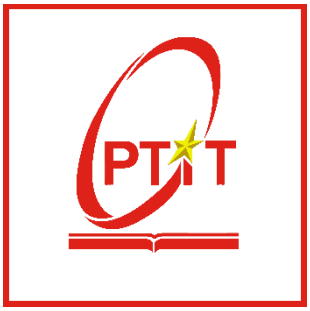
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Trương Gia Hiển

PHÁT HIỆN SỰ BẤT THƯỜNG TRONG HOẠT ĐỘNG CỦA MẠNG BLOCKCHAIN DỰA TRÊN HỌC MÁY

CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN

MÃ SỐ: 8.48.01.04

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

HÀ NỘI - 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Nguyễn Trung Kiên

(Ghi rõ học hàm, học vị)

Phản biện 1: ……

Phản biện 2: ……..

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng .... năm .....

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

Bối cảnh và đặt vấn đề

- Bối cảnh: Sự phát triển nhanh chóng của Tài chính Phi tập trung (DeFi) trên blockchain (Ethereum) đã hình thành một hạ tầng tài chính mới, không cần cấp phép và có khả năng lập trình cao.
- Vấn đề: Tính mở và phi tập trung khiến DeFi trở thành mục tiêu hấp dẫn của các hoạt động gian lận và tấn công có chủ đích (vay chớp nhoáng, rug pull, khai thác lỗ hổng...), gây thiệt hại lớn.
- Thách thức dữ liệu: Dấu vết giao dịch gian lận trên chuỗi khối có đặc điểm lệch, nhiễu cao, phụ thuộc thời gian và thiếu nhãn đầy đủ, làm các phương pháp giám sát truyền thống trở nên kém hiệu quả.
- Tính cấp thiết: Việc nghiên cứu và ứng dụng các phương pháp Học máy và Học sâu để phát hiện sớm giao dịch bất thường là quan trọng để hỗ trợ các giao thức DeFi và đơn vị quản lý rủi ro giảm thiểu thiệt hại, nâng cao an toàn hệ sinh thái.

Mục tiêu

- Đề tài: "Phát hiện sự bất thường trong hoạt động của mạng blockchain dựa trên học máy", tập trung vào giao dịch DeFi trên Ethereum với dữ liệu BCCC-DeFiFraudTrans-2025.
- Mục tiêu tổng quát: Khảo sát và đánh giá hiệu quả của các mô hình tiêu biểu trong phát hiện bất thường/gian lận giao dịch DeFi, từ đó đề xuất cách tiếp cận phù hợp cho hệ thống giám sát.

Phương pháp và phạm vi

- Phương pháp: Kết hợp nghiên cứu lý thuyết, phân tích dữ liệu thực nghiệm, thiết kế hai tuyến tiền xử lý, xây dựng pipeline huấn luyện thống nhất và phân tích, so sánh kết quả đa chiều.
- Phạm vi: Bài toán phát hiện bất thường/gian lận trong giao dịch DeFi trên Ethereum, tập trung vào bộ dữ liệu BCCC-DeFiFraudTrans-2025 với hai dạng biểu diễn. Luận văn chủ yếu xem xét các mô hình không giám sát hoặc bán giám sát.

Đóng góp chính

- Đề xuất quy trình tiền xử lý và huấn luyện có thể tái lập, với hai tuyến dữ liệu riêng biệt, tuân thủ nguyên tắc tách tập theo trật tự khối để hạn chế rò rỉ thông tin.
- Thực hiện so sánh thực nghiệm có hệ thống giữa các mô hình học máy và học sâu trên cùng một bộ dữ liệu và hệ thước đo.
- Chỉ ra sự phân hóa vai trò giữa mô hình trên dữ liệu bảng (phù hợp làm lớp lọc ưu tiên) và mô hình trên dữ liệu chuỗi (phù hợp tăng độ bao phủ), từ đó gợi ý hướng tiếp cận phân tầng cho hệ thống phát hiện gian lận thực tế.

Bố cục luận văn

- Chương 1: Cơ sở lý thuyết.
- Chương 2: Phương pháp nghiên cứu và thiết kế thực nghiệm.
- Chương 3: Triển khai thực nghiệm và đánh giá kết quả.

CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

1.1. Giới thiệu chung

- Thiết lập cơ sở lý thuyết, chuẩn hóa thuật ngữ và xác định phạm vi nghiên cứu cho bài toán phát hiện bất thường trên blockchain.
- Điểm qua các công trình liên quan để định vị đóng góp mới, tạo tiền đề cho thiết kế thực nghiệm ở các chương sau.

1.2. Tổng quan về Blockchain

1.2.1. Khái niệm của Blockchain

- Sổ cái phân tán dạng chỉ ghi, gồm các block liên kết bằng hash pointer.
- Hash pointer giúp phát hiện mọi chỉnh sửa quá khứ; được bảo vệ thêm bởi đồng thuận P2P.
- Mức bất biến tùy cơ chế.
- Blockchain công khai minh bạch toàn bộ lịch sử; blockchain cấp phép có kiểm soát truy cập.

