KHOA TOÁN – TIN HỌC
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

Báo cáo Seminar

Credit Card Fraud Detection and Analysis

Xác Định Giao Dịch Bất Thường và Phân Tích Đánh Giá

Người thực hiện:

Huỳnh Thị Bảo Trân

Giáo viên hướng dẫn:

TS. Trần Anh Tuấn

LỜI CẢM ƠN

Bài nghiên cứu này sẽ không thực hiện được nếu như không có sự hỗ trợ của thầy TS. Trần Anh Tuấn. Cảm ơn thầy đã hướng dẫn và hỗ trợ đề xuất phương pháp thực hiện phân tích gian lận thẻ tín dụng. Đồng thời, thầy đã đọc nhiều bản thảo và góp ý để bài nghiên cứu hoàn thiện, giải thích các phần nhầm lẫn và chỉ rõ sai sót.

PHỤ LỤC

Tóm tắt

Từ và thuật ngữ viết tắt

Danh sách các hình ảnh

Danh sách các bảng

CHƯƠNG 1: GIỚI THIỆU

1.1 Động lực

1.2 Mô tả vấn đề

1.3 Tổ chức

CHƯƠNG 2 : TỔNG QUAN NGHIÊN CỨU

2.1 Phương pháp tiếp cận gần đây

2.2 Khó khăn

2.3 Phương pháp đề xuất

2.4 Đóng góp

CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU

3.1 Thuật toán K-means

3.1 Thuật toán Naive Bayes

3.2 Quá trình thực hiện

CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU

4.1 Tổng quan cơ sở dữ liệu

4.2 Đánh giá từng bước

4.3 Đánh giá quá trình tổng thể

4.4 So sánh các phương pháp khác

CHƯƠNG 5: THẢO LUẬN

5.1 Lợi ích của phương pháp đề xuất

5.2 Hạn chế của phương pháp đề xuất

TỪ VÀ THUẬT NGỮ VIẾT TẮT

AUC	Area Under The Curve
ROC	Receiver Operating Characteristics

Danh sách các hình ảnh

Danh sách các bảng

Tóm tắt

Gian lận thẻ tín dụng là một trong những vấn đề nghiêm trọng trong lĩnh vực tài chính ngân hàng. Để kiểm tra xem một giao dịch là bất thường hay không rất khó khăn bởi trạng thái giao dịch khó xác định và thay đổi liên tục. Đồng thời dữ liệu về gian lận thẻ tín dụng rất sai lệch bởi số lượng giao dịch mỗi ngày cần được xử lý rất nhiều. Những điều này tạo nên thách thức rất lớn để nhận ra các giao dịch gian lận trong số nhiều giao dịch thanh toán thông thường. Bài nghiên cứu này tập trung vào việc xác định thuật toán máy học tốt nhất để phát hiện gian lận thẻ tín dụng, cụ thể là kết hợp thuật toán K-means và Naive Bayes. Hiệu quả của mô hình được đánh giá dựa trên độ chính xác và thời gian thực hiện giao dịch. Kết quả cho thấy phương pháp được đề xuất có độ chính xác nhất định với bộ dữ liệu được chọn nghiên cứu

Keywords – credit card fraud detection, machine learning, Naive Bayes, K-Means clustering.

CHƯƠNG 1: GIỚI THIỆU

1.1 Động lực

Cùng với sự phát triển của khoa học kỹ thuật, việc sử dụng thẻ tín dụng như một phương tiện thanh toán được sử dụng phổ biến bởi sự tiện lợi. Lợi dụng cơ hội này, những kẻ lừa đảo đã điều chỉnh các cách thức để nhằm chiếm dụng tài sản, tiền bạc của các chủ sở hữu. Những người này có thể sử dụng thẻ cho lợi ích cá nhân của họ, làm cạn kiệt tài nguyên của thẻ tín dụng hoặc đến khi họ bị bắt hay thẻ bị chặn bởi chủ sở hữu thực sự.

Theo báo cáo thứ 5 về gian lận thẻ tín dụng do Ngân hàng Trung ương Châu Âu công bố, tổng giá trị các giao dịch gian lận khu vực SEPA năm 2016 lên tới 1,8 tỷ EUR. Theo Báo cáo của Nilson, một sản phẩm về các hệ thống thanh toán toàn cầu, tổng thiệt hại do gian lận trong năm 2018 lên tới 27,85 tỷ USD và dự kiến đạt 35,67 tỷ USD vào năm 2023.

Để phát hiện các giao dịch gian lận một cách hiệu quả và chính xác cao, ta cần phải có cách phân tích và đánh giá các tập dữ liệu giao dịch. Vấn đề này thu hút sự quan tâm của cả giới học thuật và ngành công nghệ. Các cá thể đang cố gắng, tìm hiểu kiến thức để xác định các giải pháp giải quyết vấn đề theo kịp các hình thức tin vi được áp dụng bởi các kẻ gian lận.

1.2 Mô tả vấn đề

Thẻ tín dụng đóng một vai trò quan trọng trong các giao dịch hiện tại, cung cấp nhiều lợi thế cho khách hàng mà không cần thông qua tiền mặt. Chủ thẻ có thể mua hàng hóa với lượng tín dụng được trả trước hoặc trả góp cho khoản tín dụng có thể thanh toán sau khi được cho phép. Nhưng có một vấn đề là các nhà cung cấp thẻ tín dụng cũng không biết là có được sử dụng bởi chủ sở hữu của thẻ hay không. Trường hợp một người sử dụng thẻ tín dụng của người khác mà không xác thực gọi là gian lận thẻ tín dụng. Đây là một vấn đề lớn khó giải quyết hiện nay.

Gian lận thẻ tín dụng được thực hiện bằng nhiều cách như trộm cắp, ứng dụng gian lận, thẻ giả, gian lận trực tuyến… Trong các vụ lừa đảo, giao dịch được thực hiện từ xa và chỉ cần các thông tin về chủ thẻ. Nguyên nhân là do sự gia tăng các giao dịch điện tử để thuận tiện thanh toán, phục vụ cho đời sống con người. Bất kỳ ai tham gia vào việc thanh toán thẻ tín dụng đều có thể trở thành nạn nhân của các kẻ lừa đảo: chủ thẻ, giao dịch trực tuyến, nhà cung cấp các dịch vụ thanh toán, công ty xử lý thanh toán, tổ chức

phát hành thẻ. Mặc dù các cơ quan phát hành thẻ đã sử dụng các cơ chế phòng ngừa như chip hay pin để ngăn chặn và giảm bớt các hoạt động trộm cắp, nhưng số gian lận trực tuyến vẫn tăng nhanh và gây rất nhiều tổn thất đặc biệt về tài chính.

Phát hiện gian lận thẻ tín dụng là một lĩnh vực nghiên cứu tích cực và xoay quanh khái niệm tự động hóa; trên thực tế, không phải lúc nào cũng khả thi hoặc có thể xem xét thủ công từng giao dịch. Điều này đòi hỏi phải triển khai các phương thức tự động để phát hiện gian lận nhanh hơn và thông minh hơn, dẫn đến các kỹ thuật máy học ngày càng được thử nghiệm và triển khai. Do đó cần một nhóm chuyên gia phân tích và thu thập các thông tin về giao dịch thẻ tín dụng như số tiền, ngày giao dịch, địa điểm,... để có thể ngăn chặn và nhận ra bất kỳ hoạt động gian lận nào trong số giao dịch.

