BỘ CÔNG THƯƠNG

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI

ĐỖ XUÂN SƠN

GIẢI PHÁP ỨNG DỤNG TRÍ TUỆ NHÂN TẠO NHẰM PHÁT HIỆN GIAO DỊCH BẤT THƯỜNG TRONG HỆ THỐNG QUẢN TRỊ GIAO DỊCH TÀI CHÍNH

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ HỆ THỐNG THÔNG TIN

Hà Nội – 2024

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI

ĐỖ XUÂN SƠN

GIẢI PHÁP ỨNG DỤNG TRÍ TUỆ NHÂN TẠO NHẰM PHÁT HIỆN GIAO DỊCH BẤT THƯỜNG TRONG HỆ THỐNG QUẢN TRỊ GIAO DỊCH TÀI CHÍNH

Ngành hệ thống thông tin
Mã số 8480104

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ HỆ THỐNG THÔNG TIN

NGƯỜI HƯỚNG DẪN:
TS.Vũ Việt Thắng

Hà Nội – 2024

MỤC LỤC

DANH MỤC HÌNH ẢNH

DANH MỤC CÁC KÝ HIỆU VÀ TỪ VIẾT TẮT

Viết tắt	Tiếng Anh	Tiếng Việt
eKYC	electronic Know Your Customer	Công nghệ định danh khách hàng điện tử
MS	Model Server	Máy chủ triển khai mô hình học máy
IF	Isolation Forest	Thuật toán cô lập các điểm dữ liệu
OTS	Open Table Service	Công cụ lập kế hoạch SQL
CF	Column Family	Cơ sở dữ liệu NoSQL sử dụng hàng và cột
TID	Transaction Identification	Định danh giao dịch
AI	Artificial Intelligent	Trí tuệ nhân tạo
SemiGNN	Semi Supervised Graph Neural	Mạng nơron đồ thị bán giám sát
GTAN	Gated Temporal Attention Network	Mạng nơron sử dụng cơ chế chú ý sử dụng các thông tin thời gian
MLP	multi-layer perception Protocol	Giao thức nhận thức nhiều lớp
TGA	Temporal Graph Attention	Đồ thị sử dụng cơ chế chú ý thời gian
GDS	Graph Data Science	Khoa học dữ liệu sử dụng lý thuyết đồ thị

DANH MỤC BẢNG

LỜI CAM ĐOAN

Những nội dung được trình bày trong luận văn là những kiến thức của riêng cá nhân em tích lũy trong quá trình học tập, nghiên cứu, không sao chép lại của một công trình nghiên cứu hay luận văn của bất cứ tác giả nào.

Trong nội dung của luận văn, những phần em đã nghiên cứu, trích dẫn đều được nêu trong các tài liệu tham khảo, có nguồn gốc, xuất xứ tên tuổi của các tác giả và nhà xuất bản rõ ràng.

Những điều em cam kết hoàn toàn là sự thật, nếu sai, em xin chịu mọi hình thức kỷ luật theo quy định.

Hà Nội, ngày 19 tháng 05 năm 2024

Học viên thực hiện

Đỗ Xuân Sơn

MỞ ĐẦU

Như chúng ta đã biết hiện nay với sự phát triển bùng nổ của công nghệ thông tin đã tác động lớn đến các hoạt động thường ngày của xã hội từ việc dạy học, làm việc, đến các nhu cầu giải trí ... Do đại dịch Covid-19, các biện pháp phòng chống dịch phong toả, giữ khoảng cách, … được triển khai. Từ những vấn đề này gây ra nhiều trở ngại đối với các hoạt động cần tiếp xúc trực tiếp như lĩnh vực ngân hàng là một trong số đó. Do đó giao dịch thanh toán trực tuyến không dùng tiền mặt trở nên thuận tiện và phát triển hơn bao giờ hết.

Tại Hoa Kỳ, các hoạt động gian lận làm cho số lượng khách hàng bị thiệt hại đạt mức kỷ lục 15,4 triệu người, cao hơn 16% so với năm 2015 và gây thiệt hại khoảng 6 tỷ đô.

Theo thống kê của Ngân hàng Nhà nước Việt Nam, số lượng và giá trị giao dịch thanh toán điện tử năm 2019 qua kênh Mobile Banking tăng trưởng lần lượt là 198% và 210%; các kênh Internet Banking và ví điện từ tăng trưởng từ 37% - 86% so với cùng kỳ. Ngày 26/5/2020, Thủ tướng Chính phủ đã ban hành chỉ thị số 22/CT-TTg về việc đẩy mạnh triển khai các giải pháp phát triển thanh toán không dùng tiền mặt tại Việt Nam, qua đó thúc đẩy mạnh mẽ các phương thức giao dịch điện tử, đồng thời cho thấy xu hướng tất yếu của việc chuyển dịch hình thức thanh toán. Trong bối cảnh đó, ngân hàng và các công ty công nghệ tài chính (Fintech) đã và đang tập trung phát triển các ứng dụng mới trên nền tảng công nghệ số, cung cấp những sản phẩm, dịch vụ đi kèm với những hình thức khuyến mãi vô cùng hấp dẫn hướng tới việc đáp ứng kịp thời nhu cầu thay đổi của khách hàng.

Bên cạnh những lợi ích to lớn mà giao dịch tài chính trực tuyến mang đến, người sử dụng đã và đang phải đối mặt với những rủi ro tiềm ẩn từ việc thất thoát dữ liệu cá nhân. Nếu không có nhận thức đầy đủ, người sử dụng sẽ

dễ dàng trở thành mục tiêu của tội phạm tài chính, có nguy cơ bị lợi dụng cho những giao dịch bất chính và thiệt hại tài sản không mong muốn.

Các ngân hàng trước đây sử dụng cách bảo mật truyền thống dựa vào yếu tố con người như cán bộ, nhân viên ngân hàng và khách hàng sử dụng dịch vụ. Trong đó các cán bộ nhân viên được giao nhiệm vụ thường xuyên giám sát các quy trình kỹ thuật trong hoạt động thanh toán… để phát hiện gian lận, khách hàng thường được khuyến cáo bằng việc tự bảo vệ tài khoản của mình. Dễ dàng nhận thấy cách truyền thống có độ bảo mật an toàn không cao bởi các hoạt động gian lận hiện nay ngày càng diễn ra hết sức tinh vi và phức tạp hơn. Do đó, vấn đề đặt ra cần có các giải pháp thông minh hơn, hiệu quả hơn để tăng cường tính bảo mật trong giao dịch tài chính thông qua việc có thể tự động phát hiện ra các giao dịch bất thường (hoặc các giao dịch có nguy cơ chứa đựng gian lận) một cách nhanh chóng và hiệu quả.

