**ỨNG DỤNG HỌC MÁY VÀ HỌC SÂU TRONG PHÁT HIỆN GIAN LẬN TRÊN NỀN TẢNG FINTECH**

*Lê Nhật Tùng1, Lê Hoàng Gia Vĩ2, Hoàng Nguyên Vũ2, Nguyễn Ngọc Quỳnh Anh 2*

*1Trường Đại học Công nghệ Đồng Nai,*

*2Trường Đại học Công nghệ TP. HCM*

***Tóm tắt:*** *Nhận diện giao dịch gian lận trên các bộ dữ liệu mất cân bằng nghiêm trọng là một thách thức đối với cả mô hình học máy truyền thống và học sâu. Nghiên cứu này đánh giá hiệu quả của sáu mô hình Random Forest, XGBoost, K-Nearest Neighbors (KNN), AdaBoost, CNN và Bi-LSTM trên hai thiết lập: dữ liệu gốc chưa cân bằng và dữ liệu đã cân bằng bằng kỹ thuật SMOTE-ENN. Kết quả thực nghiệm cho thấy các mô hình học sâu vượt trội hơn nhóm học máy truyền thống về Recall, F1-score và ROC-AUC. Trên dữ liệu gốc, Bi-LSTM đạt F1-score = 0.795 và ROC-AUC = 0.958, trong khi CNN đạt F1-score = 0.777 và ROC-AUC = 0.943. Sau khi áp dụng SMOTE-ENN, tất cả các mô hình cải thiện hiệu năng đáng kể, với Bi-LSTM đạt F1-score = 0.937 và ROC-AUC = 0.990, CNN đạt F1-score = 0.931 và ROC-AUC = 0.989. Kết quả nhấn mạnh vai trò quan trọng của cân bằng dữ liệu trong việc nâng cao khả năng phát hiện các giao dịch gian lận thuộc lớp thiểu số. Nghiên cứu chỉ ra rằng kết hợp học sâu với kỹ thuật tái lấy mẫu là chiến lược hiệu quả để đạt độ nhạy cao và dự đoán ổn định trên các bộ dữ liệu mất cân bằng nghiêm trọng.*

***Từ khoá:*** *Phát hiện gian lận, dữ liệu mất cân bằng, học máy, học sâu, SMOTE-ENN*

**1. Tổng quan**

Sự phát triển mạnh mẽ của các nền tảng FinTech trong thập kỷ gần đây đã thúc đẩy sự phổ biến của các dịch vụ tài chính số như thanh toán trực tuyến, giao dịch điện tử và ví kỹ thuật số. Tuy nhiên, cùng với sự gia tăng nhanh chóng của các giao dịch, các hình thức gian lận tài chính cũng trở nên tinh vi hơn, gây thiệt hại lớn cho tổ chức tài chính và người dùng. Báo cáo của Cơ quan Ngân hàng Châu Âu cho thấy tỷ lệ gian lận trong thanh toán điện tử tăng trung bình hằng năm và có xu hướng phức tạp hơn trong môi trường giao dịch trực tuyến [1]. Điều này đặt ra yêu cầu cấp thiết đối với các hệ thống phát hiện gian lận thông minh, có khả năng hoạt động theo thời gian thực và thích ứng với dữ liệu thay đổi liên tục.

Các phương pháp phát hiện gian lận truyền thống dựa trên luật (rule-based) và thống kê mô tả tuy đơn giản và dễ triển khai nhưng không còn phù hợp trong bối cảnh dữ liệu lớn (big data) và phi tuyến tính của FinTech. Những mô hình này thường thiếu khả năng tổng quát hóa và khó phát hiện các hành vi gian lận mới xuất hiện [2]. Do đó, nhiều nghiên cứu đã hướng sang ứng dụng các thuật toán học máy (machine learning – ML), bao gồm Random Forest, Logistic Regression, Gradient Boosting và SVM, nhằm cải thiện khả năng phân loại giao dịch bất thường [3]. Mặc dù mang lại hiệu quả đáng kể, mô hình ML truyền thống vẫn gặp hạn chế trong việc trích xuất đặc trưng tự động khi dữ liệu có cấu trúc phức tạp hoặc thay đổi theo thời gian.

Trong những năm gần đây, học sâu (deep learning – DL) trở thành một phương pháp nổi bật nhờ khả năng mô hình hóa quan hệ phi tuyến và tự động học đặc trưng với độ chính xác cao. Các mô hình như LSTM và các kiến trúc mạng khác đã được chứng minh là có hiệu quả vượt trội trong việc phát hiện gian lận tài chính, đặc biệt trong bối cảnh dữ liệu mất cân bằng hoặc có tính chuỗi thời gian [4]. Song song đó, nhiều công trình gần đây cũng khảo sát và tổng hợp các phương pháp ML/DL dành cho phát hiện gian lận, chỉ ra cả tiềm năng và thách thức (ví dụ: dữ liệu mất cân bằng, yêu cầu thời gian thực, và tính giải thích) mà hệ thống thương mại và ngân hàng cần đối mặt [5], [6].

Xuất phát từ bối cảnh và thách thức trên, nghiên cứu *“Ứng dụng học máy và học sâu trong phát hiện gian lận trên nền tảng FinTech”* tập trung khảo sát, phân tích và đánh giá các mô hình ML và DL hiện đại, nhằm xây dựng cơ sở khoa học cho các phương pháp phát hiện gian lận hiệu quả, thích ứng và có khả năng triển khai trong môi trường FinTech quy mô lớn. Mục tiêu của nghiên cứu là đề xuất một khung tiếp cận vừa đảm bảo độ chính xác, vừa duy trì tính ổn định và khả năng ứng dụng thực tế trong các hệ thống tài chính số.

**2. Các nghiên cứu liên quan**

Phát hiện gian lận trong các giao dịch tài chính, đặc biệt trên nền tảng FinTech, là một chủ đề nghiên cứu sôi động. Nhiều nghiên cứu đã ứng dụng các phương pháp học máy (ML) và học sâu (DL) để cải thiện độ chính xác, khả năng phát hiện gian lận, đồng thời giải quyết các thách thức như dữ liệu mất cân bằng, yêu cầu xử lý thời gian thực và tính giải thích mô hình.