Chúng ta không thể phóng đại tầm quan trọng của máy học và khoa học dữ liệu. Trong các trường hợp quan tâm đến việc tìm hiểu các xu hướng trong quá khứ, việc xác định và gắn nhãn các thuộc tính hoặc giúp dự đoán các giá trị tương ứng, các kết quả đem lại lượng thông tin to lớn để ứng dụng vào các bài toán thực tế, mà ở đây là phát hiện gian lận thẻ tín dụng. Các thuật toán có thể xử lý lượng lớn dữ liệu, quan tâm đến cách hoàn thành một nhiệm vụ, đưa ra những dự đoán để đánh giá một cách thông minh. Tại sao phải sử dụng máy học? Để phát triển các hệ thống tự động tùy chỉnh cho phù hợp từng người dùng. Hệ thống có thể bắt chước con người và thay thế một số nhiệm vụ, đòi hỏi sự thông minh từ các cơ sở dữ liệu lớn.

Nhiều thuật toán phổ biến khác nhau đã được thử nghiệm để xác định một giao dịch đáng ngờ hay không, chẳng hạn như Random Forrest, ARIMA model, Hidden Markov model, Artificial Neural Network, Logistic Regression… Việc sử dụng các phương pháp máy học cung cấp độ chính xác cao hơn và trả về kết quả phù hợp khi xét đến các yếu tố khác nhau. Bởi các thuật toán có thể bao quát và học trên nhiều dữ liệu, bao gồm cả các chi tiết nhỏ nhất về các hành vi được liên kết với một tài khoản cụ thể. Hơn nữa, các thuật toán này được điều chỉnh phù hợp với môi trường và điều kiện tài chính thay đổi liên tục, cho phép các nhà phân tích xác định các giao dịch đáng ngờ mới và tạo ra các quy tắc mới để ngăn chặn lừa đảo giao dịch thẻ tín dụng.

Lưu ý rằng có một vấn đề cơ bản phổ biến trong các cách tiếp cận hiện nay, chính là tính chất không cân bằng của các tập dữ liệu. Trong bối cảnh phát hiện gian lận thẻ tín dụng thực tế, tập dữ liệu sẽ mất cân bằng, điều này cản trở đáng kể hiệu suất của các kỹ thuật học tập có giám sát và kể cả học tập không có giám sát. Một vấn đề khác có liên quan đến việc thiếu dữ liệu được gán nhãn phù hợp, điều này gây ra một số trở ngại đáng kể. Để phân tích vấn đề này, bài nghiên cứu đề xuất các cách tính toán tập dữ liệu trong giao dịch để tìm một cách hiệu quả và chính xác cao. Từ đó đưa ra các phương hướng để cải thiện số lượng giao dịch gian lận thẻ tín dụng.

1.3 Tổ chức

Bài nghiên cứu được trình bày như sau:

- Tổng quan nghiên cứu, trình bày và phân tích khó khăn của các phương pháp được sử dụng gần đây, đồng thời đề xuất phương pháp phù hợp hơn được đưa ra ở chương 2.
- Ở chương 3, trình bày phương pháp nghiên cứu chính, thuật toán và cách thức để thực hiện phát hiện giao dịch thẻ tín dụng.
- Chương 4 thảo luận về phát hiện và kết quả trong quá trình nghiên cứu, so sánh với các phương pháp khác.
- Thảo luận về ưu và nhược điểm của phương pháp được đề xuất, đánh giá định hướng phát triển trong tương lai được nhận định trong chương 5.

CHƯƠNG 2: TỔNG QUAN NGHIÊN CỨU

2.1 Phương pháp tiếp cận gần đây

Sự tiến bộ của công nghệ thông tin đã tạo ra một khối lượng lớn cơ sở dữ liệu và thông tin khổng lồ trong các lĩnh vực khác nhau. Các tính toán số học, mô hình thống kê, trí tuệ nhân tạo và các thuật toán máy học đã tìm các mẫu dữ liệu được khai thác sử dụng để tạo dự đoán. Kết quả có thể xảy ra, tạo ra các thông tin và có thể thực hiện được. Điều này có nghĩa rằng lượng lớn các cơ sở dữ liệu ấy nếu như được khai thác triệt để thì kết quả mang lại cực kỳ hữu ích. Việc nghiên cứu các phương pháp phân tích và khai thác thông tin, điều quan trọng là tìm hiểu dữ liệu cơ bản và được mô hình hóa bằng cách chọn một cơ chế thích hợp với xử lý bất kỳ trường hợp nào. Các tham số của thuật toán cần được điều chỉnh sao cho phù hợp với tập dữ liệu. Các ứng dụng được phát triển để tìm thấy và nâng cao việc sử dụng và hiệu quả cao trong giải quyết vấn đề.

Trong những năm gần đây, việc sử dụng thẻ tín dụng như một phương tiện thanh toán đã gia tăng mạnh mẽ do tính tiện lợi và dễ sử dụng của nó. Một giao dịch được cho là gian lận khi nó được thực hiện bởi một bên không được ủy quyền hay chủ sở hữu hợp pháp. Từ đó, những kẻ lừa đảo đã lợi dụng sự tiện ích này mà thực hiện các hành động độc hại để chiếm đoạt các khoản tiền, tài sản của chủ sở hữu thẻ tín dụng. Hơn hết, các hành động này hết sức tinh vi mà bất cứ các hệ thống phát hiện truyền thống nào cũng không thể cho kết quả tốt nhất. Điều này đòi hỏi phải triển khai các công cụ tự động để phát hiện gian lận nhanh hơn và thông minh hơn. Nó có thể hiểu rằng chúng ta cần nghiên cứu các phương pháp về học máy để xác định các giải pháp phù hợp và theo kịp những sự tinh vi của các kẻ gian lận.

Có rất nhiều nhà nghiên cứu đã nỗ lực để thực hiện các chứng minh rằng mô hình hóa tập dữ liệu bằng phương pháp phân loại (classification), phân cụm (clustering), học sâu (deep learning)…. mà đồng thời phát hiện gian lận thẻ tín dụng là cơ sở cơ bản. Classification là một trong những mô hình máy học bao gồm các chức năng dẫn xuất để phân tích dữ liệu thành các danh mục hoặc dưới dạng các lớp, đặc trưng bằng tập huấn luyện dữ liệu chứa các quan sát trong các trường hợp. Clustering là các thuật toán chia tập hợp thành các nhóm khác nhau sao cho mỗi điểm dữ liệu tương tự với các điểm dữ liệu trong cùng một nhóm và khác với các điểm dữ liệu trong các nhóm khác. Trên cơ sở sự giống nhau và không giống nhau, sau đó phân nhóm phụ thích hợp cho đối tượng. Deep learning là một phần trong một nhánh rộng hơn các phương pháp học máy dựa trên mạng thần kinh nhân tạo kết hợp với việc học biểu diễn đặc trưng (representation learning). Cụ thể về các phương pháp sẽ được trình bày ở ngay sau đây.

Trước hết, ở bài nghiên cứu của Sahin đã trình bày và đề xuất mô hình mạng nơron nhân tạo (Artificial Network model - Artificial Network model). Artificial Network model là một mô hình tính toán được mô phỏng dựa trên hoạt động của mạng nơron sinh học. Nó bao gồm số lượng lớn các nơron đơn lẻ gắn kết với nhau, xử lý thông tin bằng cách truyền cáp kết nối và tính các giá trị mới tại cái nút. Có 3 layers thuộc Artificial Network model là: input layer, hidden layer và output layer. Input layer biểu diễn thông tin đầu vào. Hidden layer gồm các nút nhận ma trận đầu vào từ layer trước, kết hợp với trọng số cùng với hàm kích hoạt phi tuyến như sigmoid, tanh để có được kết quả ở output layer. Artificial Network model bao gồm 2 quá trình tính toán cơ bản là lan truyền tiến (feedforward) và lan truyền ngược (backpropagation).