Trong bài báo [1] tác giả và nhóm cộng sự cũng nêu rõ các thiệt hại to lớn trong các giao dịch tài chính có tính bảo mật kém. Đặc biệt trong thời kỳ đại dịch covid đã làm đẩy nhanh các giao dịch tài chính trực tuyến. Trong nghiên cứu cũng giới thiệu tổng quan về các phương pháp thông minh nhằm phát hiện các giao dịch tài chính bất thường một cách hiệu quả. Cũng trong bài báo này tác giả có thống kê các giải pháp dựa trên máy học như: SVM, CNN, Naïve Bayes, LSTM, v.v. Cuối cùng các tác giả cũng nêu rõ các thách thức trong việc phát hiện ra các giao dịch tài chính bởi các đối tượng phá hoại luôn có xu hướng thay đổi công nghệ để tránh khỏi bị hệ thống phát hiện.

Trong bài báo [2] tác giả Shaosheng Cao và các cộng sự đã giới thiệu hệ thống TitAnt có khả năng phát hiện ra các gian lận trong giao dịch tài chính online trong thời gian thực. Trong nghiên cứu cũng trình bày các vấn đề, các phương pháp trích chọn đặc trưng và phát hiện giao dịch có gian lận bao gồm:

thuật toán dựa trên tập luật, thuật toán hồi quy logic, thuật toán Gradient Boosting Decision Tree (GBDT), v.v.. Để đảm bảo hệ thống có thể hoạt động trong thời gian thực nhóm tác giả đã triển khai mô hình dựa trên các công cụ hỗ trợ mã nguồn mở như: MaxCompute, KunPeng, MS và Ali-Hbase. Các công cụ này có nhiệm vụ hỗ trợ lưu trữ phân tán, huấn luyện và dự đoán các giao dịch online.

Mặc dù cũng đã có nhiều nghiên cứu khác nhau về bài toán này tuy nhiên chưa có một giải pháp nào là tuyệt đối. Theo thời gian các đối tượng thực hiện hành vi gian lận cũng sẽ thay đổi cách thức thực hiện nhằm tránh sự phát hiện của hệ thống và điều này luôn là một thách thức với những người có trách nhiệm bảo mật trong các hệ thống ngân hàng hay các công ty tài chính.

Chính vì các lý do trên đây em đã chọn đề tài "***Giải pháp ứng dụng trí tuệ nhân tạo nhằm phát hiện giao dịch bất thường trong hệ thống quản trị giao dịch tài chính***".

CHƯƠNG 1: TỔNG QUAN VỀ GIAN LẬN TRONG GIAO DỊCH TÀI CHÍNH

1.1. Giới thiệu bài toán

Ngày nay các dịch vụ tài chính đặc biệt là dịch vụ tài chính trực tuyến mang lại nhiều lợi ích kinh tế to lớn cho xã hội. Tuy nhiên cùng với đó các hành vi gian lận xuất hiện ngày càng nhiều.

Gian lận tài chính là hành vi cố ý lừa dối liên quan đến các giao dịch tài chính nhằm mục đích thu lợi cá nhân. Các trường hợp gian lận thường được thực hiện bởi các tin tặc nhằm chiếm đoạt tài khoản của người dùng sau đó thực hiện các giao dịch chuyển tiền nhằm đánh cắp tài sản của người bị hại, các chuyên gia kinh doanh có kiến thức chuyên môn và có ý định lạm dụng kiến thức tài chính của bản thân nhằm chuộc lợi, những tội phạm thực hiện các giao dịch nhằm mục đích rửa tiền mà chúng có được từ các hoạt động phi pháp,… Điển hình như tại Hoa Kỳ, số lượng khách hàng bị gian lận đạt mức kỷ lục 15,4 triệu người, cao hơn 16% so với năm 2015 và gây thiệt hại khoảng 6 tỷ đô. Trên thực tế tồn tại nhiều loại gian lận, như gian lận trong dịch vụ thẻ tín dụng, gian lận bảo hiểm, v.v.

Tất cả những hành vi gian lận này sẽ gây tổn hại nghiêm trọng đến bảo mật cho cả người dùng và nhà cung cấp dịch vụ. Vì vậy, làm thế nào để phát hiện gian lận là một vấn đề quan trọng cần được nghiên cứu.

1.2. Các loại hình gian lận trong tài chính

1.2.1. Gian lận thông qua mã khuyến mãi

Là hành vi lợi dụng một số lỗ hổng khi doanh nghiệp đưa ra các chương trình khuyến mãi từ đó sử dụng nhiều lần mã khuyến mãi nhằm chuộc lợi. Các hoạt động gian lận phổ biến như:

- Gian lận ưu đãi: khuyến mãi tại Shopee đơn giản là các hành vi tạo nhiều tài khoản người mua trên Shopee với mục đích nhận quà tặng, mã giảm giá Shopee cho khách hàng mới, ưu đãi liên kết ví Shopee Pay (trước đây là ví AirPay). Những ưu đãi này các khách hàng cũ đã được nhận thì sẽ không được nhận nữa. Trước đây tận dụng kẽ hở này có nhiều người đã tạo ra những công cụ tự động đăng ký hàng trăm, thậm chí hàng ngàn tài khoản người mua trên Shopee để đặt mua các đơn hàng 0đ. Ngoài ra các mã ưu đãi liên quan tới vận chuyển, thanh toán Shopee cho người mới cũng rất lớn, điển hình là mã freeship shopee 40,000đ cho đơn 0đ. Hiện tại hình thức này vẫn còn tồn tại nhỏ lẻ ở một số bạn chưa nắm rõ thông tin, chính sách cũng như khả năng nhận diện chính xác những tình huống này tại Shopee và dẫn đến tài khoản Shopee bị khoá. Hiện nay, Shopee đã đưa ra nhiều hình thức chống gian lận như khóa các tài khoản mới khi thấy nghi ngờ, tăng thời gian phạt đối với tài khoản cũ hoặc nghiêm trọng hơn là khóa luôn tài khoản cũ vi phạm.
- Nhân viên bán hàng gian lận: là những người trực tiếp tiếp xúc với khách hàng, thực hiện các giao dịch thanh toán liên quan tới tiền bạc nên nhân viên dễ dàng thực hiện gian lận mỗi khi có cơ hội. Đặc biệt nếu các chủ cửa hàng lơ là, lỏng lẻo trong việc quản lý thì hậu quả có thể bị thất thoát lên tới hàng chục, hàng trăm triệu đồng.
- Báo sai giá**:** Thông thường kiểu gian lận này sẽ xảy ra đối với những cửa hàng kinh doanh nhỏ, không treo bảng giá và vẫn sử dụng các cách

bán hàng truyền thống. Như vậy mỗi khi nhà quản lý vắng mặt, nhân viên bán hàng sẽ tự ý tăng giá thêm để lấy số tiền chênh lệch. Tuy hình thức gian lận này không khiến cửa hàng bị thất thoát hàng hóa hay tiền bạc nhưng sẽ khiến cho thương hiệu bị mất uy tín.

