sử dụng mô hình Beneish và (60,53%) khi áp dụng phương pháp của kiểm toán viên thủ công.

Nghiên cứu của Fu-Hsiang Chen và Hu Howard (2015) sử dụng phương pháp lai được đề xuất (Random Forest (RF) + Stepwise regression+ Decision tree C5.0) để sàng lọc các biến ngay từ đầu, sau đó áp dụng ba loại cây

quyết định bao gồm bộ phát hiện tương tác tự động Chi bình phương, cây phân loại và hồi quy và C5.0 để thiết lập một lập mô hình và tìm hiểu xem doanh nghiệp được thử nghiệm có bị thao túng thu nhập quá mức hay không. Kết quả cho thấy phương pháp lai được đề xuất có tỷ lệ phân loại tối ưu (tỷ lệ chính xác là 91,24%) và tỷ lệ xảy ra lỗi Loại I và lỗi Loại II thấp nhất.

Theo Almaqtari và cộng sự (2021) các nghiên cứu trước đây về tối ưu hóa quản lý thu nhập không mang lại sự tối ưu hóa lý tưởng cho việc quản lý thu nhập. Nghiên cứu hiện tại thông qua học máy mang lại những hiểu biết hữu ích để dự đoán và tối ưu hóa việc quản lý thu nhập và gian lận tài chính, có ý nghĩa quan trọng đối với các nhà hoạch định chính sách, thị trường chứng khoán, kiểm toán viên, nhà đầu tư, nhà phân tích và chuyên gia.

4.2. Ưu, nhược điểm các phương pháp

Học máy hiện tại đang được sử dụng nhiều trong ngành ngân hàng và kế toán để phát hiện gian lận, chẳng hạn như chuẩn đoán khủng hoảng tài chính (Bernardo và cộng sự 2013, Hsu & Pai 2013); hiệu quả hoạt động của ngân hàng (Fethi & Pasiouras 2010; Shen & Tzeng 2014), dự đoán bong bóng tài chính (Kim và cộng sự, 2022; Tam và cộng sự, 2023). Kết quả các nghiên cứu thực nghiệm đều cho thấy các mô hình máy học có khả năng dự đoán tốt hơn so với phương pháp thống kê thông thường. Ứng dụng mô hình học máy nói chung và Cây quyết định nói riêng trong nghiên cứu, dự đoán quản trị lợi nhuận đã được nghiên cứu nhiều trên thế giới nhưng tại thị trường Việt Nam các nghiên cứu mảng này vẫn khá hạn chế. Bên cạnh đó, theo các nghiên cứu cho thấy thao túng thu nhập phổ biến ở các nước đang phát triển hơn so với các nước phát triển như Hoa Kỳ và Châu Âu (Al-Duais và cộng sự, 2022). Alzoubi (2016) nhận thấy rằng các công ty ở các nền kinh tế mới nổi có quyền sở hữu tập trung, trong đó các cổ đông kiểm soát có thể sử dụng quyền kiểm soát của mình để đạt được lợi ích cá nhân, qua đó ảnh hưởng đến hoạt động thao túng thu nhập (Toumeah và cộng sự, 2021).

Do đó, ứng dụng máy học trong nghiên cứu dự đoán hành vi quản trị lợi nhuận là một hướng mới cần được nghiên cứu sâu hơn tại thị trường Việt Nam. Sử dụng hiệu quả sẽ giúp các công ty cũng như nhà đầu tư có cái nhìn trực quan hơn về tình hình hoạt động thực tế của các công ty.

Tính tích cực của mô hình học máy – Cây quyết định

Mô hình cây quyết định thể hiện một bước tiến đáng kể trong quá trình phân tích dữ liệu và ra quyết định, đặc biệt khi so sánh với các mô hình truyền thống. Các phương pháp thống kê thông thường hiện nay có một số giả định hạn chế như tính tuyến tính, tính chuẩn và tính độc lập của các biến đầu vào (Kuang-Hua Hu và cộng sự, 2016, Liou & Tzeng, 2012). Và những giả định phi thực tế này gây ra hạn chế trong việc khám phá các kịch bản phức tạp trong thực tế khi vận dụng mô hình truyền thống vào phân tích (Shen KY và Tzeng GH, 2014)