1.2.2. Đặc điểm của Blockchain

1.2.2.1. Phi tập trung

- Không có máy chủ trung tâm; nhiều nút duy trì sổ cái, tăng khả năng chống kiểm duyệt.
- Giao dịch lan truyền đa tuyến qua mempool, tạo dòng sự kiện phong phú cho phát hiện bất thường.

1.2.2.2. Bất biến

- Block tham chiếu băm block trước, ngăn sửa đổi quá khứ.
- Bất biến giúp định danh ổn định cho giao dịch/khối; hữu ích cho gán nhãn và trích đặc trưng.

1.2.2.3. Minh bạch

- Lịch sử giao dịch hoàn toàn quan sát được, dù địa chỉ chỉ là bút danh.
- Cho phép trích đặc trưng hành vi: tần suất, kích thước giao dịch, gas, mức kết nối, mẫu tương tác hợp đồng, rất hữu ích cho bài toán anomalous detection.

1.2.2.4. Đồng thuận

- Xác định thứ tự block và trạng thái sổ cái.
- Động lực PoW/PoS ảnh hưởng mempool, giá gas, độ tắc nghẽn, nguồn tín hiệu mạnh cho mô hình xếp hạng bất thường, nhất là trong bối cảnh MEV.

1.2.3. Cấu trúc dữ liệu của Blockchain

1.2.3.1. Tiêu đề khối

- Băm khối trước để liên kết chuỗi.
- Ba gốc băm Merkle-Patricia trie: Transactions root, Receipts root, State root.
- Các tham số như số khối, thời điểm, gas limit, gas used, base fee.

1.2.3.2. Thân khối

- Chứa danh sách giao dịch theo thứ tự do block producer quyết định.
- Mỗi giao dịch sinh một biên nhận (receipt) ghi lại kết quả và sự kiện.
- Là nguồn dữ liệu trực tiếp cho đặc trưng và nhãn trong bài toán phát hiện bất thường.

1.2.4. Giao dịch và vòng đời của giao dịch

1.2.4.1. Khái niệm và cấu trúc giao dịch

- Đơn vị thay đổi trạng thái sổ cái, được ký bằng khóa riêng và phát tán P2P.
- Cấu trúc giao dịch Ethereum.

1.2.4.2. Vòng đời giao dịch

- Tạo và ký giao dịch, phát tán vào mempool, được chọn vào block theo phí và MEV.
- Thực thi trong EVM, cập nhật trạng thái hoặc thất bại khi lỗi/hết gas.
- Đạt độ tin cậy sau nhiều xác nhận hoặc theo cơ chế finality.

1.2.4.3. Kết quả thực thi, biên nhận và sự kiện

- Biên nhận ghi trạng thái thành công/thất bại, gas used, logs sự kiện, địa chỉ hợp đồng mới.
- Logs/sự kiện (transfer, swap, mint/burn, borrow/repay…) là nguồn tín hiệu quan trọng trong DeFi.
- Giao dịch nội bộ (internal tx) xuất hiện trong traces, hỗ trợ phân tích hành vi sâu hơn.

1.3. Tổng quan về DeFi

1.3.1. Khái niệm của DeFi

- DeFi (Decentralized Finance) là lớp dịch vụ tài chính vận hành trực tiếp trên blockchain thông qua smart contract.
- Đặc trưng cốt lõi:
    - Không cần cấp phép: bất kỳ ai cũng có thể truy cập và sử dụng giao thức.
    - Không lưu ký: người dùng tự giữ tài sản, không phụ thuộc bên thứ ba.
    - Khả năng lắp ghép: các giao thức có thể kết hợp tự do để tạo thành chuỗi hoạt động phức tạp.

1.3.2. Mục tiêu của DeFi

- Mở rộng tiếp cận dịch vụ tài chính với rào cản gia nhập thấp trên hạ tầng công khai.
- Tối ưu hiệu quả vốn thông qua tự động hóa khớp lệnh, cho vay có thế chấp và luân chuyển thanh khoản giữa giao thức.
- Giảm nhu cầu tin cậy tập trung bằng mã nguồn công khai và cơ chế thực thi minh bạch trên chuỗi.
- Lập trình được sản phẩm tài chính mới thông qua khả năng kết hợp nhiều khối xây dựng.

1.3.3. Ứng dụng phổ biến của DeFi

- Hoán đổi tài sản (AMM DEX).
- Vay, cho vay thế chấp.
- Cung cấp thanh khoản, khai thác lợi suất.
- Phái sinh phi tập trung.
- Cầu nối và tài sản tổng hợp.