- Không ghi chép đơn hàng/ in hóa đơn**:** Đây là mánh khóe gây ảnh hưởng trực tiếp đến doanh thu của cửa hàng. Khi mỗi đơn hàng mà khách hàng thanh toán, nhân viên chỉ thanh toán và cầm tiền mà không ghi lại đơn hàng hoặc sau khi khách về thì nhấn nút "Hủy hóa đơn". Như vậy dữ liệu trong hóa đơn sẽ không được lưu trong phần mềm bán hàng và đương nhiên bạn sẽ hoàn toàn không biết có một giao dịch đơn hàng vừa diễn ra. Với cách thức gian lận này, hàng hóa thì bị mất còn số tiền thanh toán lại "về túi" nhân viên.
- Sửa hóa đơn**:** Sau khi thực hiện xong giao dịch và vẫn in hóa đơn cho khách thì nhân viên bán hàng truy cập vào hệ thống để xem và sửa lại hóa đơn. Họ có thể giảm số lượng hàng hóa hoặc xóa bớt một món hàng. Sao cho số tiền khách hàng đã trả phải nhiều hơn số tiền trên hệ thống và nhân viên sẽ lấy tiền chênh lệch. Việc này sẽ diễn ra nhiều lần nếu bạn thường xuyên không có mặt tại cửa hàng. Và việc phát hiện nhân viên tự ý chỉnh sửa hóa đơn sẽ là rất khó nếu không có các công cụ hỗ trợ.
- Lợi dụng các chương trình ưu đãi: Mỗi khi cửa hàng có chương trình giảm giá, nếu chỉ đơn giản là treo biển và thực hiện mọi thông báo theo cách thủ công thì sẽ là cơ hội lớn để nhân viên bán hàng thực hiện gian lận. Bởi nhân những lúc bạn không để ý, họ có thể đánh tráo các biển giảm giá hoặc tự ý chỉnh số tiền chiết khấu nhiều hơn quy định. Hoặc như chỉnh giá sản phẩm lên cao rồi chiết khấu đúng với giá ban đầu. Như vậy cửa hàng bạn sẽ bị mất uy tín, tổ chức chương trình ưu

đãi mà giá sản phẩm vẫn cao, khách hàng họ sẽ không muốn quay lại lần thứ 2.

- Làm hỏng các dữ liệu bán hàng: thực tế, không ít nhân viên bán hàng sau khi thực hiện nhiều lần gian lận thì đã cố tình làm hỏng sổ sách hoặc máy tính lưu trữ dữ liệu bán hàng để các chủ quản lý không phát hiện. Điều này sẽ gây hậu quả nghiêm trọng bởi không chỉ là số liệu bán hàng, mà họ còn đang làm hỏng cả những kế hoạch tương lai, những thông tin dữ liệu khách hàng, các báo cáo thống kê tình hình bán hàng cũng như số lượng hàng hóa, thu chi công nợ,… mà bạn ghi chép trong sổ sách hoặc lưu trên file máy tính.

1.2.2. Gian lận thông qua chính sách thành viên

Người bán nhận được khoản bồi hoàn vì chủ thẻ từ chối nhận đơn đặt hàng nhưng thực tế hàng hoá đã được thanh toán. Tình trạng này có thể bắt gặp các cửa hàng có kinh doanh trực tuyến trên các trang thương mại điện tử. Lợi dụng chính sách hoàn tiền của các sàn thương mại điện tử hỗ trợ cửa hàng trong trường hợp khách trả hàng hoàn tiền. Cửa hàng đã thoả thuận với người dùng hoặc tạo tài khoản tự mua bán để chuộc lợi từ chính sách này .

Ngoài những lợi ích của các chương trình khách hàng thân thiết thường được mọi người đánh giá cao, cho dù những lợi ích đó được đưa ra dưới dạng điểm, giảm giá, tiền thưởng, quà tặng hoặc dặm bay. Các doanh nghiệp đã nhận ra rằng lòng trung thành của khách hàng là rất quan trọng để duy trì và tăng thị phần, và các chương trình khách hàng thân thiết được thiết kế để thiết lập mối quan hệ lâu dài và có lợi với khách hàng.

Từ quan điểm kinh doanh, các chương trình khách hàng thân thiết giúp tăng tỷ lệ giữ chân khách hàng, giảm chi phí tiếp thị để có được khách hàng mới, tác động đến việc mua hàng của khách hàng đối với thương hiệu và xác

định mối liên hệ giữa lợi ích của khách hàng thân thiết và hành động của khách hàng (như nghiên cứu tiếp thị một phần). Từ quan điểm của khách hàng, việc đăng ký tham gia các chương trình khách hàng thân thiết mang lại cảm giác có đi có lại (nhận được nhiều thứ hơn so với lần mua ban đầu), sự công nhận đặc biệt, sự tin tưởng và cam kết của tổ chức bên cạnh các lợi ích như giảm giá.

Nhưng có một vấn đề ít được nhận ra liên quan đến giá trị của các chương trình khách hàng thân thiết: gian lận. Trong khi 81% người Mỹ đánh đồng điểm thưởng tích lũy của họ với tiền mặt, hầu hết người tiêu dùng không thường xuyên kiểm tra số dư tài khoản khách hàng thân thiết của họ. Hơn nữa, khoảng 20% thành viên nhận thưởng chưa bao giờ đổi bất kỳ điểm tích lũy nào của họ. Những điểm không được sử dụng và không được giám sát này đã trở thành mục tiêu chính cho những kẻ lừa đảo đánh cắp để sử dụng cho riêng chúng hoặc bán trên dark web. Có nhiều phương pháp khác nhau để thực hiện hành vi gian lận trong các chương trình khách hàng thân thiết và việc bảo vệ chương trình của bạn chống lại chúng cũng đòi hỏi nhiều cách tiếp cận và nỗ lực khác nhau.

