

# Nghiên cứu ứng dụng trí tuệ nhân tạo và dữ liệu lớn trong phát hiện gian lận tài chính

Nguyễn Ngọc Hiền<sup>1</sup>, Nguyễn Thị Kim Chi<sup>2</sup>

Ngày nhận bài: 20/02/2026 | Ngày gửi phản biện: 24/02/2026 | Ngày duyệt đăng: 12/3/2026

**Tóm tắt:** Gian lận tài chính là hoạt động gây ảnh hưởng nghiêm trọng đến sự ổn định của thị trường tài chính và niềm tin của nhà đầu tư. Các phương pháp phát hiện gian lận truyền thống thường không hiệu quả và không đáp ứng được các phương thức gian lận thay đổi nhanh chóng. Sự phát triển của trí tuệ nhân tạo (AI) và dữ liệu lớn đã cung cấp các giải pháp mới cho việc phát hiện gian lận tài chính. Bài viết nghiên cứu ứng dụng AI và dữ liệu lớn trong phát hiện gian lận tài chính, đồng thời phân tích các thách thức và hướng phát triển trong tương lai.

**Từ khóa:** AI, dữ liệu lớn, gian lận tài chính.

## Research on the Application of Artificial Intelligence and Big Data in Financial Fraud Detection

**Abstract:** Financial fraud is an activity that seriously impacts the stability of financial markets and investor confidence. Traditional fraud detection methods are often ineffective and fail to keep up with rapidly changing fraud techniques. The development of artificial intelligence (AI) and big data has provided new solutions for detecting financial fraud. This article examines the application of AI and big data in financial fraud detection, analyzing the challenges and future development directions.

**Keywords:** AI, big data, financial fraud.

### 1. Giới thiệu

Trong bối cảnh kinh tế hiện nay, gian lận tài chính đã trở thành chủ đề nổi bật và ảnh hưởng nghiêm trọng đến các nhà đầu tư. Với sự phát triển liên tục của thị trường tài chính và những tiến bộ vượt bậc của khoa học công nghệ, gian lận tài chính đã thể hiện những đặc điểm đa dạng, khó lường. Việc phát hiện gian lận tài chính đã nhận được sự quan tâm đáng kể từ các nhà đầu tư, tổ chức tài chính và cơ quan quản lý Nhà nước. Các phương pháp phát hiện gian lận tài chính truyền thống chủ yếu dựa vào các quy tắc và kiểm toán thủ công nên không còn đáp ứng được nhu cầu phát hiện gian lận kịp thời, hiệu quả và chính xác. Hạn chế của phương pháp này là không thể xử lý lượng lớn dữ liệu tài chính, không thể phát hiện các mô hình gian lận phức tạp và dễ bị ảnh hưởng bởi hành vi gian lận. Do đó, cần xây dựng giải pháp phát hiện gian lận thông minh và hiệu quả hơn để nâng cao độ chính xác và hiệu quả của việc phát hiện gian lận, đồng thời giảm thiểu rủi ro và tổn thất do gian lận gây ra.

### 2. Trí tuệ nhân tạo và dữ liệu lớn trong phát hiện gian lận tài chính

Trí tuệ nhân tạo là lĩnh vực khoa học máy tính nghiên cứu các hệ thống có khả năng học hỏi, suy luận và ra quyết định tự động dựa trên dữ liệu. Dữ liệu lớn (Bigdata) là hệ sinh thái các giải pháp, công cụ và kỹ thuật tiên tiến được thiết kế để thu thập, lưu trữ, xử lý và phân tích các tập dữ liệu có quy mô

<sup>1</sup> Khoa Kế toán - Tài chính, Đại học Điện lực; Email: hiennn@epu.edu.vn,

<sup>2</sup> Khoa Kế toán - Tài chính, Đại học Điện lực; Email: chintk@epu.edu.vn

lớn, tốc độ cao và đa dạng. Trong phát hiện gian lận, Bigdata đóng vai trò tích hợp dữ liệu, hỗ trợ phân tích giữa các nguồn dữ liệu và cung cấp nền tảng xử lý.