1.4. Hoạt động bất thường trong Blockchain

1.4.1. Khái niệm về hoạt động bất thường

- Hoạt động bất thường là hành vi lệch khỏi mẫu bình thường, xuất hiện ở cấp giao dịch, địa chỉ, hợp đồng hoặc cơ chế DeFi.
- Bất thường rộng hơn gian lận (không phải mọi bất thường đều ác ý).
- Gồm 4 dạng chính: điểm, ngữ cảnh, tập thể, thay đổi chế độ.

- Bài toán khó do: ít nhãn, dữ liệu nhiễu, trôi khái niệm, tác nhân đối kháng, dữ liệu đa nguồn và gần thời gian thực.
- Nhận diện hiệu quả dựa trên: đặc trưng thống kê, đồ thị, thực thi, đặc trưng kinh tế, kết hợp mô hình không giám sát/bán giám sát theo hướng xếp hạng nguy cơ.

1.4.2. Các hình thức gian lận giao dịch trong DeFi

- Vay chớp nhoáng thao túng giá.
- Khai thác lỗi hợp đồng thông minh.
- MEV, đón đầu và kẹp giá.
- Rút thanh khoản và thoát dự án.
- Tấn công cầu nối xuyên chuỗi.
- Tấn công quản trị dùng vay chớp nhoáng.
- Lừa đảo cấp quyền chi tiêu, chiếm đoạt ví.
- Thao túng thị trường và tài sản thế chấp.

1.5. Phương pháp phát hiện bất thường

1.5.1. Bài toán phát hiện bất thường trong dữ liệu giao dịch

- Mục tiêu: gán điểm rủi ro cho mỗi thực thể để xếp hạng hoặc phân loại theo ngưỡng.
- Dữ liệu gồm bảng đặc trưng giao dịch và chuỗi thời gian theo địa chỉ/phiên.
- Nhãn gian lận rất ít, nên mô hình thường học hành vi bình thường rồi phát hiện sai lệch.
- Thách thức: mất cân bằng lớp, trôi khái niệm, phụ thuộc chuỗi, phân phối nặng đuôi, và tín hiệu gian lận dễ bị che giấu.
- Cần kết hợp nhiều loại đặc trưng (thống kê, ngữ cảnh, hành vi).
- Ba hướng chính: không giám sát, bán giám sát, và có giám sát qua nhiệm vụ phụ.

1.5.2. Mô hình học máy

1.5.2.1. K-Means

- Cơ chế: Phân cụm dữ liệu thành K nhóm; điểm nằm xa các tâm cụm được xác định là bất thường.
- Ưu điểm: Phù hợp cho dữ liệu bảng (tabular) có cấu trúc rõ ràng, không yêu cầu nhãn.

- Hạn chế: Giả định các cụm có dạng cầu và nhạy cảm với thang đo (cần chuẩn hóa dữ liệu kỹ).

1.5.2.2. Isolation Forest

- Cơ chế: Dựa trên nguyên lý cô lập; các điểm bất thường (hiếm) thường có đường đi ngắn hơn trên cây nhị phân ngẫu nhiên.
- Ưu điểm: Tốc độ xử lý nhanh, mở rộng tốt với dữ liệu nhiều chiều.
- Hạn chế: Hiệu quả giảm khi phân phối dữ liệu giữa lớp bình thường và bất thường bị chồng lấn mạnh.

1.5.2.3. Histogram-Based Outlier Score

- Cơ chế: Tính toán điểm bất thường dựa trên mật độ tần suất (histogram) của từng đặc trưng riêng biệt.
- Ưu điểm: Tốc độ rất nhanh, chi phí tính toán thấp.
- Hạn chế: Giả định các đặc trưng độc lập mạnh (bỏ qua mối tương quan giữa các biến), kết quả phụ thuộc vào số lượng bin.

1.5.3. Mô hình học sâu

1.5.3.1. Dense Autoencoder

- Cơ chế: Sử dụng mạng nơ-ron để nén và tái tạo dữ liệu bình thường; mẫu có lỗi tái tạo lớn được xem là bất thường.
- Ưu điểm: Linh hoạt, xử lý tốt dữ liệu bảng nhiều chiều.
- Hạn chế: Dễ bị quá khớp (overfitting), cần áp dụng kỹ thuật điều chuẩn (regularization).

1.5.3.2. Long Short-term Memory Autoencoder

- Cơ chế: Kết hợp kiến trúc Encoder-Decoder với các lớp LSTM để học sự phụ thuộc và động học theo thời gian.
- Ưu điểm: Chuyên biệt cho dữ liệu chuỗi; phát hiện hiệu quả các hành vi lạ hoặc trình tự bất thường.
- Đặc điểm: Phù hợp khi yếu tố thời gian/ngữ cảnh đóng vai trò trọng yếu.

