# PHÁT HIỆN GIAN LẬN THẺ TÍN DỤNG BẰNG HỌC MÁY

**Hoàng Thị Thúy**
*Trường Đại học Luật Thành phố Hồ Chí Minh*
*Email: htthuy@hcmulaw.edu.vn*
**Lê Thị Xuân Thu**
*Trường Đại học Luật Thành phố Hồ Chí Minh*
*Email: ltxthu@hcmulaw.edu.vn*

Ngày nhận: 05/4/2020
Ngày nhận bản sửa: 23/4/2020
Ngày duyệt đăng: 05/3/2021

**Tóm tắt**:
*Trong thời đại khoa học – công nghệ, khách hàng chỉ với một tấm thẻ tín dụng có thể thanh toán toàn bộ các hoạt động mua sắm của mình một cách nhanh chóng và tiện lợi bất kể đang ở vị trí nào. Tuy nhiên, bên cạnh sự tiện lợi thì cũng đi kèm theo lỗ hổng với các mối đe dọa về tổn thất tài chính. Người bán không thể kiểm tra người mua có phải chủ thẻ hay không bởi vì cả thẻ và chủ thẻ đều không cần thiết có mặt tại thời điểm giao dịch. Do đó mà việc gian lận thẻ tín dụng ngày càng trở nên phổ biến với nhiều cách thức khác nhau. Trong bài viết này các tác giả tập trung phát hiện gian lận thẻ tín dụng bằng các phương pháp của học máy* là mô hình mạng nơ ron nhân tạo *và mô hình máy vec-tơ hỗ trợ và so sánh mức độ hiệu quả của hai phương pháp với bộ dữ liệu Paysim. Kết quả cho thấy mô hình mạng nơ ron nhân tạo phát hiện với độ chính xác cao hơn ( 99%) .*
**Từ khóa**: Gian lận thẻ tín dụng, học máy, mạng nơ ron nhân tạo, máy vec-tơ hỗ trợ.
**Mã JEL**: **C38, C45, B26**

**Using machine learning to detect frauds involving credit card use**

***Abstract:***
*With the development of information technology, modern customers can make payments for any purchase with no barriers. However, it also brings about some potential financial harms. In fact, the seller cannot check whether the buyer is the cardholder due to the fact that both of them are not necessarily to be present at the place of purchase. Accordingly, frauds involving credit card use are becoming increasingly. In this study, we focus on how to detect this kind of fraud with machine learning, namely Artificial Neural Network and Support Vector Machine modes. Then, some com-parisons between the effectiveness of the two methods with the Paysim dataset. The results showed that the artificial neural network model was better with an accuracy of 99%.*

***Keywords:*** *Artificial neural network, credit card fraud, machine learning, support vector machine.*
***JEL Codes: C38, C45, B26***

## 1. Giới thiệu và tổng quan nghiên cứu

Gian lận thẻ tín dụng là hình thức sử dụng các công nghệ cao để đánh cắp thông tin thẻ tín dụng (như visa, ATM, MasterCard…) từ người sở hữu khác để thực hiện các giao dịch tài chính, ngân hàng. Hai cơ chế chính để tránh gian lận và tổn thất do các hoạt động gian lận là phòng chống gian lận và hệ thống phát hiện gian lận. Phòng chống gian lận là chủ động vô hiệu hóa sự xuất hiện của gian lận. Hệ thống phát hiện gian lận phát huy tác dụng khi những kẻ lừa đảo vượt qua các hệ thống phòng chống gian lận và bắt đầu thực hiện một giao dịch gian lận.Theo đó, mục tiêu của các hệ thống phát hiện gian lận là kiểm tra mọi giao dịch

có khả năng là gian lận bất kể các cơ chế phòng ngừa, và để xác định những kẻ lừa đảo càng nhanh càng tốt sau khi kẻ lừa đảo bắt đầu thực hiện một giao dịch gian lận. Bài đánh giá về hệ thống phát hiện gian lận có thể được tìm thấy trong các nghiên cứu của Bolton & Hand (2002), Kou & cộng sự (2004), Phua & cộng sự (2005), Sahin & Duman (2010).