Điểm tích cực đầu tiên chính là sự đơn giản và dễ sử dụng của mô hình cây quyết định, xây dựng tương đối nhanh so với các phương pháp phân loại khác (Anuja Priyam và cộng sự, 2013). Các mô hình truyền thống, chẳng hạn như hồi quy tuyến tính hoặc hồi quy logistic, thường yêu cầu tiền xử lý dữ liệu mở rộng. Điều này bao gồm chuẩn hóa dữ liệu, xử lý đa cộng tuyến và chuyển đổi các mối quan hệ phi tuyến tính thành các mối quan hệ tuyến tính. Trong khi đó, xây dựng mô hình cây quyết định chỉ yêu cầu các bước tương đối đơn giản hơn như chọn tập dữ liệu, chọn các thuộc tính để phân tách và áp dụng tiêu chí phân tách để phát triển cây. Mô hình cây ít nhạy cảm hơn với quy mô, phân bố và mối quan hệ giữa các biến.

Thứ hai là không cần các hàm toán học phức tạp. Các mô hình truyền thống thường liên quan đến các hàm toán học phức tạp và các kỹ thuật tối ưu hóa để tìm ra các tham số phù hợp nhất. Ví dụ, hồi quy tuyến tính yêu cầu giải các hệ số giúp giảm thiểu tổng sai số bình phương và hồi quy logistic liên quan đến việc tối đa hóa hàm khả năng. Mặt khác, cây quyết

định sử dụng các thuật toán đơn giản như chỉ số Gini hoặc thu thập thông tin để phân chia các nút.

Thứ ba là khả năng diễn giải vấn đề. Bản chất minh họa của cây quyết định làm cho mô hình dễ hiểu, ngay cả đối với những người không phải là chuyên gia, đặc biệt đối với những người không có nền tảng toán học vững chắc. Điều này trái ngược với nhiều mô hình truyền thống có thể trừu tượng và khó diễn giải. Sự đơn giản và trực quan của cây quyết định có thể dẫn đến sự giao tiếp và hiểu biết hiệu quả hơn trong quá trình ra quyết định.

Cuối cùng là về mức độ tổng quát của đầu ra. Cây quyết định minh họa bằng đồ họa các kết quả có thể có của một loạt các lựa chọn liên quan, cho phép xem xét nhiều biến số và độ không chắc chắn. Mỗi nhánh của cây đại diện cho một quyết định, kết quả hoặc phản ứng có thể xảy ra, tạo ra một bản đồ toàn diện về khả năng cũng như nguy cơ tiềm ẩn của những lựa chọn cụ thể.

Hạn chế của mô hình học máy – Cây quyết định

Đầu tiên, mô hình cây quyết định có thể mắc phải lỗi overfitting. Một trong những nguyên nhân phổ biến là tập dữ liệu huấn luyện và tập dữ liệu kiểm tra có phân bố khác nhau khiến các quy tắc đã học trong tập dữ liệu huấn luyện không còn giá trị trong tập dữ liệu kiểm tra hoặc do khi một mô hình có quá nhiều tham số thì dữ liệu của nó không còn hiệu lực (Hung và cộng sự, 2022). Các kỹ thuật như cắt tỉa, đặt số lượng mẫu tối thiểu cần thiết tại nút lá hoặc đặt độ sâu tối đa của cây có thể giúp ngăn chặn vấn đề này.

Thứ hai, cây quyết định không tốt trong việc xử lý dữ liệu không giống nhau. Dữ liệu với các đặc trưng liên tục hoặc dữ liệu chưa được chuẩn hóa có thể gây ra vấn đề cho mô hình cây quyết định. Do đó, để tối ưu hóa kích thước của cây cũng như nguồn dữ liệu càng phong phú, tăng độ chính xác cho việc chuẩn đoán thì việc cho phép người quản lý dễ dàng cập nhật, chỉnh sửa dữ liệu là quan trọng (Hiệu & Hà, 2015), tập dữ

liệu mẫu (training) phải đủ lớn (Nam và cộng sự, 2018)