Gian lận tài chính là hành vi cố ý, trái quy định của pháp luật nhằm mục đích thu được lợi ích tài chính trái phép. Một mặt, công nghệ dữ liệu lớn cung cấp nền tảng dữ liệu cho trí tuệ nhân tạo. Lĩnh vực tài chính đã tạo ra một lượng dữ liệu khổng lồ, bao gồm dữ liệu giao dịch, dữ liệu hành vi người dùng, dữ liệu thị trường... Công nghệ dữ liệu lớn có thể lưu trữ, quản lý và xử lý dữ liệu này một cách hiệu quả, cung cấp nguồn tài nguyên dữ liệu dồi dào cho trí tuệ nhân tạo (Choi, D., & Lee, K., 2018). Mặt khác, công nghệ trí tuệ nhân tạo học các mô hình hành vi gian lận từ dữ liệu tài chính khổng lồ thông qua các phương pháp như học máy và học sâu. Trong học máy, các thuật toán có thể xây dựng các mô hình và huấn luyện chúng dựa trên các đặc điểm dữ liệu của các trường hợp gian lận đã biết và các trường hợp không gian lận, từ đó học được các đặc điểm và mô hình của hành vi gian lận (Zhou, H., Sun, G., Fu, S., Fan, X., Jiang, W., Hu, S., & Li, L., 2020). Trong học sâu, bằng cách xây dựng mô hình mạng nơ-ron đa lớp, hệ thống có thể tự động trích xuất và học các đặc điểm nâng cao của dữ liệu, từ đó đạt được khả năng phát hiện gian lận phức tạp và tinh vi hơn. Ngoài ra, công nghệ trí tuệ nhân tạo sử dụng công nghệ xử lý luồng dữ liệu thời gian thực, cho phép hệ thống giám sát các giao dịch tài chính và hành vi người dùng trong thời gian thực, đồng thời cung cấp cảnh báo kịp thời về hành vi bất thường. Bằng cách kết hợp khả năng xử lý dữ liệu thời gian thực của công nghệ dữ liệu lớn và khả năng phân tích thông minh của công nghệ trí tuệ nhân tạo, hiệu quả và độ chính xác của việc phát hiện gian lận có thể được cải thiện một cách hiệu quả. Đồng thời, gian lận tài chính thường liên quan đến nhiều nguồn dữ liệu và chiều thông tin. Dựa trên khả năng phân tích tương quan của công nghệ trí tuệ nhân tạo, hệ thống có thể khai thác mối tương quan và liên hệ giữa các nguồn dữ liệu khác nhau và phát hiện ra các mạng lưới và mô hình gian lận tiềm năng (Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J., & Gao, Y., 2021).

### **3. Ứng dụng công nghệ trí tuệ nhân tạo trong phát hiện gian lận tài chính**

#### **3.1. Phân tích yêu cầu**

Việc ứng dụng công nghệ trí tuệ nhân tạo trong phát hiện gian lận tài chính bắt đầu bằng việc phân tích các nhu cầu khác nhau. Điều này bao gồm việc hiểu được sự phức tạp của thị trường tài chính và các phương thức gian lận liên tục thay đổi, cũng như nhận thức về hiệu quả thấp của các phương pháp phát hiện gian lận truyền thống. Đồng thời, đối với các tổ chức tài chính và người dùng, việc phát hiện và ngăn chặn hành vi gian lận kịp thời là rất quan trọng để duy trì sự ổn định của thị trường tài chính và quyền lợi của người dùng. Do đó, phân tích yêu cầu tập trung vào hiệu quả cao, độ chính xác và hiệu suất thời gian thực hiện, yêu cầu hệ thống phải có khả năng xử lý dữ liệu quy mô lớn và xác định các mô hình hành vi gian lận đa dạng. Trong tình huống này, việc ứng dụng công nghệ trí tuệ nhân tạo đã trở thành một lựa chọn không thể thiếu. Nó có thể học hỏi và phát hiện các mô hình gian lận tiềm năng từ dữ liệu khổng lồ thông qua các phương pháp như học máy và học sâu, từ đó đạt được một hệ thống phát hiện gian lận tài chính thông minh và hiệu quả.

#### **3.2. Kỹ thuật trích chọn đặc trưng**

Trong phát hiện gian lận tài chính, kỹ thuật trích chọn đặc trưng bao gồm phân tích chuyên sâu dữ liệu giao dịch tài chính và dữ liệu hành vi người dùng, trích xuất các đặc trưng có ý nghĩa để phát hiện gian lận. Các đặc điểm liên quan được thu thập có thể bao gồm nhiều thông tin khác nhau như số tiền giao dịch, tần suất giao dịch, địa điểm giao dịch, thời gian giao dịch và các mẫu hành vi lịch sử của người dùng. Thông qua kỹ thuật đặc trưng, dữ liệu thô có thể được chuyển đổi thành các biểu diễn đặc trưng mà mô hình có thể hiểu và sử dụng, từ đó cải thiện hiệu suất và hiệu quả của các mô hình phát hiện gian lận. Mục tiêu của kỹ thuật đặc trưng là chọn các đặc trưng đại diện, liên quan và phân biệt, đồng thời xem xét các hiệu ứng tương quan và tương tác giữa các đặc trưng, để tối đa hóa khả năng dự đoán của mô hình. Trong các ứng dụng thực tế, kỹ thuật đặc trưng cần xem xét toàn diện các yêu cầu

kinh doanh, đặc điểm dữ liệu và thuật toán của mô hình. Thông qua việc hiểu sâu và khai thác dữ liệu, một tập hợp đặc trưng hiệu quả phù hợp cho việc phát hiện gian lận tài chính được xây dựng, cung cấp nền tảng đáng tin cậy cho việc mô hình hóa và dự đoán tiếp theo.