Gian lận thẻ tín dụng có thể được thực hiện bằng nhiều cách như trộm cắp đơn giản, các ứng dụng gian lận, thẻ giả, gian lận trực tuyến…Trong lừa đảo trực tuyến, giao dịch được thực hiện từ xa và chỉ cần các thông tin về thẻ. Mặc dù các cơ chế phòng ngừa như *chip* và *pin* làm giảm các hoạt động trộm cắp đơn giản, thẻ giả, nhưng số lượng gian lận trực tuyến vẫn tăng nhanh chóng gây nên tổn thất rất lớn về tài chính nguyên nhân là do việc sử dụng thẻ tín dụng ngày càng phổ biến. Nhiều báo cáo như Leonard (1993) và Ghosh & Reilly (1994) cho thấy các khoản thất thoát lớn ở các quốc gia khác nhau. Theo báo cáo của Visa về các nước châu Âu, khoảng 50% của toàn bộ gian lận thẻ tín dụng thua lỗ trong năm 2008 là do gian lận trực tuyến (Visa Inc., 2008). Phát hiện gian lận là công việc không hề dễ dàng nhưng nó là vấn đề cấp bách cần được giải quyết. Một trong những nguyên nhân gây nên khó khăn trong việc phát hiện gian lận thẻ tín dụng là bộ dữ liệu về giao dịch không được cung cấp và kết quả bị kiểm duyệt. Bất chấp nhiều khó khăn các nhà nghiên cứu vẫn nỗ lực để tìm kiếm và tối ưu hóa các mô hình nhằm nâng cao việc phòng chống và phát hiện gian lận. Phương pháp thường được sử dụng để phát hiện gian lận là sử dụng các thuật toán của học máy mà tiêu biểu là mạng nơ ron nhân tạo và máy vec-tơ hỗ trợ. Những kỷ thuật này có thể thực hiện độc lập hoặc kết hợp lại với nhau tạo nên các phương tiện phân loại và phát hiện các giao dịch bất thường. Dữ liệu quá khứ về thẻ tín dụng được sử dụng để tạo thành một kho dữ liệu đại diện cho hồ sơ người dùng của khách hàng. Những hồ sơ này bao gồm các biến, mỗi biến trong số đó tiết lộ một đặc tính hành vi của khách hàng. Những biến này có thể cho thấy thói quen chi tiêu của khách hàng, vị trí thực hiện giao dịch, ngày giờ giao dịch. Sau đó, các biến này được dùng để xây dựng một mô hình trong các hệ thống phát hiện gian lận. Sẽ có sự báo động những giao dịch bất thường khi mà có sai lệch đáng kể từ giao dịch mới với các giao dịch quá khứ trong hồ sơ của khách hàng.

Có rất nhiều nghiên cứu được thực hiện về phát hiện gian lận thẻ tín dụng: Shen & cộng sự (2007) chứng minh sự hiệu quả của các mô hình phân loại cho vấn đề phát hiện gian lận thẻ tín dụng và tác giả đề xuất ba mô hình phân loại là cây quyết định, mạng nơ ron nhân tạo và hồi quy logistic. Trong ba mô hình thì mạng nơ ron nhân tạo và hồi quy logistic vượt trội hơn cây quyết định. Islam & cộng sự (2007) đề xuất khung lý thuyết xác suất: lý thuyết Bayes, phân loại Bayes đơn giản, phân loại k hàng xóm gần nhất được thực hiện và áp dụng cho bộ dữ liệu hệ thống thẻ tín dụng. Sahin & Duman (2011) đã trích dẫn nghiên cứu về phát hiện gian lận thẻ tín dụng và sử dụng bảy phương pháp phân loại khác nhau, trong đó bao gồm mô hình cây quyết định và máy vec-tơ hỗ trợ để giảm rủi ro của các ngân hàng. Sahin & Duman đã đề nghị mạng nơ ron nhân tạo và hồi quy logistic là mô hình hữu ích hơn để cải thiện hiệu suất trong việc phát hiện các gian lận.

Tất cả những nghiên cứu trên đây là nền tảng cơ sở lý thuyết vững chắc cho nghiên cứu này của các tác giả. Trong khuôn khổ bài viết, ngoài việc minh họa ứng dụng của học máy thông qua sử dụng hai mô hình mạng nơ ron nhân tạo và máy vec-tơ hỗ trợ để phát hiện gian lận với bộ dữ liệu Paysim, các tác giả còn tiến hành so sánh hiệu quả của hai phương pháp tạo điều kiện linh hoạt cho các nhà quản lý lựa chọn phương pháp hiệu quả trong việc phát hiện gian lận tránh những tổn thất to lớn do việc lừa đảo, gian lận thẻ tín dụng gây ra.

## 2.Cơ sở lý thuyết và phương pháp nghiên cứu

### *2.1. Học máy*

Học máy (Machine Learning) là một phương tiện trong trí tuệ nhân tạo, sử dụng các thuật toán cho phép máy tính có thể tự học từ dữ liệu để giải quyết những vấn đề cụ thể như làm cho máy tính có khả năng nhận thức cơ bản của con người (nghe, nhìn, hiểu, giải toán,…) và hỗ trợ cho con người xử lý một lượng thông tin khổng lồ phải đối diện hằng ngày (Vũ Hữu Tiệp, 2018). Học máy đóng một vai trò quan trọng trong nhiều ngành khoa học và các ứng dụng của nó là một phần trong cuộc sống hàng ngày của chúng ta. Học máy được sử dụng để lọc thư rác điện tử, để dự đoán thời tiết, trong chẩn đoán y tế, khuyến cáo sản phẩm, nhận diện khuôn mặt, phát hiện gian lận thẻ tín dụng, v.v.

Dựa vào tính chất của tập dữ liệu, các thuật toán của học máy có thể phân thành 2 nhóm cơ bản đó là: học có giám sát (supervised learning) và học không giám sát (unsupervised learning). Học có giám sát bao gồm

các thuật toán đưa các dữ liệu đầu vào (input) thành các kết quả đầu ra (label) tương ứng. Đầu vào phải biết trước giá trị đầu ra tương ứng của chúng và được dùng để dự đoán giá trị biến đầu ra hay còn gọi là biến trả lời. Tùy thuộc vào biến đầu ra là rời rạc hay liên tục mà chúng ta có thể phân biệt hai nhiệm vụ được giám sát: phân loại (classification) hay hồi quy (regression). Phát hiện giân lận thẻ tín dụng thuộc nhóm đầu tiên bởi vì kết quả đầu ra là quan sát các giao dịch là của chủ thẻ hay là lừa đảo trong khi đó dự báo giá cổ phiếu thì thuộc nhóm hồi quy do biến đầu ra là là biến liên tục. Tuy nhiên trong cả hai nhóm thì biến đầu vào có thể là liên tục hoặc rời rạc. Dữ liệu trong các thuật toán thuộc nhánh học không giám sát chỉ có đầu vào mà không cần đầu ra. Nó được sử dụng chủ yếu để khám phá cấu trúc và mối quan hệ dữ liệu.