Trong bối cảnh dữ liệu mất cân bằng, Ndama et al. [7] đề xuất một mô hình mạng neural sâu (DNN) kết hợp kỹ thuật *resampling* như SMOTE và tối ưu trọng số để phát hiện gian lận thẻ tín dụng. Mô hình này đạt độ chính xác 99,83% và giảm tối đa số lượng false negative. Một nghiên cứu khác [8] xây dựng mô hình ensemble học sâu gồm LSTM, GRU làm base learners và MLP làm meta-learner, kết hợp SMOTE-ENN để cân bằng dữ liệu. Kết quả thực nghiệm cho thấy mô hình đạt sensitivity = 1.000 và specificity = 0.997, vượt trội so với các mô hình ML truyền thống. Ngoài ra, việc so sánh các thuật toán ML truyền thống và DL “state-of-the-art” trên dữ liệu thẻ tín dụng cho thấy các mô hình DL như CNN, DNN vượt trội về các chỉ số precision, F1 và AUC [9].

Một số nghiên cứu gần đây mở rộng phạm vi sang các ứng dụng FinTech tổng quát. Ví dụ, Stojanović et al. [10] khảo sát việc áp dụng các phương pháp anomaly detection trong ML cho nhiều loại gian lận tài chính, bao gồm giao dịch thẻ tín dụng, blockchain và thanh toán trực tuyến. Họ chỉ ra rằng false positives là một trong những thách thức chính trong việc triển khai hệ thống thực tế, ảnh hưởng đến hiệu quả hoạt động của các nền tảng tài chính. Ngoài ra, một số nghiên cứu sử dụng kỹ thuật *one-class classification* kết hợp *subspace learning* để xử lý dữ liệu mất cân bằng, giúp khắc phục vấn đề “curse of dimensionality” và phân lớp gian lận rất hiếm[11]. Song song đó, nhu cầu giải thích mô hình (Explainable AI – XAI) đang ngày càng được quan tâm trong phát hiện gian lận, nhằm đáp ứng yêu cầu minh bạch và tuân thủ pháp lý. Một số công trình đề xuất mô hình ML có khả năng giải thích cho các quyết định phát hiện gian lận, giúp các nhà phân tích tin tưởng hơn vào kết quả dự đoán [12], [13]. Chẳng hạn, mô hình xFraud kết hợp graph neural network và trình giải thích (explainer) cho phép phát hiện giao dịch gian lận với hiệu suất cao đồng thời tạo ra lời giải thích dễ hiểu cho người dùng [13].

Tổng hợp các kết quả trên, có thể nhận thấy rằng ML và DL cung cấp các hướng tiếp cận hiệu quả trong phát hiện gian lận, nhưng vẫn tồn tại ba thách thức cốt lõi: (i) dữ liệu mất cân bằng, khi tỷ lệ gian lận rất thấp so với tổng số giao dịch; (ii) xử lý thời gian thực, yêu cầu mô hình phản hồi nhanh trong các giao dịch FinTech; và (iii) tính giải thích, đặc biệt đối với các mô hình DL “black-box” trong môi trường có quy định nghiêm ngặt. Những thách thức này định hướng cho việc phát triển các mô hình ML/DL thích hợp, có khả năng ứng dụng thực tế trên nền tảng FinTech.

**3. Phương pháp nghiên cứu**

Trong nghiên cứu này, chúng tôi đề xuất một quy trình thực nghiệm nhằm đánh giá hiệu quả của các mô hình học máy (ML) và học sâu (DL) trong bài toán phát hiện gian lận giao dịch trên nền tảng FinTech. Toàn bộ quá trình được triển khai trên bộ dữ liệu công khai từ Kaggle (Credit card fraud detection) [14], vốn được sử dụng phổ biến trong các nghiên cứu về gian lận tài chính do đặc trưng mất cân bằng nghiêm trọng giữa số lượng giao dịch gian lận và giao dịch hợp lệ.

*3.1 Mô tả và tiền xử lý dữ liệu*

Bộ dữ liệu được sử dụng trong nghiên cứu bao gồm các giao dịch thẻ tín dụng của khách hàng châu Âu được ghi nhận trong tháng 9 năm 2013, với tổng cộng 284.807 giao dịch [14]. Đặc điểm nổi bật của bộ dữ liệu này là mức độ mất cân bằng lớp rất cao, khi số lượng giao dịch hợp lệ chiếm gần như toàn bộ, trong khi giao dịch gian lận chỉ xuất hiện với tần suất cực kỳ thấp. Cụ thể, chỉ có 492 giao dịch được xác định là gian lận, còn lại đều là giao dịch hợp pháp. Tập dữ liệu gồm 31 thuộc tính, bao gồm thời gian, giá trị giao dịch, nhãn phân loại và các đặc trưng V1 – V28 là các biến đã được ẩn danh và trích xuất thông qua kỹ thuật PCA. Sự chênh lệch lớn giữa hai nhóm lớp không chỉ làm tăng độ phức tạp của bài toán mà còn ảnh hưởng đáng kể đến hiệu suất của các mô hình học máy nếu không có chiến lược xử lý phù hợp. Bảng 1 mô tả chi tiết cấu trúc và đặc trưng của từng thuộc tính trong bộ dữ liệu. Do mức độ lệch phân bố quá lớn giữa hai lớp, việc áp dụng các kỹ thuật tiền xử lý và phương pháp đặc thù nhằm giảm thiểu tác động của sự mất cân bằng là điều cần thiết để đảm bảo mô hình đạt được độ chính xác và độ tin cậy cao.

*Bảng 1: Mô tả dữ liệu*

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| Time | Thể hiện số giây trôi qua giữa mỗi giao dịch và giao dịch đầu tiên trong bộ dữ liệu |
| V1, V2, …, V28 | Các đặc trưng dạng số được tạo ra từ quá trình biến đổi PCA, giúp đảm bảo tính ẩn danh và bảo mật thông tin người dùng. |
| Amount | Biểu thị giá trị của từng giao dịch và có thể được sử dụng trong các phương pháp học nhạy cảm theo chi phí phụ thuộc vào từng mẫu |
| Class | Là biến phản hồi, trong đó giá trị 1 biểu thị giao dịch gian lận và giá trị 0 biểu thị giao dịch hợp lệ. |

Sau khi hoàn thiện bảng mô tả các đặc trưng của bộ dữ liệu, nghiên cứu tiến hành một chuỗi bước tiền xử lý nhằm nâng cao chất lượng và mức độ phù hợp của dữ liệu cho quá trình phân tích. Trước tiên, giá trị của đặc trưng ‘Amount’ được chuẩn hóa về khoảng [−1,1] bằng hàm StandardScaler từ thư viện Scikit-Learn, nhằm đảm bảo sự đồng nhất về thang đo giữa các biến đầu vào. Đồng thời, để duy trì tính toàn vẹn của dữ liệu, toàn bộ các bản ghi trùng lặp và các bản ghi có giá trị thiếu đã được rà soát và loại bỏ theo quy trình hệ thống. Các bước tiền xử lý này đóng vai trò then chốt trong việc cải thiện chất lượng dữ liệu, tạo điều kiện thuận lợi cho những phân tích chuyên sâu và việc xây dựng mô hình phát hiện gian lận thẻ tín dụng.