Thứ ba, cây quyết định có thể không hiệu quả nếu có quá nhiều đặc trưng. Khi số lượng đặc trưng tăng, số lượng nhánh tăng lên một cách mạnh mẽ, điều này có thể khiến cây trở nên quá phức tạp và khó hiểu. Bên cạnh đó, xu hướng là một mối quan tâm khác với cây quyết định. Nếu một số lớp chiếm ưu thế, cây quyết định sẽ tạo ra cây sai lệch. Trong nghiên cứu, các nhà nghiên cứu thường dùng ma trận đánh giá Confusion Matrix để đánh giá hiệu quả mô hình. Trong giai đoạn này có thể tồn tại 2 loại lỗi. Một gọi là “lỗi bỏ sót” khi tình huống quản trị lợi nhuận thực sự nghiêm trọng nhưng lại được mô hình dự đoán là hợp pháp. Loại 2 thì ngược lại, tình huống được phán đoán là quản trị lợi nhuận nhưng trong thực tế nó là hoạt động hợp pháp. Trong 2 loại lỗi trên thì loại 1 được xem là nguy hiểm vì mô hình không những không phát hiện ra mà còn cho phép hoạt động này tiếp tục không bị phát hiện. Vì vậy, nên cân bằng tập dữ liệu trước khi đưa vào cây quyết định.

Thứ tư là thiếu sự ổn định trong những thay đổi nhỏ trong dữ liệu huấn luyện (Quinlan, 1993). Những thay đổi nhỏ trong dữ liệu có thể dẫn đến một cây hoàn toàn khác. Vấn đề này có thể được giảm thiểu bằng cách sử dụng các phương pháp tổng hợp, xây dựng nhiều cây quyết định (Rừng ngẫu nhiên) và tổng hợp kết quả của chúng.

Cuối cùng, đối với mô hình học máy yêu cầu cao về số lượng dữ liệu nghiên cứu, khi cỡ mẫu lớn đủ, dữ liệu nghiên cứu có thể được chia thành các tập dữ liệu huấn luyện và xác nhận, sử dụng tập dữ liệu huấn luyện để xây dựng mô hình cây quyết định và một tập dữ liệu xác thực để quyết định kích thước cây thích hợp cần thiết để đạt được mô hình cuối cùng tối ưu (Song & Lu, 2015)

Mặc dù, các mô hình truyền thống có điểm mạnh và trường hợp sử dụng cụ thể riêng nhưng tính đơn giản và khả năng tiếp cận của mô hình

cây quyết định khiến chúng trở thành công cụ có giá trị trong bối cảnh khoa học dữ liệu và học máy không ngừng phát triển hiện nay.

5. Kết luận

Các hệ thống kiểm toán truyền thống với hạn chế về thời gian, nguồn nhân lực, chi phí có thể gặp khó khăn trong việc phát hiện các hoạt động bất thường trong dữ liệu tài chính rộng lớn và phức tạp. Trong khi đó, việc phát triển mô hình dự đoán EM được xem là hữu ích trong giai đoạn hiện nay khi chi phí thấp hơn và việc khai thác dữ liệu diễn ra nhanh hơn. Tóm lại, nghiên cứu về quản trị lợi nhuận được hưởng lợi rất nhiều từ sự kết hợp giữa phương pháp nghiên cứu hiện đại và truyền thống. Bằng cách tích hợp các phương pháp tiếp cận này, các nhà nghiên cứu có thể đạt được sự hiểu biết toàn diện và sâu sắc hơn về thực tiễn quản lý lợi nhuận. Khi công nghệ kỹ thuật số tiếp tục phát triển, sự kết hợp giữa các phương pháp hiện đại và truyền thống sẽ đóng một vai trò quan trọng trong việc thúc đẩy nghiên cứu về quản trị lợi nhuận và nâng cao tính trung thực của báo cáo tài chính hiện nay. Quá trình kết hợp có thể tiến hành tuần tự như sau:

Đầu tiên, các mô hình truyền thống có thể được sử dụng như một công cụ sàng lọc sơ bộ để xác định các công ty có thể tham gia quản trị lợi nhuận. Những mô hình này cung cấp một cách đơn giản và hiệu quả để xác định các vấn đề tiềm ẩn dựa trên các tỷ lệ tài chính và đòn bích đã được thiết lập.