1.5.3.3. Transformer Variational Autoencoder

- Cơ chế: Kết hợp Self-Attention (Transformer) và mô hình xác suất (VAE) để học phụ thuộc dài hạn và mô hình hóa không gian ẩn.
- Tiêu chí: Xác định bất thường dựa trên lỗi tái tạo cao hoặc xác suất xuất hiện thấp trong không gian ẩn.
- Đánh giá: Mạnh mẽ với chuỗi đa biến phức tạp nhưng yêu cầu dữ liệu lớn và chi phí tính toán cao.

1.6. Các công trình liên quan

1.6.1. Tổng quan các công trình

- Nghiên cứu bắt đầu từ Bitcoin, sau mở rộng sang Ethereum và đặc biệt là DeFi với hành vi hợp đồng phức tạp.
- Các phương pháp không giám sát phổ biến: phân cụm, mô hình cây, mật độ, one-class.
- Nhiều nghiên cứu xác nhận hiệu quả của K-means, Isolation Forest, HBOS, Autoencoder trên dữ liệu thực.
- Một số hướng tập trung vào siêu dữ liệu và kiến trúc blockchain để phát hiện gần thời gian thực.
- Các mô hình sâu gần đây (đặc biệt Transformer + Autoencoder) cho thấy tiềm năng mô hình hóa chuỗi sự kiện DeFi.

1.6.2. Khoảng trống nghiên cứu

- Đa số nghiên cứu vẫn nghiêng về Bitcoin, trong khi rủi ro DeFi đa dạng nhưng ít được khảo sát bằng mô hình không giám sát.
- Các nghiên cứu học sâu cho DeFi còn rời rạc, thường dùng một mô hình duy nhất, thiếu đối sánh hệ thống.
- Chưa có chuẩn đánh giá thống nhất đặt mô hình cổ điển và mô hình sâu lên cùng dữ liệu, tách thời gian, với thước đo phù hợp dữ liệu lệch nhãn.
- Do đó khó đánh giá đầy đủ ưu–nhược điểm từng nhóm mô hình.

1.6.3. Đóng góp của luận văn

- Xây dựng chuẩn so sánh thống nhất cho sáu mô hình từ cổ điển đến học sâu trên cùng pipeline.
- Áp dụng tách dữ liệu theo thời gian để phản ánh trôi khái niệm và động học giao dịch DeFi.

- Thực hiện so sánh có hệ thống giữa hai họ mô hình: học máy và học sâu.
- Làm rõ vai trò của các mô hình chuỗi (LSTM-AE, Transformer-VAE) trong việc nắm bắt phụ thuộc thời gian và tương tác hợp đồng.

1.7. Kết luận chương

- Trình bày nền tảng lý thuyết: bản chất blockchain, cấu trúc giao dịch, cơ chế vận hành của blockchain công khai và khái niệm DeFi.
- Làm rõ khái niệm hoạt động bất thường, các dạng gian lận phổ biến và các tín hiệu quan sát được trong dữ liệu.
- Giới thiệu các hướng tiếp cận phát hiện bất thường bằng học máy và học sâu, kèm nguyên lý của từng mô hình liên quan.
- Xác định hướng thực nghiệm dựa trên hai dạng dữ liệu: dữ liệu bảng (tabular) và chuỗi giao dịch theo địa chỉ.
- Danh sách sáu mô hình sẽ triển khai ở các chương sau: K-Means, Isolation Forest, HBOS, Dense Autoencoder, LSTM-AE, Transformer-VAE.

CHƯƠNG 2. PHƯƠNG PHÁP NGHIÊN CỨU VÀ THIẾT KẾ THỰC NGHIỆM

2.1. Giới thiệu chung

- Mục tiêu: Chuyển hóa cơ sở lý thuyết thành một quy trình thực nghiệm thống nhất, có tính tái lập cao.
- Xử lý dữ liệu: Xác định nguồn, cách biểu diễn dữ liệu và chuẩn hóa pipeline tiền xử lý nhằm bảo toàn tín hiệu hành vi, hạn chế nhiễu.
- Thiết kế thực nghiệm: Xây dựng kiến trúc mô hình và chiến lược tối ưu đảm bảo việc đánh giá là khách quan và không thiên lệch.
- Nguyên tắc kỹ thuật: Bám sát các đặc thù dữ liệu (mất cân bằng, phụ thuộc thời gian, trôi khái niệm) và ưu tiên sự so sánh công bằng giữa hai nhóm phương pháp: Học máy và Học sâu.

2.2. Dữ liệu

2.2.1. Bộ dữ liệu

- Sử dụng bộ dữ liệu BCCC-DeFiFraudTrans-2025 gồm hơn một triệu giao dịch Ethereum.
- Dữ liệu kết hợp đặc trưng ở mức ví và mức giao dịch, đã được thẩm định nhãn và kiểm tra trùng lặp.

2.2.2. Khai phá dữ liệu

- Dữ liệu có 94 cột với tỷ lệ giao dịch gian lận là 8,65%.
- Phân bố theo block_number hỗ trợ việc tách tập theo trật tự thời gian.