Trong nhiều năm, những chương trình khách hàng thân thiết dành cho các doanh nghiệp nhỏ hơn đã dựa vào thẻ giấy được đục lỗ hoặc đóng dấu khi mua hàng được thực hiện.Các chương trình khách hàng thân thiết này rất đơn giản để bắt đầu, không tốn kém để sản xuất và dễ học. Nhưng thẻ rất dễ bị làm giả, có thể gian lận thông qua các hành vi không hợp lệ và không cung cấp cho nhà phát hành bất kỳ dữ liệu khách hàng nào có thể sử dụng được. Những thay đổi công nghệ đơn giản, chẳng hạn như máy quét khiến gian lận chương trình dễ dàng thực hiện để nhận phần thưởng. Nhiều chương trình khác đang chuyển từ các đối tượng vật lý sang cơ sở kỹ thuật số bằng cách kết hợp việc sử dụng các ứng dụng điện thoại thông minh mang lại lợi ích cho cả khách hàng và doanh nghiệp. Các ứng dụng dành cho thiết bị di động giúp giảm sự lộn xộn trong ví của khách hàng và chuỗi khóa, giảm thiểu khả năng xảy ra các hoạt

động trái phép và cung cấp các phân tích có giá trị không khả dụng với thẻ khách hàng thân thiết. Các ứng dụng này cũng có thể giúp doanh nghiệp cá nhân hóa thông tin về người bảo trợ để thiết kế các phần thưởng hoặc dịch vụ khác biệt. Ngày nay, chương trình khách hàng thân thiết điển hình cho phép các thành viên của mình tích lũy điểm khi mua hàng của khách hàng trong tài khoản cá nhân trực tuyến. Điểm có thể được đổi lấy phần thưởng như thẻ quà tặng, du lịch và bữa ăn. Mặc dù chúng không phải là tiền mặt nhưng các điểm này có giá trị tiền tệ trong thế giới thực. Ví dụ chỉ riêng ở Hoa Kỳ ước tính có khoảng 48 tỷ đô la dành cho khách hàng thân thiết.

Các chương trình khách hàng thân thiết cung cấp một mỏ vàng thông tin cho các doanh nghiệp, nhưng chúng cũng có thể cho phép những kẻ lừa đảo truy cập vào kho thông tin này với nỗ lực tối thiểu. Thông tin có trong các trang web của chương trình phần thưởng, chẳng hạn như thông tin nhận dạng cá nhân (PII) thể dễ dàng được sử dụng để thực hiện hành vi trộm cắp danh tính. PII thường bao gồm các chi tiết như tên, ngày sinh, địa chỉ email và gửi thư, số điện thoại, số thẻ tín dụng, tình trạng hôn nhân, quy mô hộ gia đình và thu nhập hàng năm.

Một số thống kê chỉ ra rằng gian lận đã ảnh hưởng đến hơn 70% các chương trình khách hàng thân thiết. Trong các chương trình này, hành vi gian lận cũng có thể xảy ra thông qua hành vi trộm cắp điểm tích lũy, thông tin hoặc thông qua hành vi đánh lừa hệ thống để tạo ra điểm. Hầu hết người tiêu dùng thường không xem lại điểm tích lũy của họ và không nhận thấy khi những điểm đó bị xâm phạm. Chỉ khi khách hàng muốn sử dụng điểm của mình, họ mới thấy số điểm đó không còn nữa.

Quy tắc của các chương trình khách hàng thân thiết thường yêu cầu cá nhân có tên trên tài khoản phải là người tạo ra điểm. Đánh lừa hệ thống trong các chương trình khách hàng thân thiết thường có nghĩa là thành viên chương

trình cho phép người khác sử dụng thẻ khách hàng thân thiết hoặc số của mình để tạo điểm tích lũy cho thành viên. Những kỹ thuật như vậy rất khó nếu cần phải nhận dạng để truy cập vào tài khoản khách hàng thân thiết, nhưng có thể cực kỳ dễ dàng khi điểm được mua trực tuyến hoặc thông qua một ứng dụng.

Thủ phạm gian lận chương trình khách hàng trung thành thường có thể được phân chia thành ba loại chính: Tin tặc (hacker), người trong cuộc và thành viên.

Tin tặc

Tin tặc là những người bên ngoài (bao gồm cả các thành viên của mạng lưới tội phạm có tổ chức), những kẻ khai thác lỗ hổng bảo mật của chương trình và mật khẩu yếu của khách hàng để đánh cắp điểm thưởng tích lũy. Những cá nhân này sử dụng các phương pháp như kế hoạch lừa đảo hoặc các hình thức kỹ nghệ xã hội khác để thu thập thông tin nhằm xâm nhập vào tài khoản của thành viên. Tiền của chương trình khách hàng thân thiết được coi là một mục tiêu dễ dàng chiếm đoạt, chủ yếu là do nhận thức của người tiêu dùng thấp liên quan đến việc giám sát. Ngoài ra bảo mật xung quanh các chương trình khách hàng thân thiết gần như không mạnh bằng các tài khoản hoạt động bằng tiền thật. Điểm bị đánh cắp có thể được tin tặc sử dụng để nhận phần thưởng miễn phí.

Ví dụ như sau một vụ tấn công vào chương trình Hilton Honors năm 2014, tài khoản của một thành viên đã được sử dụng để thanh toán cho sáu lần lưu trú khách sạn tại các cơ sở kinh doanh của Hilton.Thẻ tín dụng của công ty được liên kết với tài khoản sau đó đã được sử dụng để mua thêm điểm thưởng cho tin tặc. Điểm bị hack cũng có thể trở thành một phần của gian lận tam giác, liên quan đến khách hàng thân thiết, tin tặc và bên thứ ba (thường là trang web hợp pháp hoặc "chợ dành cho tin tặc"). Trong vụ hack Hilton Honors, nhiều

điểm rút khỏi tài khoản sau đó được rao bán trực tuyến với giá rất rẻ so với giá trị thật.

Người trong cuộc

Người trong cuộc là nhân viên của doanh nghiệp cung cấp các chương trình khách hàng thân thiết hoặc những người có quyền truy cập vào hệ thống.