Tiếp theo, các mô hình hiện đại có thể được áp dụng cho các công ty được xác định để phân tích sâu hơn. Các thuật toán học máy có thể phân tích dữ liệu một cách toàn diện hơn, xác định các mẫu và mối quan hệ phức tạp có thể không rõ ràng thông qua phương pháp truyền thống. Các công cụ phân tích dữ liệu có thể nâng cao hơn nữa khả năng phân tích này bằng cách tích hợp các nguồn dữ liệu bổ sung và cung cấp cái nhìn toàn diện về tình hình tài chính của công ty.

Tài liệu tham khảo

- AL-Duais và cộng sự (2022). Ownership structure and real earnings management: evidence from an emerging market. *Journal of Accounting in Emerging Economies*, 12(2), 380-404. © Emerald Publishing Limited 2042-1168. DOI 10.1108/JAEE-01-2021-0008
- Almaqtari, F. A., Farhan, N. H., Salmony, M. Y., Al-Ahdal, W. M., & Mishra, N. (2021, November). Earning management estimation and prediction using machine learning: A systematic review of processing methods and synthesis for future research. In *2021 International Conference on Technological Advancements and Innovations (ICTAI)* (pp. 291-298). IEEE. <https://doi.org/10.1109/ICTAI53825.2021.9673157>
- Al-Sraheen, D. A. A. D. O. (2019). The role of the audit committee in moderating the negative effect of non-audit services on earnings management among industrial firms listed on the Amman Stock Exchange. *Afro-Asian Journal of Finance and Accounting*, 9(3), 349-361. <https://doi.org/10.1504/AJFA.2019.100983>
- Alzoubi, E. S. S. (2017). Audit quality, debt financing, and earnings management: Evidence from Jordan. *Journal of International Accounting, Auditing and Taxation*. <https://doi.org/10.1016/j.intaccudtax.2017.12.001>
- Bajra, U., & Cadez, S. (2018). The impact of corporate governance quality on earnings management: Evidence from European companies cross-listed in the US. *Australian Accounting Review*, 28(2), 152-166. <https://doi.org/10.1111/auar.12176>
- Beatty, A.L., Ke, B. and Petroni, K.R. (2002). Earnings Management to Avoid Earnings Declines Across Publicly and Privately Held Banks. *Accounting Review*, 77(3), 547-570. <https://doi.org/10.2308/accr.2002.77.3.547>
- Bernardo D., Hagraş H., Tsang E. (2013). A genetic type-2 fuzzy logic based system for the generation of summarised linguistic predictive models for financial applications. *Soft Comput*, 17(12), 2185-2201. <https://doi.org/10.1007/s00500-013-1102-y>
- Bhojraj, S., Hribar, P., Picconi, M., and McNinnis, J. (2009). Making sense of cents: An examination of firms that marginally miss or beat analyst forecasts, *The Journal of Finance*, 64(5), 2361-2388. <https://doi.org/10.1111/j.1540-6261.2009.01503.x>
- Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, 24(1), 99-126. [https://doi.org/10.1016/S0165-4101\(97\)00017-7](https://doi.org/10.1016/S0165-4101(97)00017-7)
- Chen, F. H., & Howard, H. (2016). An alternative model for the analysis of detecting electronic industries earnings management using stepwise regression, random forest, and decision tree. *Soft Computing*, 20(5), 1945-1960. <https://doi.org/10.1007/s00500-015-1616-6>
- Cohen, D. A., & Zarowin, P. (2010). Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics*, 50(1), 2-19. <https://doi.org/10.1016/j.jacceco.2010.01.002>
- Đang Van Nam, Nguyen Thi Phuong Bac, Nguyen Thi Hai Yen (2018). Research and application of decision tree in the recruitment problem. *ERSD 2018*. [https://qlkh.humg.edu.vn/CongBo/Download/2373?FileName=C%C3%B4ng%20ng%E1%BB%87%20th%C3%B4ng%20tin%20v%C3%A0%E1%BB%A8ng%20d%E1%BB%A5ng%20\(2\).pdf](https://qlkh.humg.edu.vn/CongBo/Download/2373?FileName=C%C3%B4ng%20ng%E1%BB%87%20th%C3%B4ng%20tin%20v%C3%A0%E1%BB%A8ng%20d%E1%BB%A5ng%20(2).pdf)
- DeAngelo, L. E. (1986). Accounting numbers as market valuation substitutes: A study of management buyouts of public stockholders. *The Accounting Review*, 61(3), 400-420. <https://www.jstor.org/stable/247149>
- Dechow, P. M., & Skinner, D. J. (2000). Earnings management: Reconciling the views of accounting academics, practitioners, and regulators. *Accounting Horizons*, 14(2), 235-250.
- Dechow, P. M., & Sloan, R. G. (1991). Executive incentives and the horizon problem: An empirical investigation. *Journal of Accounting and Economics*, 14(1), 51-89. [https://doi.org/10.1016/0167-7187\(91\)90058-S](https://doi.org/10.1016/0167-7187(91)90058-S)
- Dechow, P. M., Hutton, A. P., Kim, J. H., & Sloan, R. G. (2012). Detecting earnings management: A new approach. *Journal of Accounting Research*, 50(2), 275-334. <https://doi.org/10.1111/j.1475-679X.2012.00449.x>
- Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings Management. *Accounting Review*, 70(2), 193-225. <https://www.jstor.org/stable/248303>