2.2.3. Tiền xử lý dữ liệu

2.2.3.1. Tuyến bảng đặc trưng cho học máy

Giữ 30 đặc trưng quan trọng nhất, chuẩn hóa chỉ theo tập huấn luyện và xuất CSV và Parquet phục vụ K-Means, Isolation Forest, HBOS và Dense AE.

2.2.3.2. Tuyến chuỗi thời gian cho học sâu theo trình tự

Nhóm theo address, sắp xếp theo block_number, tạo chuỗi độ dài 20 với nhãn có fraud. Xuất NPZ gồm X, y, features, addresses và end_block.

2.2.4. Kiểm soát chất lượng

Kiểm tra kích thước, tính hữu hạn, tính hợp lệ của nhãn và sự ổn định tỷ lệ dương tính. Mọi biến đổi đều cố định seed và chỉ khớp trên tập huấn luyện để tránh rò rỉ tương lai.

2.3. Thiết kế mô hình huấn luyện

2.3.1. Pipeline huấn luyện chung

- Mục tiêu: Thiết lập quy trình nhất quán, áp dụng đồng bộ cho cả sáu mô hình (từ nạp dữ liệu đến báo cáo kết quả).
- Nguyên tắc: Tất cả mô hình đều trả về điểm bất thường cùng chiều, tối ưu trên tập kiểm định và đánh giá mù trên tập kiểm tra.
- Các bước chính:
  - Nạp dữ liệu & Kiểm tra chất lượng.
  - Tách tập dữ liệu theo thời gian.
  - Làm sạch & Chuẩn hóa thang đo đặc trưng.
  - Sinh điểm bất thường và quy ước cùng chiều.
  - Tối ưu siêu tham số (dùng Optuna).
  - Huấn luyện toàn cục, Chọn ngưỡng và Đánh giá.

2.3.2. Thiết kế huấn luyện mô hình học máy

2.3.2.1. K-Means

- Dữ liệu CSV đã chuẩn hóa RobustScaler, tách theo thời gian 60-20-20.
- Tối ưu K ∈ [1, 500], n_init ∈ [8, 40], init {k-means++, random}, và PCA 8–32 chiều.
- Optuna (TPE) tối ưu hàm J; có bước tinh chỉnh cục bộ quanh cấu hình tốt nhất.
- Huấn luyện lại toàn bộ với PCA nếu được chọn; ngưỡng lấy theo F1 Validation.

2.3.2.2. Isolation Forest

- Dữ liệu CSV theo pipeline chung, chuẩn hóa RobustScaler.
- Tối ưu n_estimators 200–800, max_samples 0.3–1.0, max_features 0.3–1.0, PCA 8–32.
- Contamination dùng để nội chuẩn hóa, không đặt ngưỡng.
- Huấn luyện full-train, chọn ngưỡng theo F1 Validation.

2.3.2.3. Histogram-Based Outlier Score

- Dữ liệu CSV, chuẩn hóa RobustScaler, tách theo thời gian.
- Tối ưu n_bins 20–200, alpha 0.05–0.999, tol 0.001–0.5, kèm PCA 8–32 để khử tương quan.
- Optuna + refinement tinh chỉnh cục bộ.
- Huấn luyện lại full-train, ngưỡng theo F1 Validation.

2.3.3. Thiết kế huấn luyện mô hình học sâu

2.3.3.1. Dense Autoencoder

- Huấn luyện chỉ trên Train-Legit; dùng Parquet và RobustScaler (fit trên legit).
- Kiến trúc đối xứng, có GaussianNoise + BatchNorm + Dropout.
- Optuna tìm kiến trúc (w1, w2, bottleneck), dropout, L2, LR, activation.
- Huấn luyện lại full Train-Legit với EarlyStopping; tính median/MAD trên Val-Legit.
- Ngưỡng lấy theo F1 Validation và áp dụng cho Test.

2.3.3.2. Long Short-term Memory Autoencoder

- Dữ liệu NPZ chuỗi; tách theo end_block; impute-median → clip → RobustScaler.
- Encoder–Decoder dùng LSTM (có thể Bidirectional) + RepeatVector + TimeDistributed.
- Optuna tìm units, dropout, LR, batch size, Bidirectional.
- Huấn luyện lại full Train-Legit; chọn ngưỡng F1 Validation.
- Đánh giá Test bằng các chỉ số chuẩn.

2.3.3.3. Transformer Variational Autoencoder

- Chuẩn hóa như LSTM-AE; có tùy chọn bỏ scaler ngoài và dùng LayerNorm nội bộ.
- Time2Vec + PositionalEmbedding; Transformer Encoder–Decoder + lớp VAE (μ, σ).
- Optuna tối ưu embed_dim, num_heads, số khối, FF-expansion, latent-size, dropout, LR.
- Huấn luyện full; điểm Test lấy trung bình nhiều lần Monte Carlo.
- Ngưỡng cố định theo F1 Validation.