Một số thuật toán sử dụng phổ biến để phát hiện gian lận thẻ tín dụng hiện nay:

- Mạng nơ ron nhân tạo (Artifical Neural Network**)**
- Rừng ngẫu nhiên (Random Forrest)
- Logic mờ (Fuzzy Logic)
- Máy vec-tơ hỗ trợ (Support Vector Machine)
- Mạng Bayesian (Bayesian Network)
- K láng giềng gần nhất (K-nearest neighbor)
- Mô hình Markov ẩn (Hidden Markov model)
- Hồi quy Logistic (Logistic Regression)

Trong bài viết này, chúng tôi chỉ tập trung vào hai thuật toán: Mô hình mạng nơ ron nhân tạo và mô hình máy vec-tơ hỗ trợ bởi vì chúng phù hợp và đem lại hiệu quả cao đối với bộ dữ liệu được chọn nghiên cứu.

***2.2. Mô hình mạng nơ ron nhân tạo (Artifical Neural Network model)***

Mạng nơ ron nhân tạo là mô hình tính toán được mô phỏng dựa trên hoạt động của mạng nơ ron sinh học. Nó bao gồm số lượng lớn các nơ ron đơn lẻ gắn kết với nhau, xử lý thông tin bằng cách truyền các kết nối và tính các giá trị mới tại các nút. Có 3 tầng trong mạng nơ ron nhân tạo là: tầng vào (input layer), tầng ẩn (hidden layer) và tầng ra (output layer). Tầng vào biểu diễn thông tin đầu vào, tầng ẩn gồm các nút nhận ma trận đầu vào từ tầng trước, kết hợp với trọng số cùng với hàm kích hoạt phi tuyến như sigmoid, tanh để có được kết quả tầng ra. Mô hình mạng nơ ron nhân tạo gồm 2 quá trình tính toán cơ bản là: Lan truyền tiến và lan truyền ngược. Quá trình suy luận từ tầng vào cho tới tầng ra là quá trình lan truyền tiến (feedforward), tức là quá trình này chỉ có chiều hướng các nơ ron ở cùng một tầng lấy thông tin từ tầng trước mà không có chiều ngược lại:

$$\begin{aligned} a^{(0)} &= x \\ z^{(l)} &= \mathrm{W}^{(l)} a^{(l-1)} + b^{(l)}, \quad l = \overline{1, L} \quad (1) \\ a^{(\mathrm{l})} &= f^{(\mathrm{l})}(z^{(l)}) \\ \hat{y} &= a^{(L)} \end{aligned}$$

Ở đây x là đầu vào, $\mathrm{W}^{(l)}$ trọng số tương ứng ở tầng thứ $l, b^{(l)}$ hệ số điều chỉnh (bias) ở tầng $l$ hay còn gọi là một ngưỡng quyết định đầu ra. Hàm $f^{(l)}$ là hàm kích hoạt phi tuyến. Hàm kích hoạt thường dùng nhất là hàm Sigmoid $f(z) = \dfrac{1}{1+e^{-z}}$ (2) với đồ thị cân xứng thể hiện mức độ công bằng đối với các tham số. $\hat{y}$ chính là đầu ra dự đoán.

Lan truyền ngược (back propagation) là phương pháp để tính đạo hàm của hàm mất mát từ tầng cuối cùng đến tầng đầu tiên. Tầng cuối cùng được tính toán trước vì nó ảnh hưởng trực tiếp đến đầu ra. Hàm mất mát đạt giá trị nhỏ khi đầu ra dự đoán gần với đầu ra thực sự. Tùy theo mục đích là phân loại hay hồi quy ta có thể thiết kế hàm mất mát phù hợp. Giả sử hàm mất mát $J(\mathrm{W}, b, X, Y)$ với W, b lần lượt là ma trận trọng số và điều chỉnh. X, Y là cặp dữ liệu của tập huấn luyện.

Đạo hàm riêng của hàm mất mát theo một thành phần ma trận trọng số của tầng đầu ra L:

$$\frac{\partial J}{\partial \mathrm{w}_{\mathrm{ij}}^{(\mathrm{L})}} = \frac{\partial J}{\partial z_j^{(L)}} \frac{\partial z_j^{(L)}}{\partial \mathrm{w}_{\mathrm{ij}}^{(\mathrm{L})}} \quad (3)$$