- DeGeorge, F., Patel, J., and Zeckhauser, R. (1999). Earnings management to exceed thresholds. *The Journal of Business*, 72(1), 1-33. <https://doi.org/10.1086/209601>
- Efendi, J., Srivastava, A. & Swanson, E.P. (2007). Why do corporate managers misstate financial statements? The role of option compensation and other factors. *Journal of Financial Economics*, 85(3), 667-708. <https://doi.org/10.1016/j.jfineco.2006.05.009>
- Fang, V. W., Huang, A. H., & Karpoff, J. M. (2016). Short selling and earnings management: A controlled experiment. *The Journal of Finance*, 71(3), 1251-1294.
- Fethi MD., Pasiouras F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *The European Journal of Operational Research*, 204(2), 189-198. <https://doi.org/10.1016/j.ejor.2009.08.003>
- Fethi, M. D., & Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European journal of operational research*, 204(2), 189-198. <https://doi.org/10.1016/j.ejor.2009.08.003>
- Gunny, K. A. (2010). The relation between earnings management using real activities manipulation and future performance: Evidence from meeting earnings benchmarks. *Contemporary Accounting Research*, 27(3), 855-888. <https://doi.org/10.1111/j.1911-3846.2010.01029.x>
- Hazarika, S., Karpoff, J. M., & Nahata, R. (2012). Internal corporate governance, CEO turnover, and earnings management. *Journal of financial Economics*, 104(1), 44-69. <https://doi.org/10.1016/j.jfineco.2011.10.011>
- Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics*, 7(1-3), 85-107. [https://doi.org/10.1016/0165-4101\(85\)90029-1](https://doi.org/10.1016/0165-4101(85)90029-1)
- Hsu, M. F., & Pai, P. F. (2013). Incorporating support vector machines with multiple criteria decision making for financial crisis analysis. *Quality & Quantity*, 47, 3481-3492. <https://doi.org/10.1007/s11135-012-9735-y>
- Hu, K. H., Chen, F. H., & Chang, W. J. (2016). Application of correlation-based feature selection and decision tree to detect earnings management and accounting fraud relationship. *ICIC Express Letters, Part B: Applications*, 7(11), 2361-2366. <http://www.icalb.org/ellb/contents/2016/11/elb-07-11-11.pdf>
- Jiraporn, P., Miller, G. A., Yoon, S. S., & Kim, Y. S. (2008). Is earnings management opportunistic or beneficial? An agency theory perspective. *International Review of Financial Analysis*, 17(3), 622-634. <https://doi.org/10.1016/j.irfa.2006.10.005>
- Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, 29(2), 193-228. <https://doi.org/10.2307/2491047>
- Kothari, S. P., & Sloan, R. G. (1992). Information in prices about future earnings: Implications for earnings response coefficients. *Journal of Accounting and Economics*, 15(2-3), 143-171. [https://doi.org/10.1016/0165-4101\(92\)90016-U](https://doi.org/10.1016/0165-4101(92)90016-U)
- Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics*, 39(1), 163-197. <https://doi.org/10.1016/j.jacceco.2004.11.002>
- Lassoued, N., Attia, M. B. R., & Sassi, H. (2017). Earnings management and ownership structure in emerging market: Evidence from banking industry. *Managerial Finance*, 43(10), 1117-1136. <https://doi.org/10.1108/MF-11-2015-0312>
- Le, T., Pham, T., & Dao, S. (2021). Using Machine Learning to Predict the Defaults of Credit Card Clients. *Fintech with Artificial Intelligence, Big Data, and Blockchain*, 133-152. https://doi.org/10.1007/978-981-33-6137-9_4
- Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: an international comparison. *Journal of Financial Economics*, 69(3), 505-527. [https://doi.org/10.1016/S0304-405X\(03\)00121-1](https://doi.org/10.1016/S0304-405X(03)00121-1)
- Liou, J. J., & Tzeng, G. H. (2012). Comments on “Multiple criteria decision making (MCDM) methods in economics: an overview”. *Technological and Economic Development of Economy*, 18(4), 672-695. <https://doi.org/10.3846/20294913.2012.753489>
- Ngoc Hung Dang, Thi Thuy Van Vu, Thi Nhat Le Dao (2022). Accounting information and stock returns in Vietnam securities market: Machine learning approach. *Contabilidad Y Negocios* 17(33), 94-118. <https://doi.org/10.18800/contabilidad.202201.004>