*3.2 Thiết lập thử nghiệm và huấn luyện mô hình*

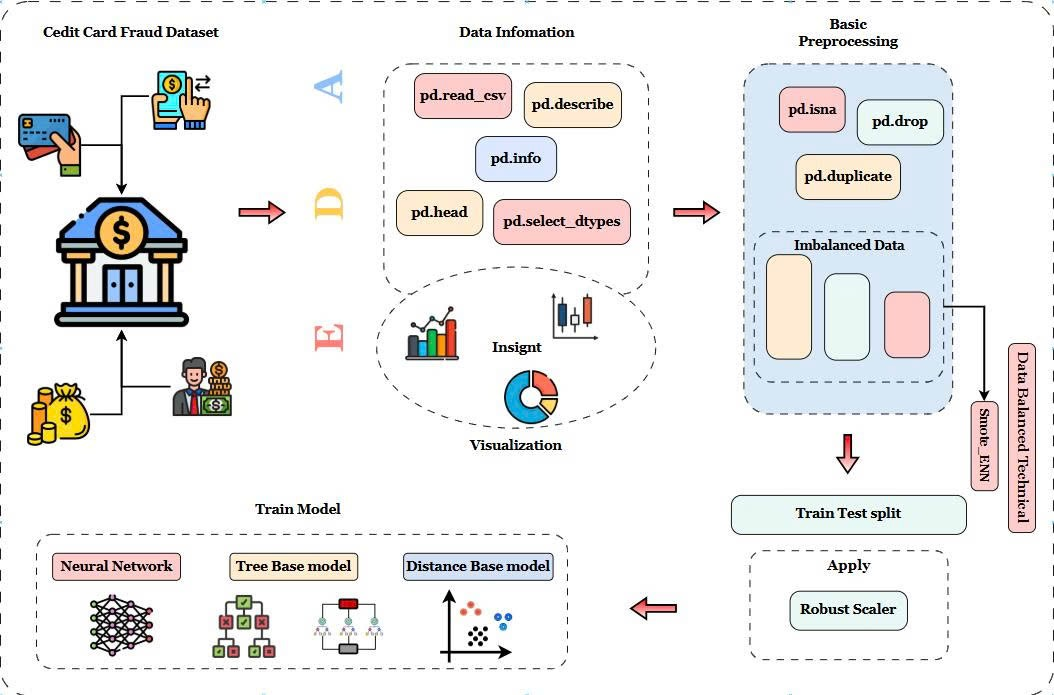
Trong nghiên cứu này, hai thiết lập thử nghiệm được triển khai nhằm đánh giá hiệu quả của các mô hình học máy (ML) và học sâu (DL) trong phát hiện gian lận giao dịch, đồng thời phân tích tác động của việc xử lý mất cân bằng dữ liệu.

* Thiết lập 1 – Huấn luyện trên dữ liệu gốc chưa cân bằng:

Dữ liệu thô từ bộ Credit Card Fraud Detection được sử dụng trực tiếp mà không áp dụng bất kỳ phương pháp xử lý mất cân bằng nào. Mục tiêu của thiết lập này là cung cấp cơ sở để so sánh, đồng thời đánh giá khả năng phát hiện gian lận trong điều kiện dữ liệu thực tế, vốn có sự chênh lệch nghiêm trọng giữa lớp giao dịch gian lận và giao dịch hợp lệ.

* Thiết lập 2 – Huấn luyện trên dữ liệu đã cân bằng:

Dữ liệu huấn luyện được xử lý mất cân bằng bằng kỹ thuật SMOTE-ENN [8], bao gồm việc sinh thêm các mẫu lớp thiểu số và loại bỏ các điểm nhiễu của lớp đa số. Việc này giúp mô hình học được biên quyết định rõ ràng hơn và cải thiện khả năng nhận diện các giao dịch gian lận.



*Hình 1: Sơ đồ tổng quan quy trình đề xuất.*

Các mô hình được triển khai trong cả hai thiết lập bao gồm các thuật toán học máy truyền thống như Random Forest [15], XGBoost [16], K-Nearest Neighbors (KNN) [17], AdaBoost [18], cùng với hai kiến trúc mạng neural sâu là Base Convolutional Neural Network (CNN) [19] và Bidirectional Long Short-Term Memory (BiLSTM) [20].

*3.3 Phương pháp đánh giá*

Hiệu suất của các mô hình học máy (ML) và học sâu (DL) được đánh giá thông qua một tập hợp các chỉ số tiêu chuẩn phổ biến trong bài toán phát hiện gian lận, bao gồm:

* Precision [21]: Tỷ lệ các dự đoán gian lận đúng trên tổng số giao dịch được mô hình dự đoán là gian lận. Chỉ số này phản ánh mức độ đáng tin cậy của các dự đoán dương tính.

Precision = (1)

* Recall [21]: Tỷ lệ các giao dịch gian lận được phát hiện trên tổng số giao dịch gian lận thật sự. Recall đặc biệt quan trọng trong bài toán phát hiện gian lận, vì một giao dịch gian lận bị bỏ sót có thể gây thiệt hại nghiêm trọng.

Recall = (2)

* F1-score [21]: Trung bình điều hòa của Precision và Recall, cung cấp một thước đo cân bằng giữa độ chính xác và khả năng phát hiện. F1-score đặc biệt phù hợp khi dữ liệu mất cân bằng, giúp so sánh hiệu suất tổng thể của mô hình.

F1-score = (3)

* Accuracy: Tỷ lệ dự đoán đúng trên tổng số mẫu. Mặc dù Accuracy là chỉ số phổ biến, trong trường hợp dữ liệu mất cân bằng, chỉ số này có thể gây hiểu nhầm, vì mô hình có thể đạt Accuracy cao bằng cách luôn dự đoán lớp chiếm đa số (giao dịch hợp lệ). Do đó, Accuracy được sử dụng như tham chiếu bổ sung, không phải là chỉ số chính.

accuracy = (4)

* AUC-ROC (Area Under the Receiver Operating Characteristic Curve) [22]: Đo khả năng phân biệt hai lớp của mô hình ở nhiều ngưỡng dự đoán khác nhau. Giá trị AUC càng cao, khả năng phân biệt giữa giao dịch gian lận và hợp lệ càng tốt.