Với $\frac{\partial z_j^{(L)}}{\partial \mathrm{w}_{\mathrm{ij}}^{(\mathrm{L})}} = a_i^{(L-1)}$ (4) bởi vì $z_j^{(L)} = \mathrm{w}_j^{(L)T} a^{(L-1)} + b_j^{(L)}$ (5) và $\frac{\partial J}{\partial z_j^{(L)}}$ thường là một đại lượng không khó để tính toán. Đối với đạo hàm riêng theo trọng số ở các tầng $l < L$, bằng quy nạp ngược từ cuối:

$$\frac{\partial J}{\partial \mathrm{w}_{\mathrm{ij}}^{(\mathrm{l})}} = \frac{\partial J}{\partial z_j^{(\mathrm{l})}} \frac{\partial z_j^{(\mathrm{l})}}{\partial \mathrm{w}_{\mathrm{ij}}^{(\mathrm{l})}} = \frac{\partial J}{\partial z_j^{(\mathrm{l})}} a^{(\mathrm{l}-1)} \quad (6)$$

Trong đó $\frac{\partial J}{\partial z^{(l)}} = \left( (\mathrm{W}^{(\mathrm{l}+1)})^T \frac{\partial J}{\partial z^{(l+1)}} \right) \frac{\partial a^{(\mathrm{l})}}{\partial z^{(l)}}$ (7) với $\frac{\partial J}{\partial z^{(l+1)}}$ được tính ở vòng lặp ngay trước đó.

***2.3. Mô hình máy vec-tơ hỗ trợ (Support Vector Machine Model)***

Máy vec-tơ hỗ trợ (Support vector machine, SVM) là một thuật toán hiệu quả, đặc biệt khi tính toán trên bộ dữ liệu lớn với mục đích là phân chia dữ liệu thành các nhóm riêng biệt.Ý tưởng của SVM là đi tìm một siêu phẳng phân tách dữ liệu tốt nhất. Gọi khoảng cách nhỏ nhất từ một điểm thuộc một lớp đến mặt phân chia là lề (margin). Cần tìm một siêu phẳng sao cho lề của hai lớp là như nhau. Độ rộng của lề càng lớn thì khả năng phân loại lỗi càng thấp. Do đó bài toán tối ưu trong SVM chính là đi tìm siêu phẳng phân chia có lề lớn nhất (Vũ Hữu Tiệp, 2018).

Bài toán tối ưu của SVM là đi tìm $\mathrm{w}, b$ sao cho lề đạt giá trị lớn nhất

$$(\mathrm{w}, b) = \arg\max \left\{ \frac{1}{\|\mathrm{w}\|_2} \min y_n (\mathrm{w}^T \mathrm{x}_n + \mathrm{b}) \right\} \quad (8)$$

Bởi vì khoảng cách từ mỗi điểm đến mặt phân chia không đổi, nên ta có thể giả sử với những điểm gần mặt phân chia nhất thỏa mãn điều kiện:

$$y_m (\mathrm{w}^T \mathrm{x}_m + \mathrm{b}) = 1 \quad (9)$$

Bài toán tối ưu bây giờ có thể đưa về dạng tối ưu có ràng buộc

$$\begin{cases} (\mathrm{w},\mathrm{b}) = \arg\max \dfrac{1}{\|\mathrm{w}\|_2} \\ y_n (\mathrm{w}^T \mathrm{x}_n + \mathrm{b}) \geq 1, \forall \mathrm{n} = \overline{1, \mathrm{N}} \end{cases} \quad (10)$$

Đây là bài toán quy hoạch toàn phương với hàm mục tiêu là hàm lồi chặt. Tuy nhiên, khi số chiều của không gian dữ liệu và số điểm dữ liệu *N* lớn, ta thường phải giải quyết thông qua bài toán đối ngẫu của bài toán này và sử dụng phương pháp nhân tử Lagrange. Lúc này bài toán đối ngẫu Lagrange đi tìm các giá trị $\lambda$ thỏa mãn:

$$\lambda = \arg\max_{\lambda} \sum_{n=1}^{N} \lambda_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \lambda_n \lambda_m y_n y_m x_n^T x_m \quad (11)$$

$$\lambda \geq 0$$

$$\sum_{n=1}^{N} \lambda_n y_n = 0$$

Đặt $S = \{\mathrm{n}: \lambda_n \neq 0\}$ và $N_s$ là số phần tử của S. Sau khi tìm được $\lambda$ thì ta có các tham số:

$$\mathrm{w} = \sum_{m \in S} \lambda_m y_m x_m \quad (12)$$

$$b = y_m - \mathrm{w}^T x_m \quad (13)$$

Ở đây $(x_m, y_m)$ là điểm dữ liệu bất kỳ nào đó nằm trên đường biên gốc, ta còn gọi Support Vector (vec-tơ hỗ trợ). Trong thực tế *b* thường được tính bằng trung bình cộng của các *b* theo mỗi $n \in S$ vì ổn định hơn

trong quá trình tính toán:

$$b = \frac{1}{N_s}\sum_{n\in S}\left(y_n - \mathrm{w}^T x_n\right) = \frac{1}{N_s}\sum_{n\in S}\left(y_n - \sum_{m\in S}\lambda_m y_m x_m^T x_n\right) \quad (14)$$