2.4. Các thước đo đánh giá hiệu năng

2.4.1. Ma trận nhầm lẫn

Mọi chỉ số quyết định đều được dẫn xuất từ ma trận nhầm lẫn, bao gồm 4 thành phần cơ bản khi so sánh nhãn dự đoán với nhãn thực tế:

- True Positive (TP - Dương tính thật): Phát hiện chính xác gian lận.
- False Positive (FP - Dương tính giả): Báo động giả.
- True Negative (TN - Âm tính thật): Phân loại đúng hợp pháp.
- False Negative (FN - Âm tính giả): Bỏ lọt gian lận.

2.4.2. Nhóm chỉ số đánh giá xếp hạng

2.4.2.1. Precision-Recall Curve (PR-AUC / AP)

- Quan trọng nhất cho dữ liệu lệch.
- Tóm tắt trade-off giữa Precision – Recall ở mọi ngưỡng.
- AP càng cao thì mô hình càng đưa giao dịch gian lận lên đầu danh sách.

2.4.2.2. ROC Curve (ROC-AUC)

- Đo mức phân tách giữa hai lớp qua TPR–FPR.
- Phản ánh độ ổn định tổng quát.

2.4.2.3. Precision@1%

- Định hướng vận hành: ưu tiên top giao dịch nghi nhất.
- Tỷ lệ gian lận thực sự nằm trong top 1% điểm cao nhất.
- Precision@1% cao thì giảm tải điều tra thủ công.

2.4.3. Nhóm chỉ số đánh giá quyết định

2.4.3.1. Precision (Độ chính xác dương tính)

Cho biết độ tin cậy của cảnh báo; Precision thấp thì nhiều báo động giả.

2.4.3.2. Recall (Độ nhạy / Độ phủ)

Cho biết mức độ bao phủ rủi ro; Recall thấp thì nhiều gian lận bị lọt..

2.4.3.3. F1-Score

- Trung bình điều hòa của Precision & Recall.
- Dùng làm tiêu chí chọn ngưỡng tối ưu trên Validation.

2.4.3.4. Balanced Accuracy (Độ chính xác cân bằng)

Giúp đánh giá công bằng khi lớp thiểu số rất nhỏ.

2.4.4. Chiến lược tối ưu hóa tổng hợp

Tinh chỉnh siêu tham số bằng hàm mục tiêu lai:

$$J = 0.5 \times AP + 0.3 \times ROC - AUC + 0.2 \times Precision@1\%$$

2.5. Kết luận chương

Chương 2 đã thiết lập hoàn chỉnh khung phương pháp luận và quy trình thực nghiệm cho bài toán phát hiện gian lận DeFi. Các đóng góp chính bao gồm:

- Quy trình xử lý dữ liệu chặt chẽ.
- Thiết kế mô hình đa dạng.
- Chiến lược tối ưu và đánh giá định hướng vận hành.

CHƯƠNG 3. TRIỂN KHAI THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. Giới thiệu chung

- Chuyển hóa thiết kế phương pháp luận (Chương 2) thành quy trình thực nghiệm cụ thể.
- Thực hiện huấn luyện, tối ưu hóa và đánh giá sáu mô hình đề xuất trên bộ dữ liệu BCCC-DeFiFraudTrans-2025.

3.2. Môi trường và công cụ

3.2.1. Cấu hình phần cứng

- CPU: Dual Intel Xeon Gold 6148 (Đảm bảo khả năng xử lý song song).
- RAM: 256 GB DDR4 (Hỗ trợ xử lý bộ dữ liệu lớn).
- GPU: NVIDIA GeForce RTX 4090 (Tối ưu huấn luyện mô hình học sâu).
- Lưu trữ: SSD NVMe 1TB (Tăng tốc độ đọc/ghi dữ liệu).

3.2.2. Công cụ và thư viện phần mềm

Nền tảng: Python 3.10.

Quản lý/Tiền xử lý dữ liệu:

- NumPy, Pandas, PyArrow (hỗ trợ định dạng Parquet).
- Scikit-learn (chuẩn hóa, giảm chiều PCA, chia tập).

Mô hình học máy:

- Scikit-learn (K-Means).
- PyOD (Isolation Forest, HBOS - chuyên biệt cho phát hiện bất thường).

Mô hình học sâu:

- TensorFlow 2.10+ / Keras.
- Tối ưu hóa GPU: Kích hoạt Mixed Precision (float16) và trình biên dịch XLA để tăng tốc độ huấn luyện.

Tối ưu và Đánh giá:

- Optuna: Tự động hóa quá trình tìm kiếm siêu tham số (sử dụng chiến lược TPE).

- Matplotlib, Seaborn: Trực quan hóa kết quả đánh giá (ROC/PR Curve).