**4. Kết quả và thảo luận**

Phần này trình bày kết quả thực nghiệm của các mô hình học máy (ML) và học sâu (DL) bao gồm Random Forest, XGBoost, K-Nearest Neighbors (KNN), AdaBoost, cùng với hai kiến trúc sâu là Base Convolutional Neural Network (CNN) và Bidirectional Long Short-Term Memory (BiLSTM). Các mô hình được đánh giá trong hai thiết lập: (i) dữ liệu gốc có phân bố cực kỳ mất cân bằng và (ii) dữ liệu đã được tái lấy mẫu bằng SMOTE-ENN nhằm cân bằng lại tỷ lệ giữa hai lớp.

*4.1 Kết quả trên dữ liệu gốc (Imbalanced Data)*

Trong phần này, chúng tôi trình bày hiệu suất của các mô hình học máy (ML) và học sâu (DL) khi được huấn luyện và kiểm thử trên bộ dữ liệu gốc chưa xử lý mất cân bằng. Do tỷ lệ số lượng giao dịch gian lận chỉ chiếm 0,172% tổng số mẫu, các mô hình có xu hướng nghiêng về lớp đa số, gây ảnh hưởng đáng kể đến các chỉ số phản ánh khả năng phát hiện đúng gian lận (class 1). Vì vậy, nghiên cứu sử dụng các thước đo nhạy cảm đối với mất cân bằng như Precision, Recall, F1-score, và ROC-AUC, bên cạnh Accuracy và Confusion Matrix nhằm cung cấp cái nhìn toàn diện hơn về hiệu quả dự đoán.

*Bảng 2: Hiệu suất của các mô hình trên dữ liệu gốc*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mô hình** | **Precision** | **Recall** | **F1-score** | **Accuracy** | **ROC-AUC** |
| Random Forest | 0.730 | 0.684 | 0.707 | 0.999 | 0.842 |
| XGBoost | 0.962 | 0.526 | 0.680 | 0.999 | 0.899 |
| KNN | 0.970 | 0.674 | 0.795 | 0.999 | 0.889 |
| AdaBoost | 0.732 | 0.632 | 0.678 | 0.999 | 0.816 |
| CNN | 0.850 | 0.716 | 0.777 | 0.999 | 0.943 |
| Bi-LSTM | 0.864 | 0.737 | 0.795 | 0.999 | 0.958 |

Bảng 2 trình bày hiệu suất dự đoán của sáu mô hình khi được huấn luyện và kiểm thử trực tiếp trên bộ dữ liệu gốc chưa xử lý mất cân bằng. Kết quả cho thấy sự khác biệt đáng kể giữa nhóm mô hình học sâu (CNN, BiLSTM) và nhóm mô hình học máy truyền thống (Random Forest, XGBoost, KNN, AdaBoost).

Cụ thể, BiLSTM là mô hình đạt hiệu suất tổng thể cao nhất trên dữ liệu gốc, với F1-score đạt 0.795 và ROC-AUC đạt 0.958, phản ánh khả năng nhận diện mẫu gian lận tốt hơn so với các mô hình khác. CNN cũng cho kết quả tích cực, đạt F1-score 0.777 và ROC-AUC 0.943, cho thấy kiến trúc tích chập vẫn có thể khai thác hiệu quả các đặc trưng PCA ẩn danh trong bộ dữ liệu. Trong nhóm mô hình học máy truyền thống, KNN cho kết quả tốt nhất với F1-score 0.795, cao hơn cả XGBoost và Random Forest. Tuy nhiên, ROC-AUC của KNN (0.889) cho thấy khả năng phân tách ranh giới giữa hai lớp vẫn kém hơn so với các mô hình DL.Trong nhóm mô hình học máy truyền thống, KNN cho kết quả tốt nhất với F1-score 0.795, cao hơn cả XGBoost và Random Forest. Tuy nhiên, ROC-AUC của KNN (0.889) cho thấy khả năng phân tách ranh giới giữa hai lớp vẫn kém hơn so với các mô hình DL. Mặc dù XGBoost đạt Precision rất cao (0.962), Recall chỉ đạt 0.526, cho thấy mô hình phát hiện được rất ít mẫu gian lận và thường chỉ dự đoán gian lận khi rất chắc chắn. Do đó, F1-score của XGBoost chỉ đạt 0.680, thấp hơn nhiều so với nhóm DL. Điều này phù hợp với đặc điểm thực nghiệm khi gradient boosting thường nhạy với mất cân bằng và yêu cầu kỹ thuật tái mẫu để đạt hiệu quả tối ưu. Random Forest và AdaBoost đều cho F1-score ở mức trung bình (0.707 và 0.678), đồng thời có ROC-AUC thấp nhất trong nhóm ML (0.842 và 0.816), cho thấy các mô hình này gặp khó khăn trong việc mô hình hóa lớp gian lận có kích thước quá nhỏ.

Nhìn chung, các kết quả trên cho thấy:

* Các mô hình học sâu (CNN, BiLSTM) thích hợp hơn cho dữ liệu mất cân bằng và có đặc trưng trừu tượng như PCA.
* Các mô hình ML truyền thống đều bị ảnh hưởng mạnh bởi phân bố lớp lệch, đặc biệt là XGBoost và AdaBoost, vốn giảm mạnh Recall.
* Mặc dù Accuracy đều đạt 0.999, nhưng đây không phải thước đo phù hợp trong bối cảnh dữ liệu mất cân bằng và không phản ánh năng lực thực sự của mô hình.

Những kết quả này nhấn mạnh sự cần thiết của việc áp dụng các kỹ thuật xử lý mất cân bằng để cải thiện khả năng phát hiện giao dịch gian lận trước khi huấn luyện mô hình.