Khi đó một điểm của dữ liệu sẽ được phân loại dựa vào dấu của biểu thức:

$$\sum_{n\in S}\lambda_m y_m x_m^T x + b \quad (15)$$

***2.4. Ứng dụng của mô hình mạng nơ ron nhân tạo và mô hình máy vec-tơ hỗ trợ phát hiện gian lận thẻ tín dụng***

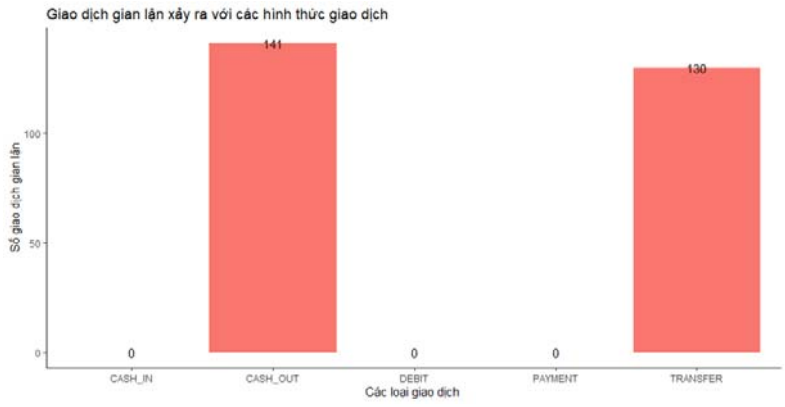
*2.4.1. Thu thập dữ liệu*

Trong thực tế thì rất khó thu thập được dữ liệu thực mô tả hành vi khách hàng giao dịch thông qua thẻ tín dụng bởi vì nó là các thông tin bảo mật của ngân hàng. Vì vậy, dữ liệu trong nghiên cứu này được các tác giả sử dụng là Paysim. Paysim là một bộ dữ liệu của Edgar & cộng sự (2016) được tạo bằng trình giả lập. Dữ liệu được tổng hợp từ bộ dữ liệu riêng để tạo thành bộ dữ liệu giống với hoạt động bình thường của các giao dịch, trong đó có các giao dịch bất thường, gian lận. Dữ liệu mô phỏng các giao dịch trên điện thoại di động dựa trên một mẫu các giao dịch thực được trích từ nhật ký tài chính tháng của một dịch vụ tiền điện thoại di động được thực hiện ở một quốc gia châu Phi. Nhật ký ban đầu được cung cấp bởi một công ty đa quốc gia, nhà cung cấp dịch vụ tài chính di động hiện đang hoạt động tại hơn 14 quốc gia trên toàn thế giới. Tuy nhiên, bộ dữ liệu khá lớn nên ở đây tác giả chỉ trích xuất 55% từ bộ dữ liệu gốc để thực hiện nghiên cứu trong phạm vi bài viết này.

Một số biến trong dữ liệu:

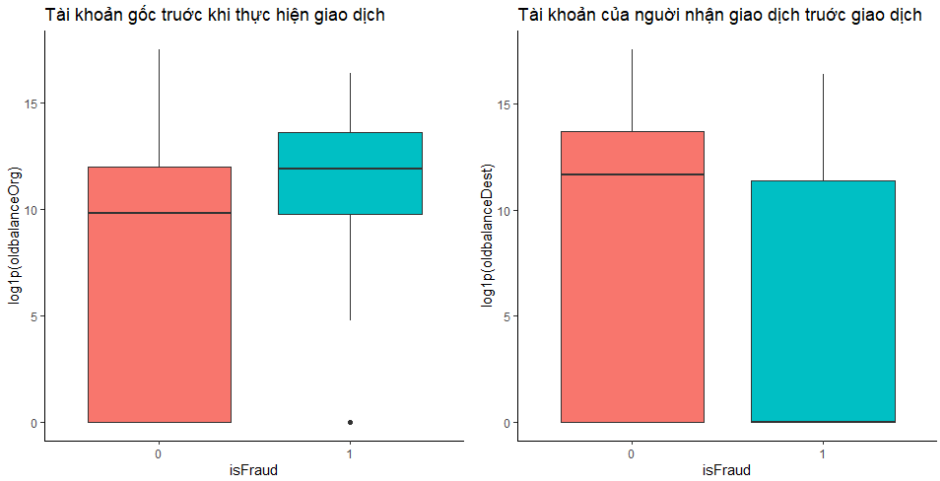
- StepMaps: đơn vị thời gian, 1 bước là 1 giờ.
- Type: Các loại hình giao dịch: CASH IN (tiền vào), CASH OUT (tiền ra), DEBIT (ghi nợ), PAYMENT (thanh toán), TRANSFER (chuyển khoản).
- Amount: lượng tiền trong mỗi lần giao dịch (nội tệ).
- NameOrig: Người thực hiện giao dịch.
- OldbalanceOrg: Số dư tài khoản gốc trước khi giao dịch.
- NewbalanceOrig: Số dư tài khoản gốc sau khi giao dịch.
- NameDest: Người nhận giao dịch.
- OldbalanceDest: Số dư tài khoản của người nhận trước khi giao dịch.
- NewbalanceDest: Số dư tài khoản của người nhận sau khi giao dịch.
- IsFraud: Xác nhận một giao dịch có gian lận hay không (0 là bình thường, 1 là giao dịch gian lận).

**Hình 1: Giao dịch gian lận xảy ra với các hình thức giao dịch**




*Nguồn: Các tác giả thực hiện*

**Hình 2: Lượng tiền liên quan đến giao dịch gian lận**




*Nguồn: Các tác giả thực hiện*

- IsFlaggedFraudflags: Chuyển bất hợp pháp hơn 200,000 trong một giao dịch.