3.3. Kết quả huấn luyện mô hình

3.3.1. Mô hình học máy

3.3.1.1. K-Means

- PR-AUC ≈ 0.7801, ROC-AUC ≈ 0.9146, Precision@1% = 1.0000.
- F1-score ≈ 0.6657, Precision ≈ 0.5707, Recall ≈ 0.7987.

3.3.1.2. Isolation Forest

- PR-AUC ≈ 0.1902, ROC-AUC ≈ 0.7168, Precision@1% ≈ 0.2065.
- F1-score ≈ 0.3215, Precision ≈ 0.2464, Recall ≈ 0.4623.

3.3.1.3. Histogram-Based Outlier Score

- PR-AUC = 0.2213, ROC-AUC = 0.7642, Precision@1% ≈ 0.3877.
- F1-score ≈ 0.3366, Precision ≈ 0.2347, Recall ≈ 0.5944.

3.3.2. Mô hình học sâu

3.3.2.1. Dense Autoencoder

- PR-AUC = 0.6574, ROC-AUC = 0.8395, Precision@1% = 1.0000.
- F1-score ≈ 0.5509, Precision ≈ 0.4479, Recall ≈ 0.7152.

3.3.2.2. Long Short-term Memory Autoencoder

- PR-AUC = 0.5137, ROC-AUC = 0.7514, Precision@1% ≈ 0.6765.
- F1-score ≈ 0.6419, Precision ≈ 0.4944, Recall ≈ 0.9149.

3.3.2.3. Transformer Variational Autoencoder

- PR-AUC = 0.4258, ROC-AUC = 0.6768, Precision@1% ≈ 0.7332.
- F1-score ≈ 0.6632, Precision ≈ 0.5273, Recall ≈ 0.8936.

3.4. Đánh giá và so sánh kết quả mô hình

Bảng tổng hợp cho thấy sự phân hóa rõ rệt giữa hai nhóm mô hình.

Bảng 3. 1: Tổng hợp hiệu năng chi tiết của các mô hình

Mô hình	PR-AUC	ROC-AUC	P@1%	Precision	Recall	F1-Score	TP	FP	FN	TN
K-Means	**0.7801**	**0.9146**	**1.0000**	0.5707	0.7987	**0.6657**	35,399	26,627	8,923	134,424
DAE	0.6574	0.8395	**1.0000**	0.4479	0.7152	0.5509	31,699	39,068	12,623	121,983
LSTM-AE	0.5137	0.7514	0.6765	0.4944	**0.9149**	0.6419	**57,295**	58,593	**5,330**	74,877
Trans-VAE	0.4258	0.6768	0.7332	0.5273	0.8936	0.6632	55,962	50,177	6,663	83,293
HBOS	0.2213	0.7642	0.3877	0.2347	0.5944	0.3366	10,556	34,415	7,203	153,200
Isolation Forest	0.1902	0.7168	0.2065	0.2464	0.4623	0.3215	8,210	25,106	9,549	**162,509**

3.4.1. Chất lượng xếp hạng

- K-Means và DAE có PR-AUC cao nhất (≈ 0.78 và 0.66).
- LSTM-AE và Transformer-VAE đạt mức trung bình (≈ 0.51 và 0.43).
- HBOS và Isolation Forest thấp nhất (≈ 0.22 và 0.19).

3.4.2. Khả năng phân tách tổng quát

- K-Means dẫn đầu (ROC-AUC ≈ 0.91); DAE xếp thứ hai (≈ 0.84).
- LSTM-AE và Transformer-VAE đạt 0.75 và 0.68.
- HBOS (≈ 0.76) khá ổn; Isolation Forest thấp nhất (≈ 0.72).

3.4.3. Hiệu quả tại vùng cảnh báo trọng điểm

- K-Means và DAE đạt tuyệt đối 1.0.
- Transformer-VAE và LSTM-AE đạt khá (≈ 0.73 và ≈ 0.68).
- HBOS và Isolation Forest thấp (≈ 0.39 và 0.21).

3.4.4. Năng lực bao phủ và cân bằng phát hiện

3.4.4.1. So sánh Recall

- LSTM-AE (≈ 0.915): Recall cao nhất, bao phủ gian lận tốt, phù hợp hệ thống ưu tiên hạn chế bỏ sót.
- Transformer-VAE (≈ 0.894): Gần tương đương LSTM-AE, điểm số ổn định hơn nhờ attention.
- Dense Autoencoder (≈ 0.715): Không dùng chuỗi nhưng vẫn nhận diện tốt, chứng minh AE tabular đủ mạnh.

- K-Means (≈ 0.799): Hiệu quả đáng kể dù chỉ phân cụm và không dùng thông tin thời gian.
- HBOS (≈ 0.594): Chỉ bắt tốt các outlier đơn giản, bỏ sót trường hợp phức tạp.
- Isolation Forest (≈ 0.462): Thấp nhất, do phân phối điểm hai lớp chồng lấn mạnh.