### **3.3. Xây dựng mô hình**

Trong phát hiện gian lận tài chính, xây dựng mô hình là khâu cốt lõi của ứng dụng công nghệ trí tuệ nhân tạo. Thông qua xây dựng mô hình, các công nghệ học máy và học sâu được sử dụng để khám phá các mẫu gian lận tiềm năng từ dữ liệu tài chính, nhằm đạt được một hệ thống phát hiện gian lận thông minh và hiệu quả. Trong quá trình xây dựng mô hình, cần phải lựa chọn các thuật toán và cấu trúc mô hình phù hợp, chẳng hạn như hồi quy logistic, cây quyết định, máy vectơ hỗ trợ, mạng nơ-ron..., và huấn luyện cũng như tối ưu hóa mô hình để nâng cao khả năng khái quát hóa và độ chính xác. Đồng thời, cần xem xét kết quả của kỹ thuật trích xuất đặc trưng và đưa các đặc trưng có ý nghĩa đã trích xuất vào mô hình để huấn luyện. Trong quá trình xây dựng mô hình, cần liên tục đánh giá hiệu suất và hiệu quả của mô hình, điều chỉnh và tối ưu hóa dựa trên nhu cầu thực tế và thông tin phản hồi để đảm bảo mô hình có thể xác định và dự đoán chính xác hành vi gian lận, bảo vệ sự ổn định của thị trường tài chính cũng như quyền và lợi ích của người dùng.

### **3.4. Giám sát và cảnh báo sớm**

Thông qua hệ thống giám sát và cảnh báo sớm, công nghệ xử lý luồng dữ liệu thời gian thực được sử dụng để liên tục giám sát dữ liệu giao dịch tài chính và hành vi người dùng, kịp thời phát hiện hành vi giao dịch bất thường và thực hiện các biện pháp cảnh báo sớm để ngăn ngừa rủi ro gian lận. Hệ thống giám sát và cảnh báo sớm tương ứng kết hợp công nghệ học máy và học sâu, có thể xác định các mẫu và mô hình bất thường trong luồng dữ liệu theo thời gian thực và cung cấp cảnh báo sớm về hành vi gian lận tiềm ẩn. Khi hệ thống phát hiện các giao dịch đáng ngờ hoặc hành vi bất thường, nó sẽ ngay lập tức kích hoạt cơ chế cảnh báo, gửi thông báo cảnh báo đến các nhân viên có liên quan để kịp thời thực hiện các biện pháp điều tra và xử lý. Việc triển khai hệ thống giám sát và cảnh báo sớm không chỉ giúp các tổ chức tài chính phát hiện và ngăn chặn hành vi gian lận kịp thời, bảo vệ an ninh tài sản của khách hàng mà còn cải thiện tính minh bạch và ổn định của thị trường tài chính, tăng cường niềm tin thị trường và cung cấp sự đảm bảo vững chắc cho sự phát triển bền vững của ngành tài chính.