Ví dụ: nếu thẻ bấm lỗ được sử dụng trong các chương trình khách hàng thân thiết và những người trong cuộc có thể dễ dàng bấm thêm vào thẻ của bạn bè họ. Ngay cả với sự ra đời của các thiết bị chương trình tinh vi hơn thì nhân viên vẫn có thể thao túng hệ thống điểm trung thành. Nếu khách hàng không phải là thành viên của chương trình khách hàng thân thiết hoặc quên sử dụng liên kết khách hàng thân thiết của mình khi mua hàng, nhân viên (chẳng hạn như đại lý trung tâm cuộc gọi, tiếp viên hàng không và nhân viên quầy làm thủ tục) có thể ghi có giao dịch mua cho họ. Tài khoản cá nhân của mình hoặc của các thành viên gia đình hoặc bạn bè. Tùy thuộc vào mức độ làm việc của họ, nhân viên cũng có thể có quyền điều chỉnh hoặc thêm điểm vào tài khoản khách hàng như một biện pháp hỗ trợ trong trường hợp có vấn đề với thẻ hoặc thiết bị đầu cuối tại điểm bán hàng. Điều này có thể bị lạm dụng bằng cách đưa ra các khoản tín dụng không chính đáng và khả năng chuyển điểm từ thẻ này sang thẻ khác cũng có thể bị lạm dụng. Đó có thể là một hoạt động cần thiết và hợp pháp của nhân viên khi thẻ của khách hàng bị mất hoặc bị đánh cắp, nhưng đó không phải là điểm được chuyển từ thẻ không hoạt động.

Các thành viên

Các thành viên là những khách hàng tham gia chương trình thực hiện hành vi gian lận khi họ cố gắng "đánh lừa hệ thống" để có lợi cho họ. Một thành viên chương trình cố gắng đổi điểm đồng thời qua điện thoại với đại diện công ty và thông qua tài khoản trực tuyến của họ. Hoặc thay vì đổi điểm các thành viên có thể cố gắng tích lũy điểm một cách gian lận thông qua việc đính kèm

số tài khoản phần thưởng của họ với giao dịch mua mà họ không thực hiện. Hầu hết các chương trình cho phép các thành viên tặng điểm hoặc phần thưởng của họ cho người khác, nhưng việc bán điểm thường bị cấm theo chính sách của chương trình khách hàng thân thiết. Do đó, các thành viên bán hoặc trao đổi điểm của họ đang vi phạm gian lận chương trình. Các thành viên cũng được biết là thực hiện các giao dịch mua tạo ra số lượng lớn điểm thưởng và sau đó hủy giao dịch mua nhưng không phải trước khi điểm được đổi để nhận giải thưởng tiền mặt. Một số chương trình khách hàng thân thiết cho phép kiếm điểm cho các tương tác trên mạng xã hội như chuyển tiếp tin nhắn, đánh giá và giới thiệu. Để đạt được điểm, các thành viên có thể "chia sẻ quá nhiều", đăng các bài đánh giá không đáng kể hoặc giới thiệu một số lượng lớn các cá nhân không có khả năng trở thành thành viên của chương trình khách hàng thân thiết. Trong những tình huống này, các thành viên đang thu được giá trị từ chương trình khách hàng thân thiết bằng cách tham gia vào các hoạt động không tạo ra giá trị gia tăng cho doanh nghiệp.

1.2.3. Giả mạo danh tính

Kẻ xấu có thể mua thông tin từ các nguồn thông tin bị rò rỉ hoặc đánh cắp thông tin của người khác để thực hiện hành vi lừa đảo, chiếm đoạt tài sản. Ngoài ra nhân viên có thể tham gia gian lận bằng cách sử dụng trái phép thông tin khách hàng khi có quyền truy cập vào hệ thống. Một trong những dấu hiệu tài khoản bị đánh cắp có nhiều lần chuyển điểm thưởng, chuyển tiền trong một khoảng thời gian ngắn đặc biệt là cho người không có tên trên tài khoản, vào những cung giờ khác biệt cũng có thể cho thấy hoạt động đáng ngờ. Hàng hóa trên các kênh mua sắm trực tuyến được vận chuyển đến một địa chỉ không được liên kết với tài khoản của thành viên cũng có thể cho thấy khả năng gian lận.

Ngoài ra với sự phát triển của các công nghệ giả mạo danh tính do AI tạo ra cũng dẫn đến tình trạng giả mạo danh tính của các tổ chức cơ quan chức năng nhằm đánh vào tâm lý sợ hãi của người dùng để yêu cầu người dùng thực hiện các hành vi theo chỉ định của chúng nhằm lừa đảo chiếm đoạt tài khoản và tài sản của bị hại. Các trường hợp khác kẻ xấu sẽ sử dụng thông tin thu thập được từ mạng xã hội và công cụ AI giả mạo danh tính bạn bè, người thân của người dùng để lừa người dùng chuyển khoản, thanh toán hộ để chiếm đoạt tài sản .

Công nghệ định danh khách hàng điện tử (electronic Know Your Customer - eKYC) là việc thiết lập mối quan hệ và định danh khách hàng bằng các phương tiện điện tử, bao gồm kênh trực tuyến và kênh di động, mà không cần phải gặp mặt trực tiếp đã mang đến sự tiện lợi cho cả khách hàng và những công ty/tổ chức sử dụng công nghệ này. Nhờ việc áp dụng công nghệ eKYC, các tổ chức có thể định danh khách hàng từ xa để thể thu thập thông tin về đặc điểm sinh trắc học của khách hàng và xác thực với các nguồn dữ liệu cơ sở như thông tin trên giấy tờ tùy thân, cơ sở dữ liệu dân cư, cơ sở nhận dạng ... Do đó, trong quy trình định danh điện tử có hai yếu tố đặc biệt quan trọng là: nguồn dữ liệu tin cậy làm cơ sở đối chiếu và độ chính xác của nền tảng công nghệ áp dụng để thu thập các thông tin sinh trắc học của khách hàng. Để đảm bảo an toàn và hiệu quả khi định danh điện tử, nhà nước đã đưa ra khung pháp lý để các doanh nghiệp, tổ chức tuân thủ như: (Thông tư số 16/2020/TT-NHNN ngày 04/12/2020 của Thống đốc Ngân hàng Nhà nước (NHNN) sửa đổi, bổ sung một số điều của Thông tư số 23/2014/TT-NHNN ngày 19/8/2014 của Thống đốc NHNN hướng dẫn việc mở và sử dụng tài khoản thanh toán tại tổ chức cung ứng dịch vụ thanh toán), ngành Ngân hàng đã lần lượt ứng dụng eKYC vào quy trình nhận biết khách hàng và cung cấp các sản phẩm, dịch vụ tài chính. Chỉ trong năm 2020, hàng loạt ngân hàng liên tiếp công bố đã hoàn thiện quy trình

công nghệ, sẵn sàng thực hiện mở tài khoản không gặp mặt trực tiếp khách hàng.