Quá trình suy luận từ input layer tới output layer là feedforward, tức là quá trình này chỉ có chiều hướng các nơron ở cùng một tầng lấy thông tin từ tầng trước mà không có chiều ngược lại:

$$a^{(0)} = x$$

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)},\ l = \overline{1,\ L}$$

$$a^{(l)} = f^{(l)}(z^{(l)})$$

$$\widehat{y} = a^{(L)}$$

Ở đây x là đầu vào, $W^{(l)}$ trọng số tương ứng ở tầng thứ l, $b^{(l)}$ là 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 $f(z) = \frac{1}{1+e^{-z}}$ với đồ thị cân xứng thể hiện mức độ công bằng đối với các tham số. $\hat{y}$ chính là đầu ra của dự đoán.

Back propagation là phương pháp để tính đạo hàm của hàm mất mát từ layer cuối cùng đến layer đầu tiên. Layer cuối cùng được tính toán trước vì nó ảnh hưởng trực tiếp đến đầu ra. Hàm mất mát đạt giá trị nhỏ nhất 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(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 layer đầu ra L:

$$\frac{\partial J}{\partial w_{ij}^{(L)}} = \frac{\partial J}{\partial z_j^{(L)}} \frac{\partial z_j^{(L)}}{\partial w_{ij}^{(L)}}$$

Với $\frac{\partial z_j^{(L)}}{\partial w_{ij}^{(L)}} = a_i^{(L-1)}$ bởi vì $z_j^{(L)} = w_j^{(L)T} a^{(L-1)} + b_j^{(L)}$ 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 layer $l < L$, bằng quy nạp ngược từ cuối:

$$\frac{\partial J}{\partial w_{ij}^{(l)}} = \frac{\partial J}{\partial z_j^{(l)}} \frac{\partial z_j^{(l)}}{\partial w_{ij}^{(l)}} = \frac{\partial J}{\partial z_j^{(l)}} a^{(l-1)}$$

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

Thứ hai, ở bài nghiên cứu đã sử dụng thuật toán Support Vector Machine. Đây 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 Support Vector Machine là đi tìm một siêu phẳng phân tách dữ liệu tốt nhất. Gọi là 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 Support Vector Machine chính là đi tìm siêu phẳng phân chia có lề lớn nhất.

Bài toán tối ưu của Support Vector Machine là đi tìm w, b sao cho lề đạt được giá trị lớn nhất.

$$(w, b) = argmax \left\{ \frac{1}{\|w\|_2} miny_n \left(w^T x_n + b \right) \right\}$$

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:

$$y_m \left(w^T x_m + b \right) = 1$$

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

$$(w,\ b) = argmax \frac{1}{\|w\|_2}$$

$$y_m \left(w^T x_m + b \right) \geq 1,\ \forall n = \overline{1,\ N}$$

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 λ thỏa mãn:

$$\lambda = argmax \sum_{n=1}^{N} \lambda_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \lambda_n \lambda_m y_n y_m x_n^T x_m$$

$$\lambda \geq 0$$

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

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

$$w = \sum_{m \in S} \lambda_m y_m x_m$$

$$b = y_m - w^T x_m$$

Ở đây (x_m, y_m) là điểm dữ liệu bất kỳ nào đó trên đường biên gốc, ta còn gọi là Support Vector. 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:

$$b = \frac{1}{N_S} \sum_{n \in S} \left(y_n - W^T x_n \right) = \frac{1}{N_S} \sum_{n \in S} (y_n - \sum_{m \in S} \lambda_m y_m x_m^T x_n)$$

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

$$\sum_{m \in S} \lambda_m y_m x_m^T x + b$$

Hơn nữa, ở một bài nghiên cứu khác, ta thấy được rằng có một số tác giả đã sử dụng phương pháp mô hình Markov ẩn - Hidden Markov model. Hidden Markov model là một tập hợp hữu hạn các trạng thái; mỗi trạng thái được liên kết với một phân phối xác suất. Chuyển đổi giữa các trạng thái này được điều chỉnh bởi xác suất thiết lập được gọi là xác suất chuyển đổi. Trong một trạng thái cụ thể, đây là biểu tượng liên quan đến các quan sát phân phối xác suất. Nó chỉ là kết quả, không phải là trạng thái mà người quan sát bên ngoài có thể nhìn thấy và do đó các trạng thái là "ẩn" bên ngoài; do đó được gọi là Hidden Markov model.

Do đó, Hidden Markov model là một giải pháp hoàn hảo để giải quyết vấn đề phát hiện giao dịch gian lận thông qua thẻ tín dụng. Một lợi ích to lớn của phương pháp tiếp cận dựa trên Hidden Markov model là giảm đáng kể số lượng giao dịch False Positives được hệ thống phát hiện giao dịch thẻ tín dụng đó có là độc hại hay không, mặc dù chúng thực sự có thật. Trong quá trình dự đoán này, Hidden Markov model chủ yếu xem xét ba

phạm vi giá trị như: Low (l), Medium (m) và High (h). Đầu tiên, mô hình sẽ được yêu cầu tìm ra số tiền giao dịch thuộc về một danh mục cụ thể hoặc nó sẽ ở một phạm vi thấp, trung bình hoặc cao.

Thêm một phương pháp nữa được sử dụng ở trong bài nghiên cứu của Khandare là ARIMA model. Hai mô hình được sử dụng rộng rãi cho chuỗi thời gian là mô hình tự hồi quy (autoregressive - AR) và trung bình trượt (moving average - MA), có thể được sử dụng cùng nhau dưới dạng mô hình trung bình trượt tự hồi quy (autoregressive moving average - ARMA). ARMA(p, q) là sự kết hợp của các mô hình AR(p) và MA(q) và có thể được sử dụng với chuỗi thời gian đơn biến.

Mô hình AR(p) được xác định theo phương trình dưới đây; nó giả định rằng có một mối quan hệ tuyến tính phụ thuộc giữa quan sát và các giá trị của một số quan sát lagged (previous) cụ thể cộng với error term.

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \omega_t$$

trong đó $\phi = (\phi_1, \phi_2, ..., \phi_n)$ là các hệ số của mô hình, p là số nguyên không âm, c là hằng số và $\omega_t \sim N(0, \sigma^2)$.

Mô hình MA(p) được xác định bởi phương trình dưới đây; nó sử dụng sự phụ thuộc giữa một quan sát và các lỗi còn lại do việc áp dụng mô hình trung bình động cho các quan sát lagged.

$$X_t = \mu + \sum_{j=1}^{q} \theta_j \omega_{t-j} + \omega_t$$

trong đó μ là giá trị trung bình của chuỗi, $\theta = (\theta_1, \theta_2, ..., \theta_n)$ là các hệ số của mô hình và q là bậc và $\omega_t \sim N(0, \sigma^2)$.

Mô hình ARMA, là kết quả của sự kết hợp của hai mô hình này, được định nghĩa như sau:

$$X_t = c + \omega_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j X_{t-j}$$

trong đó p là thứ tự của mô hình AR và q là thứ tự của mô hình MA. Giả định chính trong phân tích chuỗi thời gian là chuỗi thời gian không đổi, nghĩa là giá trị trung bình và phương sai của nó không đổi theo thời gian; tuy nhiên, đây không phải là trường hợp trong nhiều tình huống thực tế. Giải pháp cho vấn đề này có thể được tìm thấy trong sự

khái quát hóa của Mô hình ARMA: mô hình trung bình động tích hợp tự hồi quy (ARIMA). ARIMA giới thiệu khả năng áp dụng sự khác biệt cho các điểm dữ liệu của chuỗi thời gian để làm cho nó đứng yên.

ARIMA hiện là một trong những mô hình phổ biến, linh hoạt và đơn giản nhất để phù hợp với chuỗi thời gian. Trong bối cảnh phát hiện gian lận, chuỗi thời gian có thể được sử dụng như một công cụ khi làm việc với các thuộc tính. Thường được sử dụng để lấy các tính năng mới từ các tính năng ban đầu để cung cấp cho mô hình một số thông tin phù hợp hơn. Số lượng giao dịch hàng ngày hoặc tổng số tiền chi tiêu trong một tuần là những ví dụ về thuộc tính.