Các biến tham gia vào mô hình mạng nơ ron nhân tạo và máy véc-tơ hỗ trợ: Amount, OldbalanceOrg, NewbalanceOrig, OldbalanceDest, NewbalanceDest, IsFraud.

Trong số lượng giao dịch 574255 lượt có 271 lượt gian lận chiếm 0.05%, trong đó chỉ có 2 loại hình giao dịch bị gian lận (lừa đảo, trộm cắp) là Cash out (tiền rút ra) và Transfer (chuyển khoản) (*xem Hình 1*).

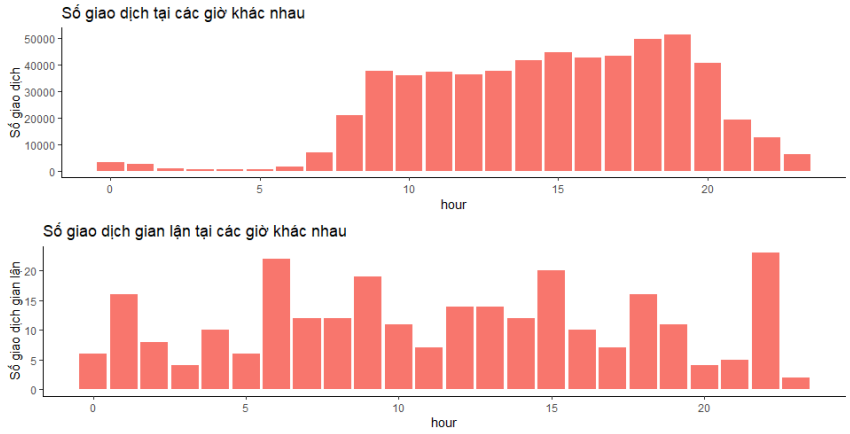
Phần lớn ở các giao dịch gian lận: số dư của tài khoản nguồn bị đánh cắp cao hơn số dư ở những tài khoản nguồn không bị đánh cắp. Trong khi đó ở tài khoản được chuyển tới thì số tiền giao dịch gian lận được chuyển tới các tài khoản có số dư thấp. Điều này cũng đúng thực tế, những tài khoản nhiều tiền hơn mặt bằng chung thường bị đánh cắp chuyển qua những tài khoản ít tiền hơn (so với mặt bằng chung) (*xem Hình 2*).

Về thời gian xảy ra giao dịch gian lận: Dựa vào đồ thị có thể thấy khung giờ từ 00 giờ đến 07 giờ các giao dịch thông thường rất ít xảy ra. Nhưng đó lại là thời điểm giao dịch gian lận xảy ra cao nhất (*xem Hình 3*).

*2.4.2. Phương pháp nghiên cứu*

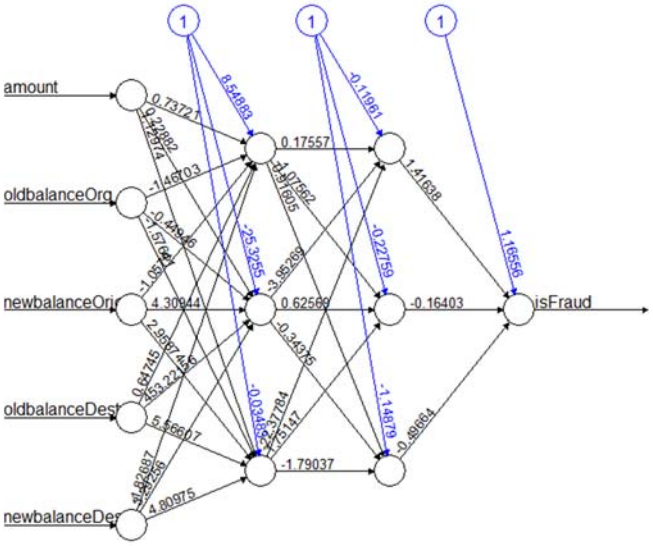
Từ bộ dữ liệu tác giả đã sử dụng các thuật toán học máy khác nhau để phát hiện gian lận trong hệ thống thẻ tín dụng: Mạng nơ ron nhân tạo, máy vec-tơ hỗ trợ. Từ đó xác định thuật toán phù hợp nhất để xác định gian lận, đồng thời so sánh hiệu quả của hai phương pháp trên cơ sở đo lường định lượng chẳng hạn như độ