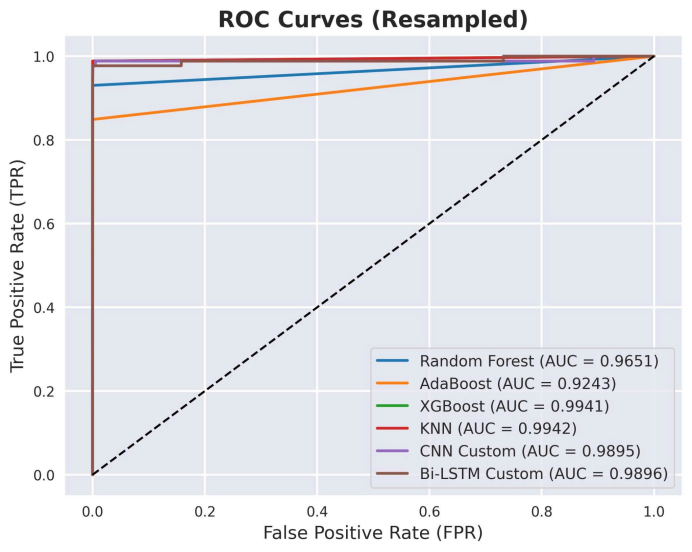
*4.2 Kết quả trên dữ liệu đã cân bằng (Balanced Data – SMOTE-ENN)*

Bảng 3 trình bày hiệu suất của các mô hình sau khi áp dụng kỹ thuật tái lấy mẫu SMOTE-ENN. Kết quả cho thấy các mô hình cải thiện đáng kể so với dữ liệu gốc chưa cân bằng, đặc biệt là về khả năng phát hiện giao dịch gian lận, thể hiện rõ qua sự tăng mạnh của chỉ số Recall. Việc cân bằng dữ liệu giúp mô hình nhận diện nhiều mẫu gian lận hơn, làm tăng độ nhạy và nâng cao hiệu quả phát hiện lớp thiểu số. Hầu hết các mô hình đều đạt Accuracy = 1.000, cho thấy dự đoán trở nên ổn định khi không còn chênh lệch lớn giữa các lớp; tuy nhiên, Accuracy không phản ánh đầy đủ khả năng nhận diện gian lận, do đó các chỉ số chính vẫn là Precision, Recall, F1-score và ROC-AUC. Xét về Precision, các mô hình XGBoost (0.960), KNN (0.950) và CNN (0.920) thể hiện khả năng hạn chế dự đoán dương tính giả. Về Recall, nhóm học sâu nổi bật với CNN đạt 0.942 và Bi-LSTM đạt cao nhất 0.953, chứng tỏ SMOTE-ENN giúp mô hình phát hiện nhiều giao dịch gian lận hơn. Chỉ số F1-score củng cố quan sát này, khi Bi-LSTM đạt 0.937 và CNN đạt 0.931. Xét theo ROC-AUC, CNN (0.989) và Bi-LSTM (0.990) vẫn vượt trội, cho thấy mô hình học sâu khai thác hiệu quả cấu trúc dữ liệu sau khi phân phối lớp được điều chỉnh hợp lý. Như vậy, sự cải thiện mạnh mẽ về Recall minh chứng cho hiệu quả của SMOTE-ENN trong việc tăng khả năng phát hiện các giao dịch gian lận.

*Bảng 3: Hiệu suất của các mô hình trên dữ liệu đã cân bằng bằng SMOTE-ENN*

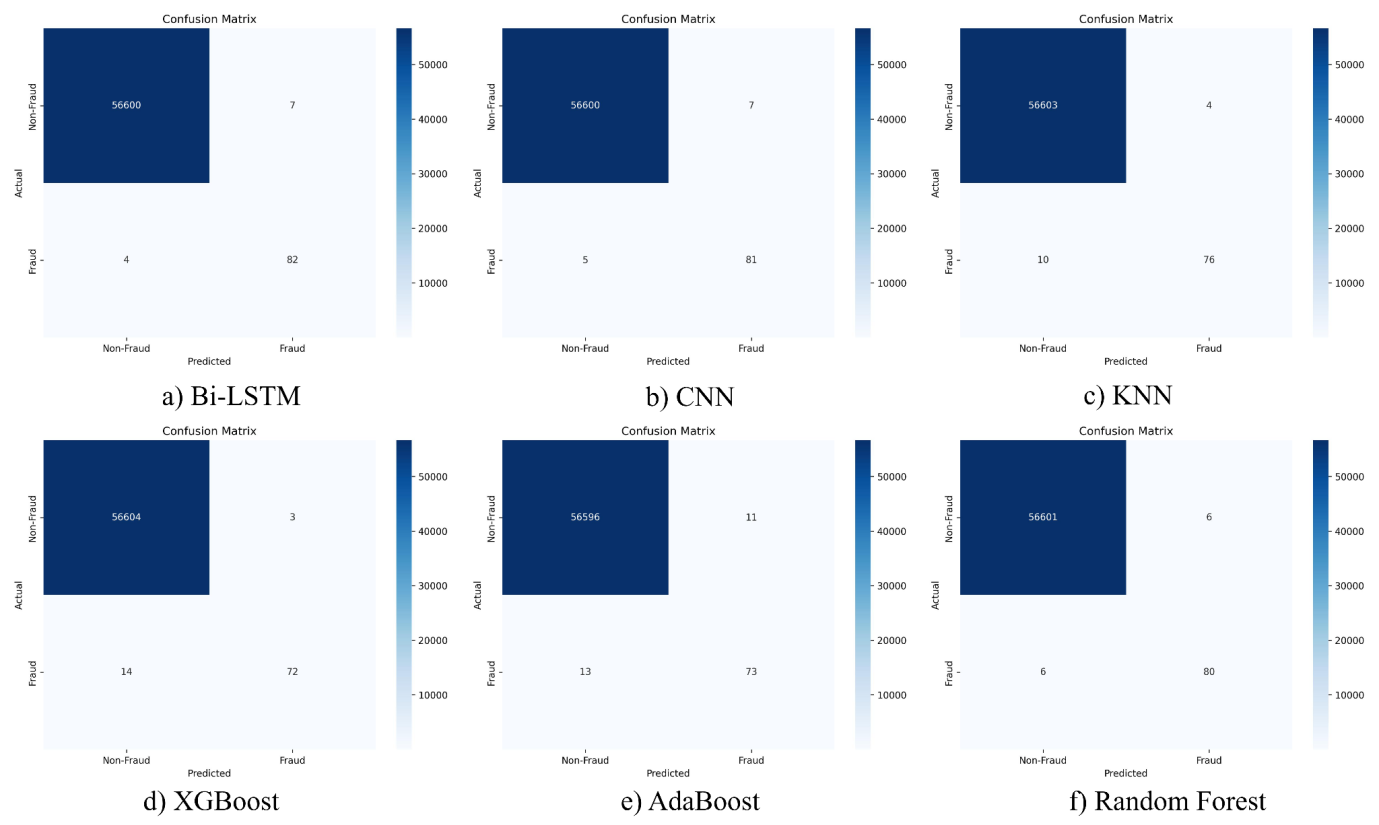
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mô hình** | **Precision** | **Recall** | **F1-score** | **Accuracy** | **ROC-AUC** |
| Random Forest | 0.930 | 0.930 | 0.930 | 1.000 | 0.965 |
| XGBoost | 0.960 | 0.837 | 0.894 | 1.000 | 0.994 |
| KNN | 0.950 | 0.884 | 0.916 | 1.000 | 0.994 |
| AdaBoost | 0.869 | 0.849 | 0.859 | 1.000 | 0.924 |
| CNN | 0.920 | 0.942 | 0.931 | 1.000 | 0.989 |
| Bi-LSTM | 0.921 | 0.953 | 0.937 | 1.000 | 0.990 |