2.2 Khó khăn

Sự gia tăng tổn thất do sự tăng nhanh chóng của các giao dịch điện tử cũng như các phương pháp gian lận ngày càng tinh vi hơn khi phát hiện các sự gian lận thẻ tín dụng truyền thống. Hơn nữa, tất cả người tham gia vào quy trình thanh toán thẻ tín dụng đều có khả năng trở thành nạn nhân của những kẻ lừa đảo, như là chủ sở hữu thẻ, giao dịch trực tuyến, nhà cung cấp cổng thanh toán, công ty phát hành thẻ… Do đó, ta cần phải xác định các giao dịch có là đáng ngờ hay không và báo cáo kết quả đó cho nhà phân tích trong các giao dịch thông thường được xử lý tự động. Hiện tại đang chuyển sang nghiên cứu các phương pháp máy học để cải thiện vấn đề.

Ở phần **2.1 Phương pháp tiếp cận gần đây** có đề cập đến các phương pháp đang được sử dụng để phát hiện gian lận thẻ tín dụng phổ biến hiện nay. Mỗi phương pháp, có những sự ưu và nhược điểm riêng và phù hợp với từng tập dữ liệu nhất định. Nhưng nhìn chung, các phương pháp đều có thể giải quyết được bài toán phát hiện gian lận thẻ tín dụng, chỉ là độ chính xác của các phương pháp có sự chênh lệch. Các thuật toán máy học có độ chính xác cao hơn và trả về kết quả phù hợp khi xét về các yếu tố bổ sung. Do các thuật toán máy học có thể xem trên nhiều tập dữ liệu hơn, bao gồm các chi tiết nhỏ nhất về các dữ liệu hành vi được liên kết với một tài khoản cụ thể.

Với thuật toán Artificial Network model, ta có thể thấy vấn đề ở đây là phát hiện dựa trên quy tắc kiểm tra các danh mục tài khoản không cung cấp đủ chứng thực với việc sử dụng thẻ tín dụng ngày càng tăng. Mặc dù phương pháp này phân loại mạnh mẽ nhưng chưa đủ chính xác. Dẫn đến hạn chế lớn đó là mặc dù sử dụng tập dữ liệu số thực nhưng không phải tất cả dữ liệu đều được chọn lọc. Ở bài nghiên cứu của Sahin đã giữ tỷ lệ của giao dịch gian lận so với các giao dịch bình thường. Đồng thời thời gian giao dịch được thu thập bị bỏ qua, không được phân tích.

Với thuật toán Hidden Markov model, vấn đề to lớn là lừa đảo trong giao dịch trực tuyến, cần giải quyết các điểm giá trị ngoại lai (outlier) của dữ liệu trong tiền xử lý dữ liệu (pre-processing), số lượng phát hiện bất thường cũng tăng. Với mục tiêu là triển khai hệ thống phát hiện gian lận thẻ tín dụng với xác suất phát hiện cao. Điều hạn chế của Hidden Markov model là không thể hiện được sự phụ thuộc giữa các thuộc tính với nhau. Do đó cần phát triển quy mô dữ liệu, tức là xử lý trên khối lượng dữ liệu lớn.

Thuật toán Support Vector Machine dẫn đến bài toán tối ưu cấp 2 toàn cục với các ràng buộc lồi (interior point methods), ánh xạ cung cấp một khung xử lý cho hầu hết các kiến trúc mô hình. Do đó có thể sử dụng Euclide có trọng số giảm chiều dữ liệu trong tập dữ liệu lớn. Điều này giúp cải tiến ranh giới của các mẫu dữ liệu rõ ràng, rút gọn thời gian lập mô hình và cải thiện độ chính xác của phân loại.

Với phương pháp ARIMA model, vấn đề của phương pháp này là bài toán học không giám sát (unsupervised learning) trong việc phát hiện gian lận thẻ tín dụng. Lý do chính là do tập trung vào mô hình chuỗi thời gian do thiếu dữ liệu hành vi gian lận, bởi các vấn đề bí mật, đây cũng là trở ngại lớn trong quá trình phát triển của máy học. Một vấn đề chính trong cách tiếp cận là mô hình ARIMA giả định rằng dữ liệu đến từ các quan sát cách đều nhau về thời gian. Tuy nhiên, giả định này không đúng trong nghiên cứu vì thời gian giao dịch cách nhau không đều. ARIMA model hoạt động tốt hơn khi có một số lượng đáng kể các vụ gian lận xảy ra trong cùng một ngày.

Ta có thể tóm tắt bằng bảng thông tin sau:

Phương pháp	**Khó khăn**
Artificial Network model	Thực hiện dựa trên quy tắc kiểm tra các danh mục đầu tư không cung cấp đủ chứng thực với việc sử dụng thẻ tín dụng ngày càng tăng.
Hidden Markov model	Cần giải quyết các điểm outlier trong tiền xử lý dữ liệu. Số lượng phát hiện bất thường tăng.
Support Vector Machine	Việc phân lớp chỉ là việc cố gắng tách các đối tượng vào hai lớp được phân tách bởi siêu phẳng Support Vector Machine.
ARIMA model	Giả định rằng dữ liệu đến từ các quan sát cách đều nhau về thời gian, giả định này không đúng trong nghiên cứu vì thời gian giao dịch cách nhau không đều.

Bảng 2.1: Sự khó khăn của các phương pháp trong Bài toán phát hiện gian lận

2.3 Phương pháp đề xuất

Cùng với sự phát triển vượt bậc của Internet và thương mại điện tử, việc sử dụng thẻ tín dụng là điều tất yếu. Do việc sử dụng thẻ tín dụng ngày càng tăng, các vụ lừa đảo liên quan đến việc này cũng tăng. Dẫn đến việc có nhiều phương pháp được sử dụng để phát hiện gian lận. Nếu có bất kỳ sự khác biệt nào xảy ra trong mô hình giao dịch gian lận, các hành vi đó được dự đoán là đáng ngờ và được xem xét thêm để tìm ra các hành vi gian lận. Nói chung, vấn đề phát hiện gian lận thẻ tín dụng có một lượng lớn dữ liệu, được khắc phục bằng phương pháp được đề xuất. Đạt được độ chính xác cao nhất, tỷ lệ nhận biết gian lận cao là nhiệm vụ chính của phương pháp này**.**

Mục tiêu chính là xây dựng một hệ thống cho các biểu hiện của hành vi lừa đối tội phạm. Nhận dạng gian lận là một vấn đề phức tạp đòi hỏi một lượng kỹ năng đáng kể cho đến khi đưa các thuật toán liên quan đến máy học vào đó. Đây gần như là triển khai kết hợp các thuật toán trí tuệ nhân tạo hoặc cụ thể hơn là các thuật toán máy học đảm bảo rằng tiền của khách hàng được an toàn và không bị thao túng bởi các hành động sai trái. Đề cập đến hệ thống xác định gian lận hiệu quả tùy thuộc vào các phương pháp máy học. Quá trình phản hồi liên quan đến việc nâng cao tỷ lệ phát hiện cũng như hiệu quả của bộ phân loại.

Dữ liệu gian lận thẻ tín dụng rất mất cân bằng (imbalanced) và nó đã tạo ra một phần lớn các giao dịch không gian lận và một phần nhỏ các giao dịch gian lận. Các biện pháp được sử dụng để đánh giá tính xác thực của các thuật toán phát hiện trở nên quan trọng đối với việc triển khai một mô hình chấm điểm chính xác các giao dịch gian lận có tính đến sự mất cân bằng trường hợp và chi phí xác định trường hợp là thật khi trên thực tế, trường hợp đó là giao dịch gian lận.