**Hình 3: Thời gian giao dịch được phân theo nhóm giao dịch bình thường và giao dịch có xảy ra gian lận**




*Nguồn: Các tác giả thực hiện.*

**Hình 4: Mô hình mạng thần kinh cho tập dữ liệu thu được**




*Nguồn: Các tác giả thực hiện*

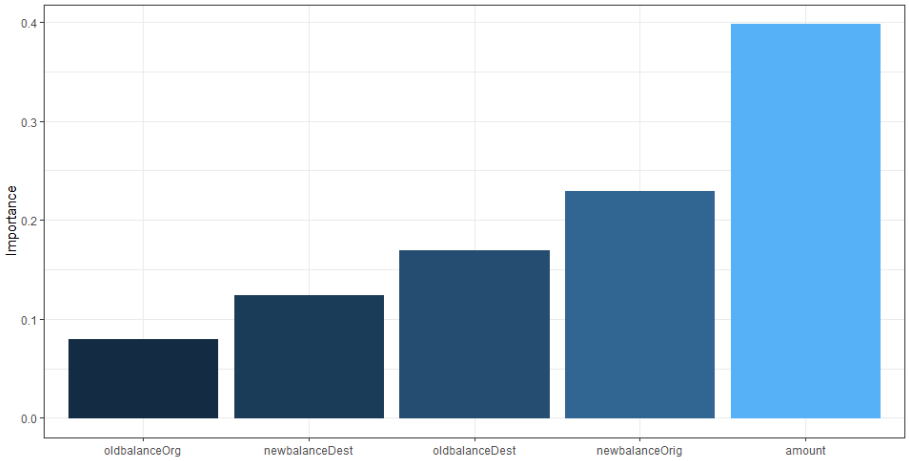
chính xác, tỷ lệ lỗi, độ nhạy, độ đặc hiệu.

Các bước được xử lý để phát hiện thuật toán tốt nhất cho tập dữ liệu đã cho:

- Bước 1: Đọc dữ liệu và khám phá dữ liệu.
- Bước 2: Lấy mẫu ngẫu nhiên trên tập dữ liệu để làm cân bằng dữ liệu.
- Bước 3: Chia tập dữ liệu thành hai phần, tập dữ liệu huấn luyện và tập kiểm tra và so sánh kết quả dự báo.
- Bước 4: Cho chạy bộ dữ liệu với mô hình mạng nơ ron nhân tạo và mô hình máy vec-tơ hỗ trợ.
- Bước 5: Tính độ chính xác và vẽ mô hình đánh giá độ chính xác của phép đo liên tục.
- Bước 6: Lấy thuật toán tốt nhất dựa trên hiệu quả với tập dữ liệu đã cho.

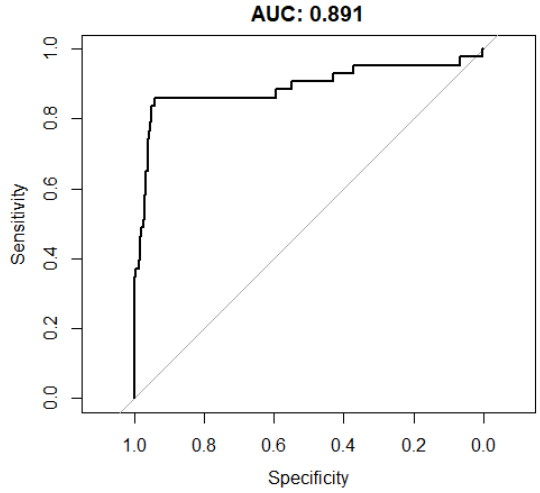
Trong bài nghiên cứu này các tác giả đã sử dụng phần mềm R để phân tích dữ liệu và thực hiện các bước trên.

**Hình 5: Mức độ quan trọng của các biến trong mô hình mạng nơ ron nhân tạo**




*Nguồn: Các tác giả thực hiện.*

**Hình 6: Đồ Thị biểu diện độ chính xác của mô hình Máy vec-tơ hỗ trợ**




*Nguồn: Do các tác giả thực hiện.*

### 3. Kết quả nghiên cứu

Dữ liệu huấn luyện: lấy 80% từ tập dữ liệu (459404 lượt giao dịch).

Dữ liệu kiểm tra: phần còn lại từ tập dữ liệu (114851 lượt giao dịch).

Để đánh giá và so sánh hiệu suất của hai mô hình nghiên cứu, các tác giả đã sử dụng các thông số sau:

Độ chính xác được tính theo công thức sau: $\frac{TP+TN}{TP+FP+FN+TN}$, trong đó TP (True Positive) là *dương tính thật,* TN (True Negative) là âm tính thật, FP (False Positive) là *dương tính giả,* FN (False Negative) là âm tính *giả. TP* là số lượng giao dịch thực sự là gian lận và cũng được phân loại (đúng) là gian lận. *FP* là số lượng giao dịch bình thường nhưng được phân loại (sai) là giao dịch gian lận. *TN* là số lượng giao dịch bình thường (không gian lận) và cũng được phân loại (đúng) là giao dịch bình thường. *FN* là số giao dịch gian lận và được phân loại (sai) là giao dịch bình thường. Như vậy độ chính xác ở đây chính là tỷ lệ các trường hợp phân loại đúng trên tổng số các trường hợp phân loại.

Mô hình mạng nơ ron nhân tạo đã tạo ra 114799 số dương tính thật và âm tính thật trong tổng số 114851 lượt giao dịch được dùng để phân loại. Do đó mô hình này có độ chính xác đến 99%. Mức độ quan trọng của các biến trong mô hình mạng nơ ron thần kinh được biểu diễn ở hình 5 và qua đó cũng cho ta thấy lượng tiền giao dịch là biến quan trọng nhất để dự báo và phát hiện gian lận.

Với mô hình máy vec-tơ hỗ trợ được thể hiện dưới dạng mô hình đánh giá độ chính xác Hình 6.