## **4. Ứng dụng công nghệ dữ liệu lớn trong phát hiện gian lận tài chính**

### **4.1. Lưu trữ dữ liệu phân tán**

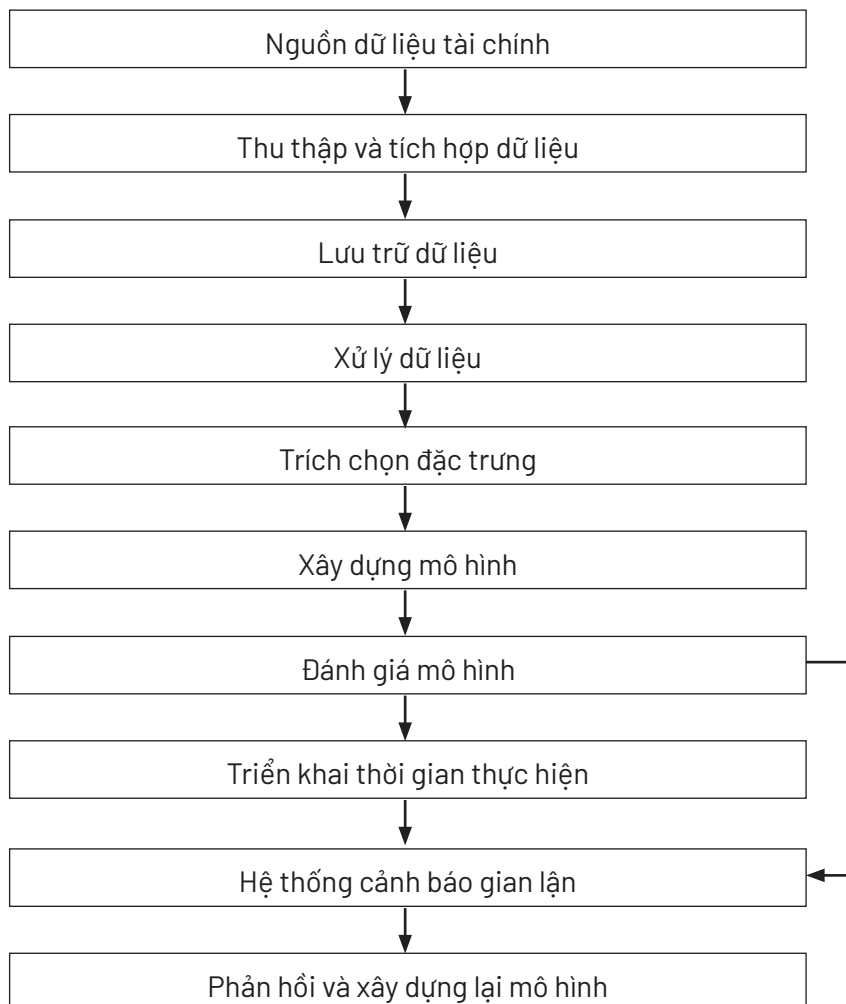
Trong việc phát hiện gian lận tài chính, việc ứng dụng công nghệ dữ liệu lớn là rất quan trọng và lưu trữ dữ liệu phân tán là một thành phần quan trọng. Lĩnh vực tài chính đã tạo ra một lượng lớn dữ liệu giao dịch, dữ liệu hành vi người dùng,... Những dữ liệu này cần được lưu trữ và quản lý hiệu quả để hỗ trợ công việc phân tích và khai thác tiếp theo. Công nghệ lưu trữ dữ liệu phân tán có thể đáp ứng nhu cầu lưu trữ của ngành tài chính đối với lượng dữ liệu khổng lồ và có các đặc điểm về độ tin cậy cao, khả năng mở rộng cao (Yu, T. R., & Song, X., 2021). Công nghệ lưu trữ dữ liệu phân tán có thể tận dụng tối đa không gian lưu trữ của mỗi máy chủ trong cụm bằng cách lưu trữ dữ liệu theo cách phân tán trên nhiều máy chủ, do đó giải quyết vấn đề dung lượng lưu trữ hạn chế trên một máy chủ duy nhất. Kiến trúc lưu trữ này có thể hỗ trợ lưu trữ dữ liệu quy mô lớn và đáp ứng nhu cầu lưu trữ của các tổ chức tài chính đối với lượng dữ liệu khổng lồ. Bên cạnh đó, công nghệ lưu trữ dữ liệu phân tán có độ tin cậy cao. Trong các hệ thống lưu trữ phân tán, dữ liệu thường được sao chép thành nhiều bản và được lưu trữ trên các nút khác nhau. Khi một nút bị lỗi, hệ thống vẫn có thể lấy dữ liệu từ các nút khác, đảm bảo độ tin cậy và tính khả dụng của dữ liệu, điều này rất quan trọng đối với ngành tài chính. Ngoài ra, công nghệ lưu trữ phân tán dữ liệu cũng có khả năng mở rộng cao. Với sự tăng trưởng liên tục của dữ liệu tài chính, hệ

thống lưu trữ phân tán có thể tự động mở rộng dung lượng lưu trữ theo nhu cầu mà không cần thời gian ngừng hoạt động hoặc di chuyển dữ liệu, đảm bảo hoạt động liên tục và hiệu suất ổn định của hệ thống.