Đến cuối năm 2021, đã có 24 tổ chức tín dụng chính thức triển khai mở tài khoản thanh toán eKYC, với khoảng 3,37 triệu tài khoản thanh toán mở bằng phương thức này đang hoạt động, đây được coi là một trong những giải pháp đột phá giúp thúc đẩy tài chính toàn diện, đưa ngân hàng đến gần hơn với khách hàng, đồng thời góp phần thực hiện quá trình chuyển đổi số của ngân hàng. Đặc biệt trong bối cảnh dịch bệnh Covid-19, việc ban hành kịp thời chính sách này đã giúp khách hàng tiếp cận, sử dụng dịch vụ thanh toán mà không phải đến quầy giao dịch của ngân hàng.

Mặc dù nhà nước đã đưa ra thông tư và có những khung hình phạt với tội danh về chiếm đoạt tài sản tuy nhiên vẫn có nhiều kẻ xấu lợi dụng việc áp dụng rộng rãi công nghệ eKYC để giả mạo danh tính của người khác. Tiêu biểu như hành vi một người giả mạo danh tính của một người thật khác bằng cách sử dụng tài liệu bị đánh cắp, kết hợp với thông tin được làm giả, thay thế thông tin giả mạo đó lên trên các giấy tờ của một người bằng hình ảnh của kẻ mạo danh từ đóng đánh lừa hệ thống eKYC. Trong quy trình định danh điện tử, bằng các thủ thuật tinh vi đối tượng mạo nhận danh tính của một người để mở tài khoản hoặc đánh cắp mật khẩu, thông tin của người dùng nhằm thực hiện giao dịch điện tử. Hiện nay, chất lượng hình ảnh và mức độ chân thực từ hình ảnh được tạo ra bởi các công nghệ như công nghệ "deepfake" đang có những bước tiến lớn tạo điều kiện cho kẻ xấu có thể sử dụng để sử dụng các hình ảnh giả mạo, chỉnh sửa để vượt qua các bước xác thực về sinh trắc học của hệ thống eKYC, đã trở thành mối đe dọa lớn. Khác với việc gặp mặt khách hàng trực tiếp thuận lợi để đánh giá chất lượng của bản gốc giấy tờ tùy thân cũng như nhận diện trực tiếp cử chỉ, đặc điểm sinh trắc và chữ ký của khách hàng thì eKYC lại nhận dạng các giấy tờ qua các hình ảnh ghi lại trong quá trình định

danh dễ bị làm giả hoặc sử dụng công nghệ cắt ghép ảnh hơn nữa độ phân giải camera của các thiết bị được sử dụng để eKYC cũng ảnh hưởng lớn đến việc xác thực mức độ tin cậy về hình ảnh do khách hàng cung cấp. Theo đó, rủi ro là không thể tránh khỏi khi áp dụng nền tảng công nghệ mới như eKYC việc không phải khách hàng nào cũng có kiến thức về công nghệ để tự thực hiện được việc định danh điện tử cá nhân mà phải nhờ sự hỗ trợ của người khác cũng là khó khăn và rủi ro lớn. Ngoài ra khách hàng cùng cần có ý thức về bảo vệ thông tin cá nhân hạn chế chia sẻ các thông tin nhạy cảm về cá nhân như thông tin về giấy tờ tùy thân và đặc điểm sinh trắc học cho người khác và lên mạng xã hội để kẻ xấu có thể lợi dụng.

1.2.4. Đánh cắp tài khoản

Hiện nay, tình trạng tin tặc (hacker) đánh cắp tài khoản ngân hàng vẫn đang diễn ra phổ biến trên toàn cầu. Các tin tặc thường sử dụng các kỹ thuật phần mềm độc hại, mạo danh và lừa đảo để truy cập vào tài khoản ngân hàng của người dùng và đánh cắp thông tin cá nhân, thông tin tài khoản, mật khẩu và số tiền trong tài khoản.

Một số kỹ thuật phổ biến được sử dụng bởi các hacker bao gồm:

- Phishing: là một kỹ thuật lừa đảo trực tuyến, mà kẻ tấn công sử dụng các email giả mạo, tin nhắn văn bản, trang web giả mạo hoặc các tin nhắn trên mạng xã hội để lừa đảo người dùng cung cấp thông tin cá nhân, thông tin tài khoản, mật khẩu hoặc số tiền trong tài khoản của họ. Tình trạng phishing chiếm đoạt tài khoản vẫn là một vấn đề lớn trên toàn cầu, với hàng nghìn người bị lừa đảo và mất tiền mỗi năm. Các kẻ tấn công thường sử dụng các email giả mạo của các tổ chức, ngân hàng, hoặc các nhà cung cấp dịch vụ để tạo ra một cảm giác tin cậy và đáng tin cậy cho người nhận. Sau đó, họ yêu cầu người dùng truy cập vào một trang web

giả mạo hoặc nhập thông tin cá nhân, thông tin tài khoản, hoặc mật khẩu của họ. Với hình thức tấn công Phishing giả mạo email tin tặc sẽ gửi email cho người dùng dưới danh nghĩa một đơn vị/tổ chức uy tín, dụ người dùng click vào đường link dẫn tới một website giả mạo. Những email giả mạo thường rất giống với email chính chủ, chỉ khác một vài chi tiết nhỏ, khiến cho nhiều người dùng nhầm lẫn và trở thành nạn nhân của cuộc tấn công. Để làm cho nội dung email giống thật nhất có thể, kẻ tấn công luôn cố gắng "ngụy trang" bằng nhiều yếu tố như: địa chỉ người gửi (VD: địa chỉ đúng là **congtyABC@gmail.com** thì địa chỉ giả mạo này có thể gần tương tự như **congtyABC1@gmail.com**), chèn Logo chính thức của tổ chức để tăng độ tin cậy, thiết kế các cửa sổ pop-up giống y hệt bản gốc (cả về màu sắc, font chữ,…), sử dụng kĩ thuật giả mạo đường dẫn (link) để lừa người dùng (VD: text là vietcombank.com.vn nhưng khi click vào lại điều hướng tới vietconbank.com.vn). Ngoài ra hình thức giả mạo Website cũng xuất hiện khá phổ biến. Bản chất của việc giả mạo website trong tấn công Phishing chỉ là làm giả một landing page chứ không phải toàn bộ website. Trang được làm giả thường là trang đăng nhập để cướp thông tin của nạn nhân. Kỹ thuật làm giả website có một số đặc điểm sau: Thiết kế giống tới 99% so với website gốc, đường link (url) chỉ khác một ký tự duy nhất. Luôn có những thông điệp khuyến khích người dùng nhập thông tin cá nhân vào website (call-to-action). Hiện nay, các nhà cung cấp dịch vụ email như Google hay Microsoft đều có những bộ lọc email spam/phishing để bảo vệ người dùng. Tuy nhiên những bộ lọc này hoạt động dựa trên việc kiểm tra văn bản (text) trong email để phát hiện xem email đó có phải phishing hay không. Hiểu được điều này, những kẻ tấn công đã cải tiến các hình thức tấn công Phishing lên một tầm cao mới.