- Nguyễn Văn Hiệu và Đỗ Thị Thu Hà (2015). A System for Diagnosing Autism Based on the Decision Tree. *Tạp chí Khoa học và Công nghệ-Đại học Đà Nẵng*, 11(96), 101-104. <https://jst-ud.vn/jst-ud/article/view/3415>
- Priyam, A., Abhijeeta, G. R., Rathee, A., & Srivastava, S. (2013). Comparative analysis of decision tree classification algorithms. *International Journal of Current Engineering and Technology*, 3(2), 334-337. <http://inpressco.com/category/ijcet>
- Quinlan, J. R. (1993). *Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
- Raman, K., & Shahrur, H. (2008). Relationship-specific investments and earnings management: Evidence on corporate suppliers and customers. *The Accounting Review*, 83(4), 1041-1081. <https://doi.org/10.2308/accr.2008.83.4.1041>
- Rodriguez-Ariza, L., Martínez-Ferrero, J., & Bermejo-Sanchez, M, (2016). Consequences of earnings management for corporate reputation: Evidence from family firms. *Accounting Research Journal*, 29(4), 457-474. <https://doi.org/10.1108/ARJ-02-2015-0017>
- Shen, K. Y., & Tzeng, G. H. (2015). A decision rule-based soft computing model for supporting financial performance improvement of the banking industry. *Soft Computing*, 19, 859-874. <https://doi.org/10.1007/s00500-014-1413-7>
- Song, Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130-135. <https://doi.org/10.11919/j.issn.1002-0829.215044>
- Tam, P. H., Tram, N. D. L., Anh, N. T. N., Nghia, N. Q. T., Linh, H. T., & Van Thanh, T. (2023). Application of machine learning in classification of overinvestment: Evidence from listed firms in Vietnam stock exchange market. *Science & Technology Development Journal: Economics-Law & Management*, 7(4), 4814-4833. <https://doi.org/https://doi.org/10.32508/stdjelm.v7i4.1269>
- Toumeh (2021). The Influence of Enterprise Resource Planning System Implementation on Accrual-Based Earnings Management: Empirical Evidence from Jordan. *Global Business Review*, DOI: 10.1177/09721509221116692 journals.sagepub.com/home/gbr
- Tran, K. L., Le, H. A., Nguyen, T. H., & Nguyen, D. T. (2022). Explainable machine learning for financial distress prediction: Evidence from Vietnam. *Data*, 7(11), 160. <https://doi.org/10.3390/data7110160>
- Verikas A., Kalsyte Z., Bacauskiene M., Gelzinis A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computer*, 14(9), 995-1010. <https://doi.org/10.1007/s00500-009-0490-5>
- Zaarour, B. D. I. (2017). Financial statements earnings manipulation detection using a layer of machine learning. *International Journal of Innovation, Management and Technology*, 8(3), 172-179. <https://www.ijimt.org/vol8/723-DT0010.pdf>