3.4.4.2. So sánh F1-score

- K-Means (≈ 0.666) và Transformer-VAE (≈ 0.663): Hai mô hình cân bằng tốt nhất giữa Precision và Recall.
- LSTM-AE (≈ 0.642): F1 khá cao, ưu tiên bao phủ hơn độ chính xác.
- Dense Autoencoder (≈ 0.551): Recall tốt nhưng Precision thấp làm giảm F1.
- HBOS (≈ 0.33) và Isolation Forest (≈ 0.32): Hiệu năng thấp cả ở phát hiện lẫn độ chính xác.

3.4.5. Đánh giá hiệu năng tính toán và khả năng triển khai thực tế

Bảng 3. 2: Tổng hợp hiệu năng tính toán của các mô hình

Mô hình	**Độ trễ đơn lẻ (ms/mẫu)**	**Thông lượng (mẫu/giây)**	**Thời gian xử lý Batch (s)**	**Kích thước (MB)**
HBOS	**0,29**	**1.277.705**	0,039	9,42
K-Means	0,53	262.249	0,192	2,39
Isolation Forest	25,07	25.774	1,941	**1.448,07**
DAE	119,42	264.592	0,193	3,63
LSTM-AE	131,00	15.496	0,653	32,40
Trans-VAE	120,00	20.076	0,500	**0,59**

- Độ trễ thấp nhất: HBOS và K-Means (<1 ms).
- Thông lượng cao: HBOS (≈ 1.28M mẫu/s), DAE và K-Means ≈ 260k mẫu/s.
- Mô hình nhẹ: Transformer-VAE (0.59 MB), K-Means và DAE (<4 MB).
- Isolation Forest rất nặng (≈ 1.45 GB) và thông lượng thấp.
- Đề xuất triển khai:
    - Tầng 1 – Lọc nhanh: HBOS hoặc K-Means.
    - Tầng 2 – Phân tích sâu: LSTM-AE hoặc Transformer-VAE.
    - Hạn chế Isolation Forest trong thực tế.

3.5. Kết luận chương

- Tổng kết thực nghiệm: Đã hoàn thành quy trình huấn luyện và đánh giá toàn diện 6 mô hình trên bộ dữ liệu BCCC-DeFiFraudTrans-2025.
- Phân hóa hiệu năng:
    - Nhóm Tabular (K-Means, DAE): Vượt trội về khả năng xếp hạng rủi ro (PR-AUC, ROC-AUC, Precision@1% cao). Phù hợp làm lớp lọc sơ cấp (lớp 1).
    - Nhóm Chuỗi (LSTM-AE, Transformer-VAE): Đạt Recall (độ phủ) cao nhất. Phù hợp làm lớp phân tích chuyên sâu (lớp 2) để hạn chế bỏ sót gian lận.
    - Nhóm Baseline (HBOS, Isolation Forest): Bộc lộ nhiều hạn chế và kém hiệu quả trên dữ liệu DeFi phức tạp.
- Kiến nghị: Khẳng định không có mô hình đơn lẻ nào tối ưu toàn diện; đề xuất hướng tiếp cận hiệu quả nhất là kiến trúc phân tầng, kết hợp ưu điểm của cả mô hình tabular (tốc độ, xếp hạng) và mô hình chuỗi (độ phủ).

KẾT LUẬN VÀ KIẾN NGHỊ

Tổng kết kết quả nghiên cứu

- Thực nghiệm: Triển khai toàn diện 6 mô hình học máy/học sâu trên bộ dữ liệu BCCC-DeFiFraudTrans-2025 với hai pipeline xử lý riêng biệt (bảng và chuỗi).
- Phân hóa vai trò mô hình:
    - Nhóm Tabular (K-Means, DAE): Tối ưu cho nhiệm vụ xếp hạng (Precision@1% tuyệt đối), phù hợp lọc ưu tiên.
    - Nhóm Sequence (LSTM-AE, Transformer-VAE): Tối ưu cho độ phủ (Recall cao), giúp giảm thiểu bỏ sót gian lận.
    - Nhóm Baseline (HBOS, Isolation Forest): Hiệu quả thấp hơn, chủ yếu dùng để tham chiếu.
- Đề xuất giải pháp: Kiến nghị áp dụng kiến trúc phân tầng, kết hợp sức mạnh của cả mô hình tabular và mô hình chuỗi để đạt hiệu quả tối ưu.

Hướng phát triển và đóng góp

- Hướng mở rộng:
    - Tích hợp thêm nguồn dữ liệu (biến động giá, thanh khoản, cross-chain).
    - Ứng dụng kiến trúc tiên tiến hơn (GNN, Transformer cho chuỗi dài).
    - Tăng cường cơ chế giải thích mô hình (Explainability) và duy trì kiểm định định kỳ.
- Giá trị thực tiễn: Luận văn cung cấp một đánh giá thực nghiệm có hệ thống và đề xuất khung giải pháp khả thi cho hệ thống giám sát an ninh trong DeFi.