Hình 2 trình bày đường cong ROC của sáu mô hình trên tập kiểm tra đã qua xử lý mất cân bằng. Có thể thấy rõ đường cong của Bi-LSTM nằm sát nhất về góc trên-trái và đạt giá trị AUC cao nhất (0,9896), tiếp theo sát nút là CNN Custom (AUC = 0,9895), XGBoost (AUC = 0,9941) và KNN (AUC = 0,9942). Hai mô hình còn lại là Random Forest (AUC = 0,9651) và AdaBoost (AUC = 0,9243) có diện tích dưới đường cong thấp hơn đáng kể.



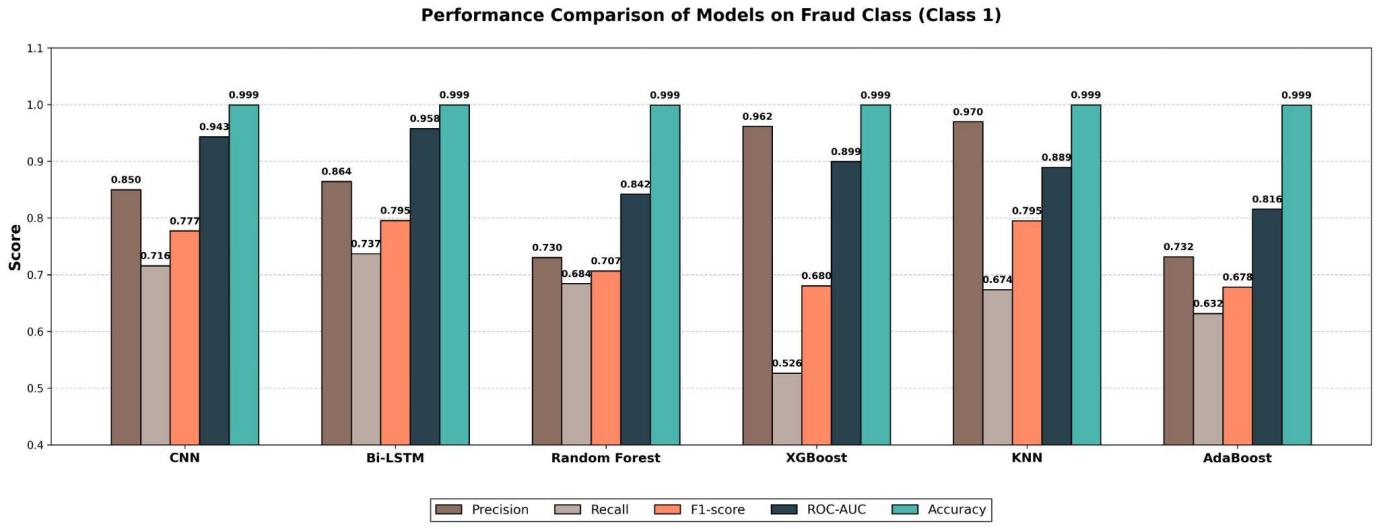
*Hình 2: Đường cong ROC của các mô hình trên tập dữ liệu đã được cân bằng bằng kỹ thuật SMOTE-ENN*

Hình 3 minh họa rõ ràng tác động tiêu cực của mất cân bằng lớp nghiêm trọng lên khả năng phát hiện gian lận thực tế. tất cả sáu mô hình đều thể hiện hiệu suất phân loại giao dịch không gian lận (Non-Fraud) rất cao, với số lượng True Negatives đạt khoảng 56.600 và False Positives dưới 11, cho thấy độ đặc hiệu gần như hoàn hảo. Tuy nhiên, hiệu quả phát hiện lớp thiểu số là giao dịch gian lận (Fraud) có sự khác biệt rõ rệt. Trong bối cảnh này, việc giảm thiểu False Negatives (FN) để tránh bỏ sót các giao dịch gian lận thực tế là yếu tố then chốt. Trong số các mô hình, Bi-LSTM nổi bật nhất với FN thấp nhất (4) và True Positives cao nhất (82), đạt Recall 95,3%, chứng tỏ khả năng phát hiện hầu hết các giao dịch gian lận trên tập dữ liệu đã cân bằng. Các mô hình khác như Random Forest cũng đạt hiệu suất tốt (FN = 6, TP = 80), trong khi AdaBoost có FN cao nhất (13), bỏ sót nhiều gian lận hơn. Kết quả này nhấn mạnh rằng, mặc dù tất cả mô hình đều duy trì độ đặc hiệu cao, Bi-LSTM là mô hình hiệu quả nhất, kết hợp tốt giữa khả năng phát hiện gian lận và độ chính xác trong dự đoán.