Chúng thường sử dụng ảnh hoặc video để truyền tải thông điệp lừa đảo thay vì dùng text như trước đây để vượt quá các bộ lọc này.

- Keylogging: là một phương thức tấn công mà kẻ tấn công sử dụng phần mềm độc hại để ghi lại các ký tự được nhập từ bàn phím của người dùng. Khi người dùng nhập thông tin cá nhân như tên đăng nhập, mật khẩu hoặc thông tin tài khoản ngân hàng, các thông tin này sẽ được lưu lại trong bộ nhớ của máy tính của kẻ tấn công. Từ đó, họ có thể sử dụng các thông tin này để truy cập vào các tài khoản của người dùng và chiếm đoạt tiền trong tài khoản. Tình trạng keylogging chiếm đoạt tài khoản vẫn đang diễn ra trên toàn cầu và đã gây ra nhiều thiệt hại cho các cá nhân và tổ chức. Các kẻ tấn công thường sử dụng phần mềm độc hại như Trojan hoặc spyware để cài đặt keylogger trên máy tính của nạn nhân mà không bị phát hiện. Kỹ thuật đánh cắp tài khoản dùng keylog thường được nhiều hacker sử dụng vì vừa đơn giản vừa hiệu quả. Tùy vào các loại keylogger khác nhau thì nó sẽ có khả năng thu thập thông tin khác nhau, nhưng thường thì các phần mềm theo dõi máy tính này đều có thể khai thác các thông tin như: Ghi lại mật khẩu bạn đã đăng nhập trên thiết bị, Gửi bản báo cáo đã ghi thông qua email đến địa chỉ email, FTP, HTTP, Chụp ảnh màn hình thiết bị với chu kì cố định, Các ứng dụng đang chạy trên thiết bị đều được ghi lại, Chụp các website bạn đã truy cập, ghi lại URL bạn đã vào bằng trình duyệt, Chụp bản sao email bạn đã gửi, Chụp bản ghi màn hình của tất cả tin nhắn từ Zalo, What's app, Facebook Messenger, Viber,… Keylogger ghi lại tất cả thao tác phím và còn chụp màn hình từ thiết bị. Khi đã lấy được thông tin, phần mềm Keylogger có thể lưu trữ dữ liệu trên ổ cứng hoặc chuyển thông tin về một máy được chỉ định trước (hoặc server khác). keylogger được viết ra với chỉ có một loại duy nhất là giúp các bạn giám sát con cái, người thân xem họ làm gì với máy

tính, với internet, khi chat với người lạ nhưng cách sử dụng và chức năng của keylogger hiện tại trên thế giới khiến người ta thường hay phân loại keylogger theo mức độ nguy hiểm phụ thuộc vào những điểm người dùng gặp phải như: nhiễm vào máy không qua cài đặt/Cài đặt vào máy cực nhanh (quick install), Có thuộc tính ẩn/giấu trên trình quản lý tiến trình (process manager) và trình cài đặt và gỡ bỏ chương trình (Add or Remove Program), Có thêm chức năng Capturescreen hoặc ghi lại thao tác chuột, khó gỡ cài đặt, Có khả năng lây nhiễm, chống tắt (kill process). Cứ mỗi câu trả lời "có", cho một điểm. Điểm càng cao, keylogger càng vượt khỏi mục đích giám sát (monitoring) đến với mục đích theo dõi gián điệp (spying) và tính nguy hiểm nó càng cao. Keylogger có thể được phân loại theo số điểm: chạy công khai thông báo cho người bị giám sát đúng với mục đích giám sát, chạy ngầm hướng đến mục đích theo dõi gián điệp hơn là giám sát (nguy hại đến các thông tin cá nhân như là tài khoản cá nhân, mật khẩu, thẻ tín dụng vì người dùng không biết), ẩn giấu hoàn toàn theo dõi trên một phạm vi rộng với mục đích do thám rõ ràng(loại rất nguy hiểm), thường được mang theo bởi các trojan-virus cực kỳ khó tháo gỡ là loại keylogger nguy hiểm nhất. Thông thường, một chương trình keylogger sẽ gồm có ba phần chính. Phần một là chương trình điều khiển (Control Program) dùng để theo điều phối hoạt động, tinh chỉnh các thiết lập, xem các tập tin nhật ký cho keylogger. Phần này là phần được giấu kỹ nhất của keylogger, thông thường chỉ có thể gọi ra bằng một tổ hợp phím tắt đặc biệt. Phần 2 là tập tin hook, hoặc là một chương trình monitor dùng để ghi nhận lại các thao tác bàn phím, capture screen (đây là phần quan trọng nhất). Phần thứ ba là tập tin nhật ký (log), nơi chứa đựng/ghi lại toàn bộ những gì hook ghi nhận được. Các loại keylogger thông thường khi cài đặt vào máy cũng giống như

mọi chương trình máy tính khác, đều phải qua bước cài đặt. Đầu tiên nó sẽ cài đặt các tập tin dùng để hoạt động vào một thư mục đặc biệt (rất phức tạp), sau đó đăng ký cách thức hoạt động rồi đợi người dùng thiết lập thêm các ứng dụng. Sau đó nó bắt đầu hoạt động nhưng đặc biệt với loại keylogger theo virus có thể vào thẳng máy của người dùng bỏ qua bước cài đặt, dùng tính năng autorun để chạy cùng với hệ thống. Một số loại tự thả (drop) vào các chương trình khác, để người dùng sử dụng các chương trình này keylogger sẽ tự động chạy theo.