Trong đó độ nhạy (sensitivity) là tỷ lệ của những trường hợp được phân loại gian lận đúng với tổng số các trường hợp gian lận của mẫu nghiên cứu. Độ nhạy $= \frac{TP}{TP+FN}$. Độ nhạy 100% (hay là 1 theo đồ thị) có nghĩa là toàn bộ các giao dịch gian lận đều được phát hiện. Tuy nhiên một mình độ nhạy không cho chúng ta biết toàn bộ thông tin về mô hình bởi vì 100% độ nhạy có thể có được nếu ta gán cho toàn bộ các giao dịch đều là gian lận. Do đó chúng ta cần biết thông tin về độ đặc hiệu của mô hình.

Độ đặc hiệu (specificity) là tỷ lệ giữa những trường hợp giao dịch bình thường được phân loại đúng và tổng số những trường hợp giao dịch bình thường của mẫu nghiên cứu. Độ đặc hiệu $= \frac{TN}{TN+FP}$.

Biểu đồ ROC (receiver operating characteristic) mô tả mối liên hệ giữa độ nhạy và độ đặc hiệu, thường được dùng để đánh giá một phương pháp hay mô hình tiên lượng.

Dựa vào đồ thị hình 6: Diện tích dưới đường cong ROC (còn gọi area under the curve, AUC) là 0,891. Chỉ số độ nhạy, đặc hiệu hay AUC phản ánh độ chính xác của mô hình máy vec-tơ hỗ trợ. So sánh hai mô hình

với hai giải thuật khác nhau, dễ thấy mô hình mạng nơ ron thần kinh cho kết quả phát hiện gian lận với độ chính xác cao hơn (99%) với bộ dữ liệu nghiên cứu.

**4. Kết luận**

Việc phát hiện gian lận thẻ tín dụng luôn là công việc không hề dễ dàng đặc biệt trong đời đại bùng nổ công nghệ và việc sử dung thẻ tín dụng ngày càng phổ biến như hiện nay. Các hình thức gian lận thẻ tín dụng ngày càng tinh vi hơn, bắt buộc các mô hình để phát hiện gian lận cũng cải tiến và đem lại hiệu quả cao, nâng cao độ chính xác và giảm bớt rủi ro tổn thất tài chính do gian lận thẻ tín dụng đem lại. Bài nghiên cứu này của các tác giả một lần nữa khẳng định ứng dụng rất lớn của các giải thuật máy học trong phát hiện gian lận thẻ tín dụng với dữ liệu mô phỏng bộ dữ liệu thực đồng thời cung cấp những biến cụ thể cho mô hình nghiên cứu, cái mà những nghiên cứu trước không có nhắc đến. Kết quả cho thấy mô hình mạng nơ ron thần kinh cung cấp khả năng phát hiện gian lận đến 99%.Với kết quả này sẽ giúp các nhà quản lý lựa chọn các mô hình linh hoạt hơn nhằm nâng cao khả năng phát hiện gian lận thẻ tín dụng, góp phần giảm thiểu tổn thất tài chính và tối ưu hóa làm lợi nhuận.

**Tài liệu tham khảo**


Bolton, R.J. & Hand, D.J. (2002), 'Statistical fraud detection: A review', *Statistical Science*, 28(3), 235-255.

Edgar, A. Lopez-Rojas, Ahmad, E. & Stefan, A. (2016), 'Paysim: a financial mobile money simulator for fraud detection', *28th European Modeling and Simulation Symposium Proceedings*, CAL-TEK SRL, Italy, 249-255.

Ghosh, S. & Reilly, D.L. (1994), 'Credit card fraud detection with a neural network', *The 27th Annual Hawaii International Conference on system Sciences Proceedings,* IEEE Computer Society Press, USA, 3, 621-630.

Islam, M.J., Wu, Q.M.J., Ahmadi, M. & SidAhmed, M.A. (2007), 'Investigating the Performance of Naive-Bayes Classifiers and K Nearest Neighbor Classifiers', *IEEE International Conference on Convergence Information Technology*, IEEE, 1541-1546.

Kou, Y., Lu, C.T., Sirwongwattana, S. & Huang, Y.-P. (2004), 'Survey of fraud detection techniques', *The 2004 IEEE International Conference on Networking, Sensing and Control Proceedings*, Taipei, Taiwan, 749-754.

Leonard, K.J. (1993), 'Detecting credit card fraud using expert systems', *Computers and Industrial Engineering*, 25, 103-106.

Phua, C., Lee, V., Smith, K. & Gayler, R. (2005), 'A comprehensive survey of data mining-based fraud detection research', retrieved on April 23rd 2020, DOI: 10.1016/j.chb.2012.01.002.

Sahin, Y. & Duman, E. (2010), 'An overview of business domains where fraud can take place, and a survey of various fraud detection techniques', *The 1st International Symposium on Computing in Science and Engineering Proceedings*, Aydin, Turkey, 243-249.

Sahin, Y. & Duman, E. (2011), 'Detecting credit card fraud by ANN and logistic regression', *Innovations in Intelligent Systems and Applications (INISTA) 2011 International Symposium,* Istanbul, 315-319.

Shen, A., Tong, R. & Deng, Y. (2007), 'Application of classification models on credit card fraud detection', *Service Systems and Service Management 2007 International Conference*, IEEE, China, 1-4.

Visa Inc. (2008), *Annual Report 2008*, San Francisco.

Vũ Hữu Tiệp (2018), *Machine Learning cơ bản*, Nhà xuất bản Khoa học và Kỹ thuật, Hà Nội.