#### 4.2. Tài nguyên lưu trữ đám mây

Trong việc phát hiện gian lận tài chính, ứng dụng công nghệ dữ liệu lớn không chỉ giới hạn ở việc lưu trữ dữ liệu phân tán mà còn bao gồm việc gọi các tài nguyên lưu trữ đám mây. Tài nguyên lưu trữ đám mây đề cập đến các dịch vụ lưu trữ dữ liệu được lưu trữ trên các nền tảng điện toán đám mây, chẳng hạn như Amazon S3, Google Cloud Storage, Microsoft Azure Blob Storage, v.v... Các dịch vụ lưu trữ đám mây liên quan cung cấp các giải pháp lưu trữ có độ tin cậy cao, có khả năng mở rộng và linh hoạt, cung cấp các phương pháp lưu trữ và quản lý dữ liệu thuận tiện cho các tổ chức tài chính. Mặt khác, việc gọi các tài nguyên lưu trữ đám mây cho phép các tổ chức tài chính lưu trữ dữ liệu trên đám mây mà không cần phải mua và bảo trì thiết bị phần cứng đắt tiền, giảm chi phí lưu trữ. Các tổ chức tài chính có thể tự động điều chỉnh dung lượng lưu trữ theo nhu cầu thực tế và mở rộng linh hoạt theo sự tăng trưởng của khối lượng dữ liệu, tránh lãng phí tài nguyên hoặc các vấn đề thiếu hụt do thay đổi nhu cầu lưu trữ. Mặt khác, các nhà cung cấp dịch vụ điện toán đám mây thường áp dụng công nghệ sao lưu đa bản và dự phòng dữ liệu để đảm bảo an ninh và độ tin cậy của dữ liệu.

**Hình 1: Quy trình ứng dụng AI và Bigdata trong phát hiện gian lận tài chính**



*Nguồn: Tác giả đề xuất*

Ngoài ra, các dịch vụ lưu trữ đám mây thường có các thỏa thuận mức dịch vụ với độ khả dụng cao, đảm bảo các tổ chức tài chính có thể truy cập và gọi dữ liệu được lưu trữ trên đám mây mọi lúc mọi nơi, đảm bảo hoạt động kinh doanh liên tục. Thêm vào đó, việc sử dụng tài nguyên lưu trữ đám mây cũng cho phép các tổ chức tài chính tận dụng khả năng điện toán phân tán và linh hoạt do nền tảng điện toán đám mây cung cấp để xử lý và phân tích dữ liệu quy mô lớn.

## 5. Kết luận

Tóm lại, từ lưu trữ dữ liệu phân tán đến sử dụng tài nguyên lưu trữ đám mây, công nghệ dữ liệu lớn cung cấp cho các tổ chức tài chính khả năng lưu trữ, quản lý và xử lý dữ liệu mạnh mẽ, trong khi công nghệ trí tuệ nhân tạo có thể đạt được khả năng phân tích chuyên sâu và khai thác dữ liệu. Thông qua công nghệ lưu trữ phân tán, các tổ chức tài chính có thể xử lý hiệu quả lượng dữ liệu khổng lồ, đạt được độ tin cậy và khả năng mở rộng cao của dữ liệu, và gọi các tài nguyên lưu trữ đám mây để nâng cao hơn nữa hiệu quả và tính linh hoạt trong lưu trữ dữ liệu của các tổ chức tài chính. Đồng thời, nó trang bị cho họ khả năng tính toán linh hoạt và phân tán, từ đó tăng cường khả năng giám sát và phòng chống hành vi gian lận. Dữ liệu được gọi được xác định và phân tích dựa trên các mô hình trí tuệ nhân tạo, xác định chính xác và hiệu quả hành vi gian lận tài chính, cung cấp các giải pháp thông minh dựa trên dữ liệu cho các tổ chức tài chính.

## Tài liệu tham khảo

1. Li, J., Ye, Z., & Zhang, C. (2022). Study on the interaction between big data and artificial intelligence. *Systems Research and Behavioral Science*, 39(3), 641-648.
2. Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. *Innovative Technology at the Interface of Finance and Operations: Volume I*, 223-247.
3. Ghimire, A., Thapa, S., Jha, A. K., Adhikari, S., & Kumar, A. (2020, October). Accelerating business growth with big data and artificial intelligence. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 441-448). IEEE.
4. Choi, D., & Lee, K. (2018). An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation. *Security and Communication Networks*, 2018.
5. Zhou, H., Sun, G., Fu, S., Fan, X., Jiang, W., Hu, S., & Li, L. (2020). A distributed approach of big data mining for financial fraud detection in a supply chain. *Comput Mater Continua*, 64(2), 1091-1105.
6. Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J., & Gao, Y. (2021). Internet financial fraud detection based on a distributed big data approach with node2vec. *IEEE Access*, 9, 43378-43386.
7. Yu, T. R., & Song, X. (2021). Big data and artificial intelligence in the banking industry. In *Handbook of financial econometrics, mathematics, statistics, and machine learning* (pp. 4025-4041).