Các phương pháp cổ điển đã không thể phát hiện ra các gian lận thông qua các giao dịch bởi sự phức tạp của bộ dữ liệu hay nói cụ thể hơn là do sự tinh vi của những kẻ lừa đảo nhằm chiếm đoạt tài sản hay tiền bạc của chủ sở hữu thẻ tín dụng. Để giải quyết nhược điểm này, bài nghiên cứu này đề xuất hệ thống được triển khai bằng thuật toán phân cụm K-means kết hợp với thuật toán phân loại Naive Bayes để phát hiện gian lận thẻ tín dụng. Kết quả cho thấy phương pháp được đề xuất có độ chính xác nhất định với bộ dữ liệu được chọn nghiên cứu.

2.4 Đóng góp

Chủ sở hữu thẻ tín dụng có thể mua hàng với số lượng tín dụng được trả trước hoặc góp cho khoản tín dụng có thể thanh toán sau được biểu thị bằng tỷ lệ đồng ý trước đó.

Nhưng các nhà cung cấp thẻ cũng không biết được thẻ có được sử dụng bởi chủ sở hữu của nó hay không. Một người sử dụng thẻ tín dụng của người khác mà không xác thực được gọi là gian lận thẻ tín dụng. Do đó, một khoản tiền hay tài sản khổng lồ sẽ không thể thu hồi được xảy ra ở ngân hàng. Người mua hàng trực tuyến cũng bị ảnh hưởng bởi gian lận thẻ tín dụng vì họ phải thanh toán qua thẻ. Để phát hiện các giao dịch gian lận thẻ tín dụng, nghiên cứu này đề xuất phương pháp kết hợp thuật toán phân cụm K-means với thuật toán phân loại Naive Bayes.

Phương pháp này được lựa chọn vì kết quả phân cụm cuối cùng của K-means phụ thuộc nhiều vào việc lựa chọn các trọng tâm ban đầu, do đó cần phải là một phương pháp có hệ thống phân loại để xác định các trọng tâm ban đầu giúp cho K-means hội tụ. Naïve Bayes là một trong những thuật toán phân loại máy học đơn giản và phổ biến nhất. Trong thuật toán này, kỹ thuật phân loại dựa trên Định lý Bayes với giả định về tính độc lập giữa các yếu tố dự đoán. Ngoài việc giải các bài toán có kết quả không duy nhất, thuật toán còn được áp dụng rộng rãi cho các dạng bài toán khác nhau. Phương pháp này cố gắng nâng cao thuật toán phân cụm K-means và phân loại Naive Bayes bằng cách loại bỏ một trong những nhược điểm của nó.

Thuật toán K-Means rất dễ thực hiện và đơn giản trong thực thi, khả năng mở rộng, tốc độ hội tụ và khả năng thích ứng với dữ liệu khuyết (missing value). Đây là một phương pháp tốt để xác định tính hợp lệ của giao dịch bằng thẻ tín dụng. Thuật toán được sử dụng làm tiêu chí hợp lệ cho một bộ dữ liệu số nhất định. Hầu như tất cả các số thẻ tín dụng đều được trích xuất theo tiêu chí hợp lệ này… còn được gọi là kiểm tra Mod 10. Đây là phương pháp cũng được sử dụng để xác minh một số thẻ hiện có nhất định trong tài chính ngân hàng.

Thuật toán Naive Bayes 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. Xem xét và phân tích chi phí tài chính của việc gian lận được thu thập trong dữ liệu, chứng minh là chính xác hơn để phát hiện gian lận thẻ tín dụng so với các phương pháp truyền thống. Kết quả cho thấy chiến lược này cải thiện đáng kể hiệu suất phát hiện gian lận thẻ tín dụng. Kết quả cho thấy chiến lược này cải thiện đáng kể hiệu suất phát hiện gian lận thẻ tín dụng.

CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU

3.1 Thuật toán K-means

K-Means là thuật toán học không giám sát (unsupervised learning) được sử dụng trong bài toán phân cụm (clustering), không biết nhãn (label) của từng điểm dữ liệu. Mục đích là tìm cách để phân dữ liệu thành các cụm cluster khác nhau sao cho dữ liệu trong cùng một cụm có tính chất giống nhau. Thuật toán K-means hoạt động với đầu vào được chia thành hai phần: một tập dữ liệu $X = x_1, x_2, x_3, ..., x_n$ và một số lượng cụm cần tìm K. Đầu ra là các trọng tâm J và label vector cho từng điểm dữ liệu. Thuật toán đặt trọng tâm $c_1, c_2, c_3, ..., c_n$ cho mỗi cụm J tại các vị trí ngẫu nhiên, sau đó thực hiện theo các bước trình bày dưới đây:

Bước 1: Với mỗi điểm x_n:

- Tìm trọng tâm c_j gần nhất. K-means tính toán khoảng cách Euclide giữa mỗi điểm x_n và trọng tâm c_j. Cách tiếp cận này thường được gọi là giảm thiểu quán tính của các cụm và có thể được định nghĩa như sau:

$$SS_{w_i} = \sum_n \left|\left|x_n - c_i\right|\right|^2 \ \forall i \in (1,\ K)$$

trong đó n là số điểm x và i là số trọng tâm c.

- Gán thể hiện x_n cho cụm J.

Bước 2: Với mỗi cụm J: 1, 2, ...K

- Tính trọng tâm mới c_j. Điều này đạt được bằng cách tính giá trị trung bình từ mỗi điểm x đến trọng tâm x của cụm J được gán đầu tiên.

Bước 3: Dừng lại khi đạt đến sự hội tụ; nghĩa là không còn thay đổi trong các lần lặp.

Chỉnh sửa K-means học tập với toàn bộ tập dữ liệu chỉ định hai cụm (đối với số lượng hàng ngày hợp pháp và gian lận). Cụm chứa số lượng thể hiện nhỏ nhất được coi là cụm chỉ ra lớp chính xác, chỉ một phần của các giá trị ngoại lệ trong bộ thử nghiệm được tính đến.

3.2 Thuật toán Naive Bayes

Naive Bayes là một trong những thuật toán phân loại máy học đơn giản và phổ biến nhất. Trong thuật toán này, kỹ thuật phân loại dựa trên Định lý Bayes với giả định về tính độc lập giữa các yếu tố dự đoán. Nói một cách đơn giản, mô hình phân loại này giả định rằng sự hiện diện của một đặc điểm chung trong một lớp không liên quan đến sự hiện diện của bất kỳ đặc điểm nào khác. Nó rất dễ xây dựng và đặc biệt hữu ích cho các tập dữ liệu lớn. Do đó, nó được biết vượt trội hơn cả các mô hình phân loại phức tạp cao và có thể triển khai dễ dàng. Naive Bayes được mô hình hóa dựa trên Định lý Bayes trong xác suất thống kê:

Sử dụng mô hình xác suất:

$$P(Y|X) = \frac{P(X|Y)*P(Y)}{P(X)}$$

Với P(Y|X) là posterior probability: xác suất của mục tiêu y với điều kiện có đặc trưng x, P(X|Y) là likelihood: xác suất của đặc trưng X khi có điều kiện y, P(Y) gọi là prior probability của mục tiêu y.

Ở đây, X là vector đặc trưng, có thể viết dưới dạng: $X = (x_1, x_2, x_3, ..., x_n)$

Trong mô hình Naive Bayes, có hai giả thiết được đặt ra:

- Các đặc trưng đưa vào mô hình là độc lập với nhau. Tức là sự thay đổi giá trị của một đặc trưng không ảnh hưởng các đặc trưng còn lại.
- Các đặc trưng đưa vào mô hình có ảnh hưởng ngang nhau đối với đầu ra của mục tiêu.

Khi đó, kết quả của y để P(Y|X) đạt cực đại trở thành:

$$Y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|Y)$$

Bởi sự đơn giản của hai giả thiết với việc dự đoán rất nhanh kết quả đầu ra khiến nó được sử dụng rất nhiều trong thực tế trên những bộ dữ liệu lớn, đem lại kết quả khả quan.