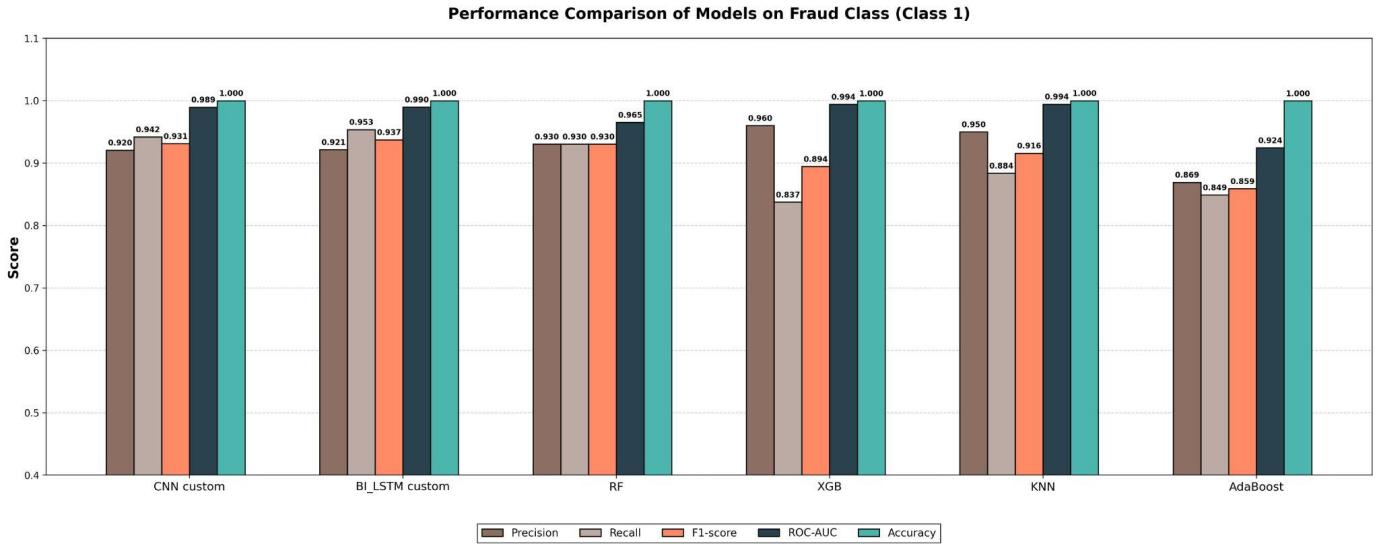


*Hình 3: Ma trận nhầm lẫn của các mô hình trên tập dữ liệu đã được cân bằng bằng kỹ thuật SMOTE-ENN*

Hình 3 và Hình 4 minh họa sự so sánh hiệu suất của các mô hình trước và sau bước tái lấy mẫu, tập trung vào các chỉ số đánh giá chính của lớp gian lận (Precision, Recall, F1-score, Accuracy và ROC-AUC). Các kết quả cho thấy tác động đáng kể của quá trình cân bằng dữ liệu đối với khả năng học và dự đoán của mô hình. Trên tập dữ liệu gốc, mặc dù Accuracy và ROC-AUC đạt giá trị cao, các mô hình lại thể hiện Recall và F1-score thấp trên lớp thiểu số, phản ánh hạn chế trong việc nhận diện các giao dịch gian lận thực sự. Sau khi áp dụng kỹ thuật tái lấy mẫu, độ nhạy (Recall) của lớp gian lận được cải thiện rõ rệt ở tất cả mô hình, kéo theo sự tăng trưởng tương ứng của F1-score. Đáng chú ý, các mô hình học sâu như Bi-LSTM và CNN đạt mức hiệu suất vượt trội, cho thấy khả năng khai thác hiệu quả hơn các đặc trưng sau khi dữ liệu được cân bằng.



*Hình 3: Biểu đồ hiệu suất các mô hình trên dữ liệu gốc*



*Hình 4: Biểu đồ hiệu suất các mô hình trên tập dữ liệu cân bằng (SMOTE-ENN)*

Nhìn chung, các kết quả từ Bảng 3 và Hình 2–4 cho thấy rõ ràng vai trò quan trọng của việc cân bằng dữ liệu trong việc cải thiện hiệu suất của các mô hình phân loại giao dịch gian lận. Việc áp dụng kỹ thuật SMOTE-ENN không chỉ nâng cao độ nhạy (Recall) và F1-score của lớp thiểu số mà còn giúp các mô hình học sâu khai thác tốt hơn cấu trúc dữ liệu, từ đó đạt được khả năng phân tách hai lớp chính xác hơn, như thể hiện qua các giá trị ROC-AUC cao. Đồng thời, quá trình tái lấy mẫu cũng làm cho các dự đoán trở nên ổn định và đồng đều, giảm thiểu tình trạng dự đoán sai lệch giữa các lớp. Những quan sát này nhấn mạnh rằng, trong bối cảnh dữ liệu mất cân bằng, việc kết hợp các mô hình học sâu với các kỹ thuật cân bằng dữ liệu là chiến lược hiệu quả để nâng cao khả năng nhận diện các giao dịch gian lận một cách toàn diện.

*4.2 Thảo luận*

Kết quả thực nghiệm cho thấy dữ liệu mất cân bằng ảnh hưởng mạnh đến khả năng nhận diện giao dịch gian lận. Trên dữ liệu gốc, mặc dù Accuracy của tất cả các mô hình đều đạt 0.999, Recall và F1-score trên lớp gian lận thấp, ví dụ XGBoost chỉ đạt Recall = 0.526 và F1-score = 0.680, trong khi các mô hình học sâu như Bi-LSTM và CNN đạt F1-score lần lượt là 0.795 và 0.777, cùng ROC-AUC cao hơn (Bi-LSTM 0.958, CNN 0.943). Sau khi cân bằng dữ liệu bằng SMOTE-ENN, hiệu năng cải thiện rõ rệt: Bi-LSTM đạt Recall = 0.953, F1-score = 0.937 và ROC-AUC = 0.990, trong khi CNN cũng đạt Recall = 0.942, F1-score = 0.931 và ROC-AUC = 0.989. Nhóm học máy truyền thống như XGBoost và KNN cũng được cải thiện đáng kể, với F1-score lần lượt 0.894 và 0.916, ROC-AUC 0.994. Các kết quả này cho thấy việc tái lấy mẫu giúp nâng cao độ nhạy, giảm dự đoán sai lệch giữa các lớp và làm dự đoán ổn định hơn. Nhìn chung, các quan sát củng cố rằng trong bối cảnh dữ liệu gian lận cực kỳ hiếm, việc kết hợp các mô hình học sâu với kỹ thuật SMOTE-ENN là chiến lược hiệu quả, vừa tăng khả năng phát hiện lớp thiểu số vừa giữ được tính ổn định và phân tách hai lớp chính xác.