3.3 Quá trình thực hiện

Các bước thực hiện xử lý phát hiện giao dịch thẻ tín dụng gian lận:

- Bước 1: Đọc và trích xuất các thông tin cơ bản về tập dữ liệu.
- Bước 2: Phân tích và trực quan hóa dữ liệu (được thực hiện bằng Microsoft Power BI).
- Bước 3: Khai thác dữ liệu cơ bản và cân bằng dữ liệu.

- Bước 4: Chia tập dữ liệu thành hai phần, tập huấn luyện (train) và tập kiểm tra (test).
- Bước 5: Thực hiện training dữ liệu với mô hình đã thiết lập (K-means và Naive Bayes).
- Bước 6: Tính độ chính xác của phương pháp đề xuất với các phương pháp khác.
- Bước 7: Đánh giá mô hình tốt nhất phù hợp với tập dữ liệu

CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU

4.1 Tổng quan cơ sở dữ liệu

Có một sự thực rằng rất khó để thu thập được dữ liệu thực mô tả hoạt động giao dịch thẻ tín dụng trực tuyến vì tính bảo mật về thông tin của ngân hàng. Dữ liệu được sử dụng là Paysim, bộ dữ liệu của Edgar và cộng sự (2016) được tạo bằng giả lập. Dữ liệu này 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. 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.

Các thuộc tính trong dữ liệu:

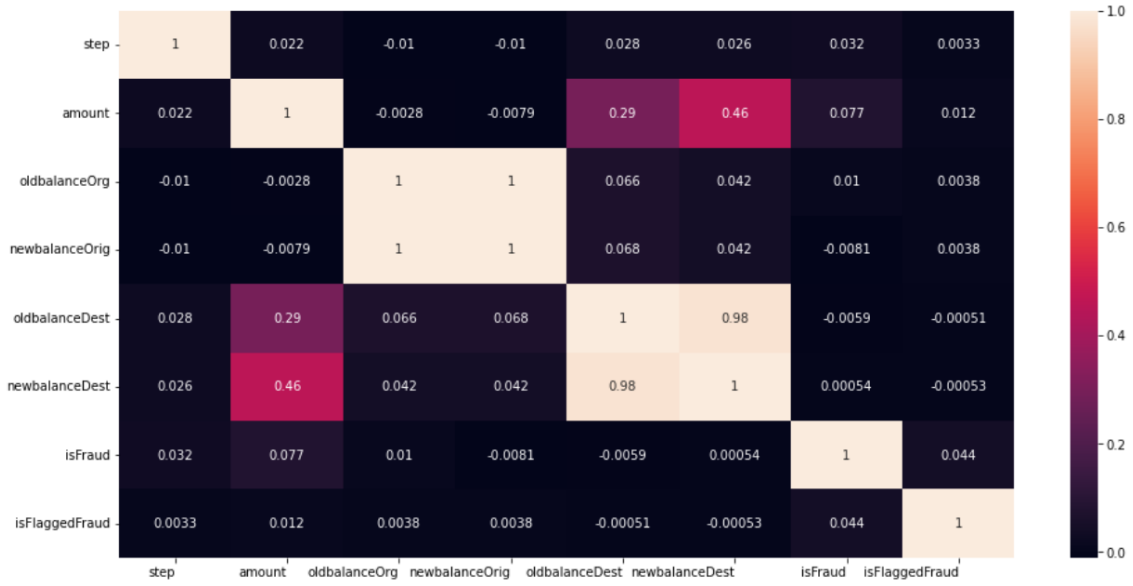
- step: Đại diện cho một đơn vị thời gian trong đó 1 bước bằng 1 giờ.
- type: Các loại hình giao dịch: PAYMENT (thanh toán), TRANSFER (chuyển khoản), CASH_OUT (tiền ra), CASH_IN (tiền vào), DEBIT (ghi nợ).
- amount: Số tiền của mỗi lần giao dịch.
- nameOrig: Khách hàng thực hiện giao dịch.
- oldbalanceOrg: Số dư tài khoản gốc của người gửi trước giao dịch.
- newbalanceOrig: Số dư tài khoản của người gửi sau giao dịch.
- nameDest: Người nhận giao dịch.
- oldbalanceDest: Số dư ban đầu của người nhận trước khi giao dịch.
- newbalanceDest: Số dư mới của người nhận sau giao dịch.
- isFraud: Xác định một giao dịch có là gian lận hay không (0 là bình thường, 1 là giao dịch gian lận).
- isFlaggedFraudflags: Chuyển bất hợp pháp hơn 200,000 trong một giao dịch.

4.2 Đánh giá từng bước

- Bước 1: Đọc và trích xuất các thông tin cơ bản về tập dữ liệu.

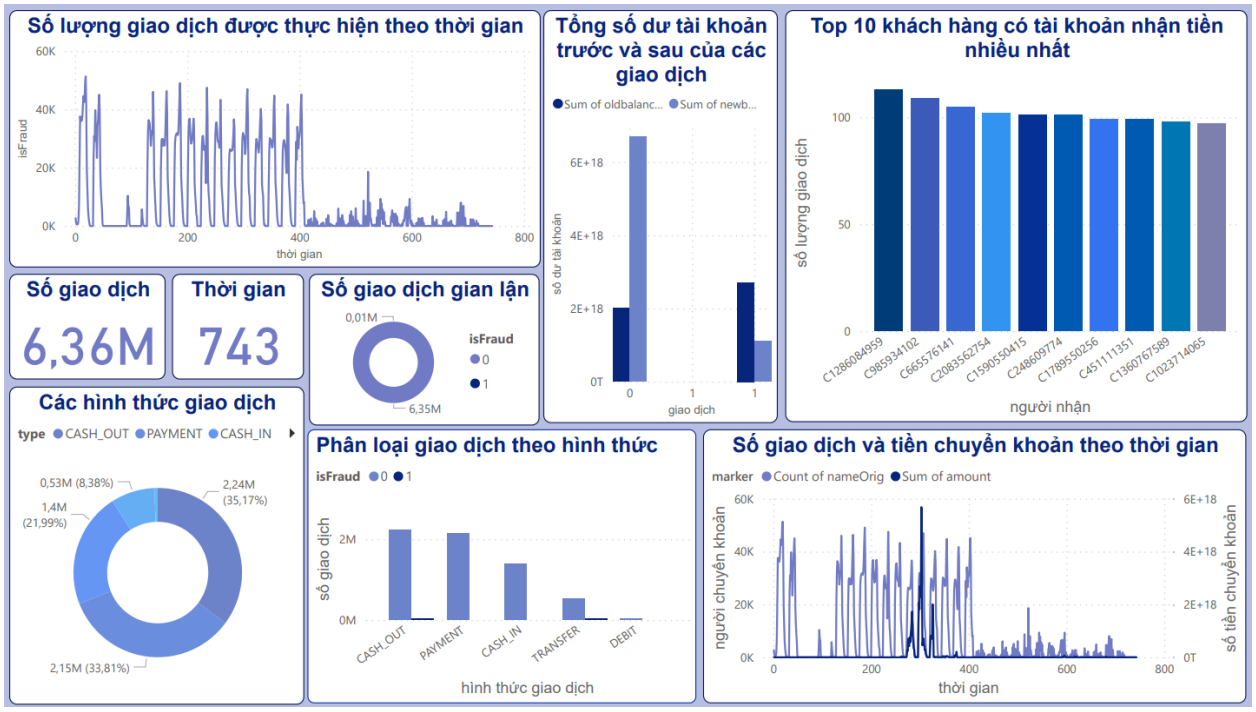
Có 6362620 giao dịch thanh toán bằng thẻ ngân hàng và 11 thuộc tính mô tả thông tin của những lần giao dịch. Tập dữ liệu không có các giá trị rỗng hay lặp lại.

Biểu đồ dưới đây thể hiện giá trị tương quan giữa các thuộc tính, cho thấy các thuộc tính có mối quan hệ tuyến tính giữa các thuộc tính.



Hình 4.1 Biểu đồ thể hiện mức độ tương quan giữa các thuộc tính của dữ liệu

- Bước 2: Phân tích và trực quan hóa dữ liệu (thực hiện bằng Microsoft Power BI).



Hình 4.2 Dashboard phân tích dữ liệu bằng Microsoft Power BI

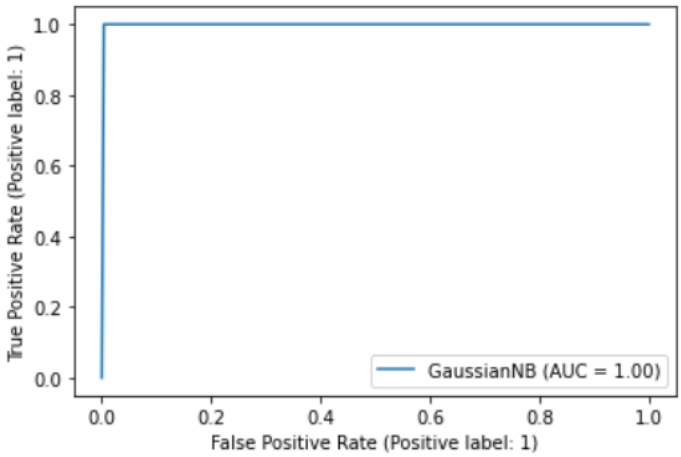
- Bước 3: Khai thác dữ liệu cơ bản và cân bằng dữ liệu.
- Bước 4: Chia tập dữ liệu thành hai phần, tập huấn luyện (train) và tập kiểm tra (test).

Dữ liệu huấn luyện: 80% tập dữ liệu.

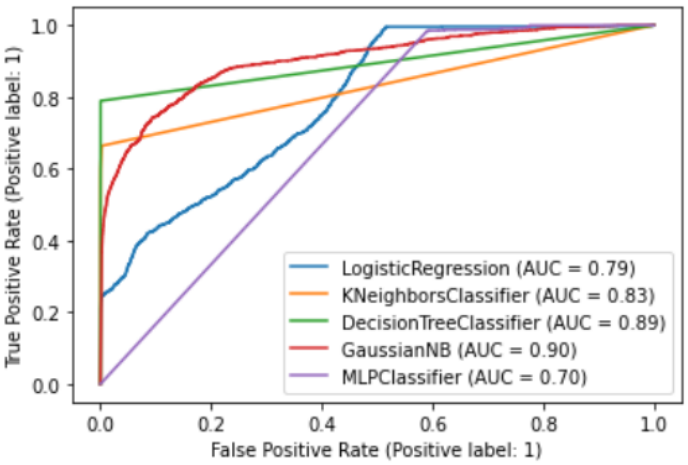
Dữ liệu kiểm tra: phần còn lại của tập dữ liệu.

- Bước 5: Thực hiện training dữ liệu với mô hình đã thiết lập (K-means và Naive Bayes).
- Bước 6: Tính độ chính xác của phương pháp đề xuất với các phương pháp khác.

Đánh giá độ chính xác dựa trên chỉ số AUC và Confusion Matrix bởi vì dữ liệu bị mất cân bằng (sự chênh lệch lớn giữa các giao dịch thông thường với giao dịch gian lận).



Hình 4.3: Biểu đồ ROC thể hiện kết quả AUC của phương pháp đề xuất



Hình 4.3: Biểu đồ ROC thể hiện kết quả AUC của các phương pháp khác

- Bước 7: Đánh giá mô hình tốt nhất phù hợp với tập dữ liệu.

Thông qua các biểu đồ ROC và Confusion Matrix, ta có thể thấy rằng phương pháp đề xuất cho kết quả tương đối tốt so với các phương pháp khác.

4.3 Đánh giá quá trình tổng thể

Trong tổng số 6362620 lượt giao dịch thì giao dịch gian lận có 8213 lượt (chiếm 0,13%). Trong số các giao dịch gian lận, gian lận chỉ xảy ra trong loại CASHOUT (rút tiền) và TRANSFER (chuyển khoản), và có 0,183% xảy ra gian lận trong loại rút tiền và 0,769% gian lận xảy ra chuyển tiền.

Chỉ có 16 hồ sơ trong tổng số hồ sơ bị gắn cờ là gian lận, tỷ lệ rất ít 0,195% tổng số giao dịch. Có tới 8197 giao dịch được gắn cờ không chính xác. Tất cả 16 giao dịch được gắn cờ gian lận thì thực sự là giao dịch gian lận. Việc gắn cờ không chính xác có thể có tác động lớn trong tương lai nếu không tính toán đúng cách, vì điều này có thể dẫn đến tỷ lệ gian lận thanh toán trực tuyến gia tăng khi mọi người ngày càng chuyển sang thanh toán trực tuyến nhiều hơn.

Số tiền giao dịch gian lận dao động trong khoảng 340.000 - 360.000, đây là một khoản tiền lớn.

Trong các giao dịch được xem là gian lận, số dư của tài khoản gốc cao hơn so với số dư ở tài khoản gốc không bị gian lận. 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.

Giao dịch gian lận ít có khả năng hoặc hiếm khi xảy ra trong quá trình chuyển chế độ thanh toán mặc dù mọi người đang sử dụng thanh toán trực tuyến nhiều hơn.

Không có nhiều thông tin được lấy từ các cột oldbalanceOrg, newbalanceOrig, nameDest, oldbalanceDest và newbalanceDest mặc dù chúng có giá trị tương quan tốt.

Thời gian xảy ra các giao dịch gian lận đa phần là từ 00 giờ đến 07 giờ, trong khi các giao dịch thông thường rất ít khi xảy ra.

4.4 So sánh các phương pháp khác

Như kết quả đã thể hiện ở **Bước 6, 4.2 Đánh giá từng bước**, ta thấy rằng chỉ số đặc hiệu AUC của việc kết hợp K-Means với Naive Bayes cho kết quả tốt hơn so với các phương pháp khác như Logistic Regression, K-nearest Neighbor, Decision Tree, Multi-layer Perceptron Classifier.

CHƯƠNG 5: THẢO LUẬN

5.1 Lợi ích của phương pháp đề xuất

Thuật toán K-means cung cấp một phương pháp dễ dàng để thực hiện giải pháp gần đúng cho phương trình này. Nguyên nhân của sự phổ biến này là sự dễ dàng và đơn giản trong thực thi, khả năng mở rộng, tốc độ hội tụ và khả năng thích ứng với dữ liệu thưa thớt. Thuật toán K-mean có thể được coi là một thủ tục giảm dần độ dốc, bắt đầu từ các trọng tâm cụm bắt đầu và sửa đổi lặp đi lặp lại các trọng tâm này để giảm hàm mục tiêu trong danh sách phương trình ở trên. Ngoài việc giải các bài toán có kết quả không duy nhất, thuật toán còn được áp dụng rộng rãi cho các dạng bài toán khác nhau. Các vấn đề về tính đồng nhất cũng như các vấn đề về phân phối điểm dữ liệu không đồng đều được giải quyết tốt hơn.

Thuật toán Naive Bayes là một thuật toán phân loại cho các vấn đề phân loại nhị phân (hai lớp) và đa lớp. Kỹ thuật này dễ hiểu nhất khi được mô tả bằng các giá trị đầu vào nhị phân hoặc phân loại. Đây là một thuật toán hiệu quả khi tính toán trên bộ dữ liệu lớn với mục đích là phân chia dữ liệu thành các loại nhãn riêng biệt. Dễ dàng và nhanh chóng để dự đoán lớp của tập dữ liệu thử nghiệm. Nó cũng hoạt động tốt trong dự đoán nhiều lớp. Khi giả định giữ độc lập, bộ phân loại Naive Bayes hoạt động tốt hơn so với các mô hình khác như hồi quy logistic và bạn cần ít dữ liệu đào tạo hơn. Nó hoạt động tốt trong trường hợp các biến đầu vào phân loại so với (các) biến số. Đối với biến số, phân phối chuẩn được giả định (đường cong hình chuông, một giả định mạnh).