**5. Kết luận**

Nghiên cứu này đánh giá hiệu quả của các mô hình học máy (Random Forest, XGBoost, KNN, AdaBoost) và học sâu (CNN, Bi-LSTM) trong việc phát hiện giao dịch gian lận trên dữ liệu cực kỳ mất cân bằng. Kết quả thực nghiệm cho thấy các mô hình học sâu vượt trội hơn các mô hình học máy truyền thống về Recall, F1-score và ROC-AUC, nhờ khả năng khai thác các đặc trưng trừu tượng từ dữ liệu. Trên dữ liệu gốc, Bi-LSTM đạt F1-score 0.795 và ROC-AUC 0.958, trong khi CNN đạt F1-score 0.777 và ROC-AUC 0.943. Sau khi áp dụng SMOTE-ENN để cân bằng dữ liệu, hiệu năng của tất cả các mô hình cải thiện rõ rệt. Bi-LSTM đạt F1-score 0.937 và ROC-AUC 0.990, CNN đạt F1-score 0.931 và ROC-AUC 0.989, chứng tỏ cân bằng dữ liệu giúp nâng cao khả năng phát hiện lớp thiểu số và làm dự đoán ổn định hơn. Các chỉ số như Accuracy không phản ánh đầy đủ năng lực của mô hình trong bối cảnh mất cân bằng, do đó Precision, Recall, F1-score và ROC-AUC là các thước đo quan trọng để đánh giá hiệu quả. Hướng phát triển trong tương lai bao gồm: (i) thử nghiệm các kỹ thuật cân bằng dữ liệu tiên tiến hơn như ADASYN hoặc kết hợp oversampling và undersampling; (ii) áp dụng các kiến trúc học sâu hiện đại hơn như Transformer hoặc Graph Neural Networks để phát hiện các mẫu gian lận phức tạp; (iii) tích hợp các đặc trưng thời gian thực và ngữ cảnh giao dịch để nâng cao độ chính xác; và (iv) xây dựng hệ thống học trực tuyến (online learning) giúp mô hình tự điều chỉnh theo phân phối dữ liệu mới. Tóm lại, kết hợp mô hình học sâu với các kỹ thuật cân bằng dữ liệu là chiến lược hiệu quả để nâng cao khả năng phát hiện giao dịch gian lận, đồng thời vẫn còn nhiều hướng nghiên cứu tiềm năng để tối ưu hóa hiệu suất và ứng dụng thực tế.

***Tài liệu tham khảo***

[1] European Banking Authority. *EBA / ECB report on payment fraud (recent editions e.g. 2024)*. [eba.europa.eu/sites/default/files/2024-08/465e3044-4773-4e9d-8ca8-b1cd031295fc/EBA\_ECB 2024 Report on Payment Fraud.pdf?](https://eba.europa.eu/sites/default/files/2024-08/465e3044-4773-4e9d-8ca8-b1cd031295fc/EBA_ECB%202024%20Report%20on%20Payment%20Fraud.pdf?)

[2] Richard J. Bolton. David J. Hand. "Statistical Fraud Detection: A Review." Statist. Sci. 17 (3) 235 - 255, August 2002. <https://doi.org/10.1214/ss/1042727940>

[3] Bhattacharyya, Siddhartha, et al. "Data mining for credit card fraud: A comparative study." *Decision support systems* 50.3 (2011): 602-613. [https://doi.org/10.1016/j.dss.2010.08.00](https://doi.org/10.1016/j.dss.2010.08.008)8

[4] Jurgovsky, Johannes, et al. "Sequence classification for credit-card fraud detection." Expert systems with applications 100 (2018): 234-245. <https://doi.org/10.1016/j.eswa.2018.01.037>

[5] Raghavan, Pradheepan, and Neamat El Gayar. "Fraud detection using machine learning and deep learning." 2019 international conference on computational intelligence and knowledge economy (ICCIKE). IEEE, 2019. [https://doi.org/](https://doi.org/10.1016/j.eswa.2018.01.037)[10.1109/ICCIKE47802.2019.9004231](https://doi.org/10.1109/ICCIKE47802.2019.9004231)

[6] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). *A Comprehensive Survey of Data Mining-based Fraud Detection Research*. arXiv:1009.6119 <https://arxiv.org/abs/1009.6119>

[7] Ndama, Oussama, Ismail Bensassi, and El Mokhtar En-Naimi. "Optimizing credit card fraud detection: a deep learning approach to imbalanced datasets." *International Journal of Electrical & Computer Engineering (2088-8708)* 14.4 (2024). <https://ijece.iaescore.com/index.php/IJECE/article/view/35029>

[8] Mienye, Ibomoiye Domor, and Yanxia Sun. "A deep learning ensemble with data resampling for credit card fraud detection." Ieee Access 11 (2023): 30628-30638. [https://doi.org/](https://doi.org/10.1016/j.eswa.2018.01.037)[10.1109/ACCESS.2023.3262020](https://doi.org/10.1109/ACCESS.2023.3262020)

[9]Alarfaj, Fawaz Khaled, et al. "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms." *Ieee Access* 10 (2022): 39700-39715. [https://doi.org/](https://doi.org/10.1016/j.eswa.2018.01.037)[10.1109/ACCESS.2022.3166891](https://doi.org/10.1109/ACCESS.2022.3166891)

[10] Stojanović, Branka, et al. "Follow the trail: Machine learning for fraud detection in Fintech applications." *Sensors* 21.5 (2021): 1594. <https://doi.org/10.3390/s21051594>

[11] Zaffar, Zaffar, et al. "Credit card fraud detection with subspace learning-based one-class classification." *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2023. [https://doi.org/](https://doi.org/10.1016/j.eswa.2018.01.037)[10.1109/SSCI52147.2023.10372038](https://doi.org/10.1109/SSCI52147.2023.10372038)

[12] Zhou, Ying, et al. "A user-centered explainable artificial intelligence approach for financial fraud detection." *Finance Research Letters* 58 (2023): 104309. <https://doi.org/10.1016/j.frl.2023.104309>

**[13] Rao, Susie Xi, et al. "xFraud: explainable fraud transaction detection." *arXiv preprint arXiv:2011.12193* (2020). <https://doi.org/10.48550/arXiv.2011.12193>**

[14] Machine Learning Group-ULB, “Credit card fraud detection,” Kaggle, 2018. <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

[15] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32. <https://doi.org/10.1023/A:1010933404324>

[16] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016. <https://doi.org/10.1145/2939672.2939785>

[17] Cover, Thomas, and Peter Hart. "Nearest neighbor pattern classification." IEEE transactions on information theory 13.1 (1967): 21-27. [https://doi.org/](https://doi.org/10.1023/A:1010933404324)[10.1109/TIT.1967.1053964](https://doi.org/10.1109/TIT.1967.1053964)

[18] Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences* 55.1 (1997): 119-139. <https://doi.org/10.1006/jcss.1997.1504>

[19] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (2002): 2278-2324. [https://doi.org/](https://doi.org/10.1023/A:1010933404324)[10.1109/5.726791](https://doi.org/10.1109/5.726791)

[20] Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional recurrent neural networks." *IEEE transactions on Signal Processing* 45.11 (1997): 2673-2681. [https://doi.org/](https://doi.org/10.1023/A:1010933404324)[10.1109/78.650093](https://doi.org/10.1109/78.650093)

[21] Powers, David MW. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." [https://doi.org/](https://doi.org/10.1023/A:1010933404324)[10.9735/2229-3981](https://doi.org/10.9735/2229-3981)

[22] Fawcett, Tom. "An introduction to ROC analysis." *Pattern recognition letters* 27.8 (2006): 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>