5.2 Hạn chế của phương pháp đề xuất

Thuật toán K-means rất tốn kém về mặt tính toán vì nó liên quan đến một số phép tính khoảng cách của từng điểm dữ liệu từ tất cả các trọng tâm trong mỗi lần lặp. Kết quả cụm cuối cùng phụ thuộc rất nhiều vào việc lựa chọn trọng tâm ban đầu khiến nó hội tụ ở mức tối ưu cục bộ. K-Means chỉ có thể được áp dụng trên dữ liệu số. Nhưng cuộc sống hàng ngày, chúng ta gặp phải các tình huống có sự kết hợp của cả giá trị dữ liệu số và phân loại. Vì vậy, công việc trong tương lai có thể được thực hiện theo hướng làm cho thuật toán K-Means áp dụng cho loại dữ liệu đa dạng.

Hạn chế lớn nhất của Naive Bayes là giả định về các yếu tố dự đoán độc lập. Dù cho trong bài toán Phát hiện gian lận thẻ tín dụng hay trong thực tế, hầu như không thể có được một tập hợp các yếu tố dự đoán hoàn toàn độc lập. Đồng thời, mô hình khó có thể huấn luyện bằng các phương pháp tối ưu mạnh và chặt chẽ, tham số của mô hình là các

ước lượng xác suất điều kiện đơn lẻ. Không tính đến sự tương tác giữa các ước lượng này.

Với phương pháp kết hợp K-Means với Naive Bayes, ta thấy việc phát hiện hành vi gian lận thẻ tín dụng rất khó phát hiện. Mà mục tiêu khi kết hợp K-Means với Naive Bayes để phát hiện các hành vi gian lận và thực hiện trên dữ liệu số thực. Việc thu thập dữ liệu số thực rất khó khăn bởi sự thỏa thuận bí mật nên các nguồn thông tin không được chính xác cũng như không được gán nhãn đúng với từng hoạt động giao dịch. Điều này ảnh hưởng đến việc xác định được hết các số liệu cũng như tính chính xác của các hoạt động giao dịch. Trong tương lai, ta cần phải triển khai các phương pháp huấn luyện dữ liệu hợp lý với cấu trúc dựa trên các phân tính.

5.3 Định hướng tương lai

Vẫn không thể ngăn chặn hoàn toàn các hành động giao dịch gian lận ngay cả với tiến bộ công nghệ. Một số gian lận thẻ tín dụng thường được thực hiện bởi một nhóm chứ không phải một cá nhân cụ thể. Xác định một thực thể chung có xu hướng kết nối các thực thể khác là một trong những kỹ thuật để nhận biết các giao dịch lừa đảo. Cách tiếp cận này dẫn đến việc tìm ra mối liên hệ giữa các thực thể được cho là gian lận. Phương pháp này thường được sử dụng và hoạt động hiệu quả trong cơ sở dữ liệu. Số lượng lớn giao dịch lớn được tạo trong một thời điểm. Vì dữ liệu cần được xử lý riêng lẻ nên sẽ mất rất nhiều thời gian để xử lý từng dữ liệu một. GPU có thể được sử dụng để thực hiện song song các tác vụ, điều này dẫn đến việc trở thành thuộc tính không thể thể thiếu của GPU. Ngoài ra các nhà nghiên cứu đã đề cập rằng có thể đạt được sự bảo vệ mạnh mẽ hơn bằng cách triển khai các thuật toán cải tiến trong nhiều lớp.

KẾT LUẬN

Phát hiện gian lận thẻ tín dụng là một lĩnh vực nghiên cứu tích cực và xoay quanh khái niệm tự động hóa. Không phải lúc nào cũng khả thi hoặc có thể xem xét bằng các hệ thống thủ công từng giao dịch. Ngoài ra, điều quan trọng là phải xem xét rằng có một thành phần quan trọng khác của con người có thể thực hiện hoặc phá vỡ nỗ lực khai thác thành công thẻ của kẻ lừa đảo: sự kịp thời của chủ thẻ trong việc báo cáo thẻ bị đánh cắp, thất lạc hoặc sử dụng đáng ngờ. Điều này đòi hỏi phải triển khai các công cụ tự động để phát hiện gian lận nhanh hơn và thông minh hơn, dẫn đến các kỹ thuật máy học ngày càng được thử nghiệm và triển khai. Bài nghiên cứu này đề xuất phương pháp ứng dụng các giải thuật máy học trong phát hiện gian lận thẻ tín dụng với dữ liệu mô phỏng bộ dữ liệu thức thời cung cấp những biến cụ thể cho mô hình nghiên cứu. Kết quả cho thấy khả năng phát hiện gian lận khá tốt, phù hợp để các nhà phát triển lựa chọn nhằm giảm thiểu tổn thất tài chính và tối ưu hóa làm lợi nhuận.

PHỤ CHÚ

AUC (Area Under The Curve): biểu diễn mức độ phân loại của mô hình. $AUC = P(score(x+) > score(x-))$. Chỉ số AUC càng cao thì mô hình càng chính xác trong việc phân loại các lớp.

ROC (Receiver Operating Characteristics): là một đường cong biểu diễn xác suất. Đường cong ROC biểu diễn các cặp chỉ số (TPR, FPR) tại mỗi ngưỡng với TPR là trục tung và FPR là trục hoành.

TPR (True Positive Rate/Sensitivity/Recall): Biểu diễn tỷ lệ phân loại chính xác các mẫu dương tính trên tất cả các mẫu dương tính. TPR càng cao thì các mẫu dương tính càng được phân loại chính xác.

FPR (False Positive Rate/Fall-out): Biểu diễn tỷ lệ gắn nhãn sai các mẫu âm tính thành dương tính trên tất cả các mẫu âm tính.

TÀI LIỆU THAM KHẢO

Bank, E.C. Fifth Report on Card Fraud; European Central Bank: Frankfurt am Main, Germany, 2019.

Nilson. The Nilson Report|News and Statistics for Card and Mobile Payment Executives. Available online: Nilsonreport.com (accessed on 1 June 2019).

P.Bhati and M. Sharma, "Credit Card Number Fraud Detection Using K-Means with Hidden Markov Method," SSRG International Journal of Mobile Computing & Application (SSRG-IJMCA) , vol. II, no. 3, pp. 104-108, 2015.

N. B. Khandare, "Credit Card Fraud Detection Using Hidden Markov Model," International Journal Of Advance Scientific Research And Engineering Trends, vol. 1, no. 4, pp. 83-86, July 2016.

Dalatu, P.I. Time Complexity of K-Means and K-Medians Clustering Algorithms in Outliers Detection. Glob. J. Pure Appl. Math. 2018, 12, 4405–4418.

Bonaccorso, G. Machine Learning Algorithms: "Popular Algorithms for Data Science and Machine Learning"; Packt: Birmingham, UK, 2018.

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

Ghosh, S., Dasgupta, A., & Swetapadma, A. (2019). "A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification". 2019 International Conference on Intelligent Sustainable Systems (ICISS). doi:10.1109/iss1.2019.8908018

Pozzolo, A.D. "Learned lessons in credit card fraud detection from a practitioner perspective". Expert Syst. Appl. 2014, 41, 4915–4928.

Adhikari, R.; Agrawal, R.K. "An Introductory Study on Time Series Modeling and Forecasting". arXiv 2013, arXiv:1302.6613.