- Brute force: là một kỹ thuật tấn công mật khẩu mà kẻ tấn công sử dụng các phần mềm độc hại để đoán các mật khẩu bằng cách thử nhiều mật khẩu khác nhau cho đến khi tìm ra mật khẩu đúng. Kỹ thuật này thường được sử dụng để tấn công các tài khoản đăng nhập bằng mật khẩu, bao gồm cả tài khoản ngân hàng, email và các dịch vụ trực tuyến khác.Tình trạng Brute force chiếm đoạt tài khoản vẫn rất phổ biến trên toàn thế giới, và các kẻ tấn công thường sử dụng các danh sách mật khẩu phổ biến để tấn công các tài khoản. Họ có thể sử dụng các công cụ tự động để thử hàng ngàn mật khẩu khác nhau mỗi giây. Vì vậy nó là một cách hiệu quả để đánh cắp tài khoản của người dùng. Mục đích chính của hình thức tấn công Brute Force là để tìm ra mật khẩu và tài khoản có giá trị cao. Các loại Brute Force phổ biến hiện nay là: Simple Brute Force Attack(sử dụng cách tiếp cận có hệ thống để "đoán" username hay password mà không cần dựa vào external logic), Hybrid Brute Force Attack(dựa vào external logic nó có thể xác định các tổ hợp password có khả năng thành công cao nhất kết hợp với Simple Brute Force Attack để thử nhiều tổ hợp nhất có thể), Dictionary Attack(sử dụng một từ điển các xâu hay cụm từ khả thi để đoán username và password của người dùng), Rainbow Table Attack(là một bảng được

tính toán trước để so khớp với kết quả của các hàm hash có thể dùng để đoán một hàm có độ dài xác định và chứa một tập hợp các kí tự cụ thể), Reverse Brute Force Attack (sử dụng một password chung hay một tập hợp các password để thử với nhiều username khả thi nhắm vào một mạng người dùng mà các hacker đã đánh cắp được dữ liệu trước đó), Credential Snuffing(sử dụng các cặp password và username đã biết trước và thử chúng trên nhiều trang web khác nhau vì có không ít người dùng có thói quen sử dụng cùng một cặp password và username trên nhiều hệ thống trang web khác nhau).

- Social engineering: là kết hợp giữa 2 từ Social (xã hội) và Engineering (kỹ thuật), thể hiện bản chất của kiểu tấn công này: các mánh khóe, kỹ thuật tấn công nhắm vào bản tính xã hội của con người, thứ mà không hề tồn tại trong máy móc. Social Engineering Attack còn được biết đến với cái tên Tấn công phi kỹ thuật, nhằm lừa đảo người dùng bằng cách tạo ra một tình huống giả mạo để kích hoạt hành động của người dùng và chiếm đoạt thông tin cá nhân hoặc tiền của họ. Social engineering thường được sử dụng để tấn công các tài khoản ngân hàng, email và các dịch vụ trực tuyến khác. Tình trạng Social engineering chiếm đoạt tài khoản đang ngày càng phổ biến và phức tạp hơn. Các kẻ tấn công sử dụng các chiêu lừa đảo tinh vi để kích hoạt hành động của người dùng, bao gồm gửi email giả mạo từ các tổ chức tài chính hoặc các dịch vụ trực tuyến phổ biến, gọi điện thoại giả mạo từ các tổ chức tài chính hoặc tổ chức chính phủ, hoặc tạo ra các trang web giả mạo để lừa đảo người dùng. Qua đó, kẻ tấn công có thể đạt được các mục đích của mình như xâm nhập vào hệ thống thông qua thông tin được khai thác, truy cập thông tin quan trọng,… mà không cần phải thực hiện những kỹ thuật tấn công quá phức tạp. Có thể thấy được tấn công phi kỹ thuật không

giới hạn hình thức, phương thức, nạn nhân và thủ phạm. Bất kỳ ai đều có thể là tội phạm và bất kỳ ai đều có thể là nạn nhân. Chúng có thể tấn công bằng việc giao tiếp trực tiếp với con người hoặc giao tiếp gián tiếp với con người thông qua các thiết bị kỹ thuật, điện tử và ngày càng sử dụng cách thức tinh vi hơn. Một vài cách thức tấn công phổ biến như: phishing, baiting(là hình thức tấn công phi kỹ thuật thường xảy ra giữa những người có mối liên hệ xã hội, người quen. Khi có được sự tín nhiệm của nạn nhân, kẻ tấn công tiến hành gửi/ mượn usb hoặc các thiết bị công nghệ có chứa mã độc khiến người dùng sử dụng thiết bị đó để đăng nhập vào hệ thống công ty), vishing (là hình thức lừa đảo mạo danh thông qua điện thoại. Kẻ tấn công gọi điện cho con nạn nhân, đóng giả làm một tổ chức hoặc cá nhân uy tín để có được lòng tin của nạn nhân. Bằng cách đó,nạn nhân sẽ không mảy may nghi ngờ và cung cấp cho chúng các thông tin nhạy cảm như số tài khoản ngân hàng, mật khẩu quan trọng…), piggybacking(là hình thức Social Engineering mà kẻ tấn công lừa người có thẩm quyền để đột nhập vào công ty. Trong hình thức này, kẻ tấn công đóng giả là nhân viên chính thức/ người thân/ thợ sửa chữa/ người có thẩm quyền, yêu cầu thông tin quan trọng hoặc các thông tin cần thiết để đăng nhập hệ thống, gắn các thiết bị theo dõi hoặc trực tiếp tấn công hệ thống/ chiếm đoạt tài sản), sử dụng các thiết bị nghe lén và camera để theo dõi các hành vi của đối tượng mà chúng muốn khai thác thông tin.

1.2.5. Gian lận thẻ tín dụng

Gian lận thẻ tín dụng là hành vi lừa đảo nhằm sử dụng trái phép thông tin thẻ tín dụng của người khác để chi tiêu hoặc rút tiền mà không được sự cho phép của chủ sở hữu thẻ. Thông thường, các kẻ gian lận sẽ sao chép thông tin

từ thẻ tín dụng của nạn nhân bằng cách sử dụng các thiết bị đọc thẻ hoặc phần mềm độc hại trên các thiết bị điện tử và sau đó sử dụng thông tin này để mua sắm trực tuyến hoặc tại các cửa hàng.Các hình thức gian lận thẻ tín dụng bao gồm cả việc sao chép thông tin thẻ tín dụng từ các máy ATM hoặc các thiết bị thanh toán điện tử, giả mạo thẻ tín dụng bằng cách tạo ra các thẻ giả mạo hoặc sử dụng các thẻ tín dụng đã bị mất hoặc bị đánh cắp. Ngoài ra, các kẻ lừa đảo còn có thể tìm cách lấy thông tin thẻ tín dụng của nạn nhân thông qua các cuộc gọi điện thoại giả mạo hoặc các email lừa đảo.

Một số cách phổ biến những kẻ lừa đảo có thể có được số thẻ tín dụng:

- Một người phục vụ đánh cắp số thẻ và sử dụng nó. Hacker hay những kẻ lừa đảo sẽ cố gắng đánh cắp thông tin về danh tính bằng cách thu hút người dùng đến một trang web giả mạo nơi người bị lừa cung cấp số thẻ của mình. Kẻ trộm sau đó sử dụng thông tin thẻ tín dụng của bạn thanh toán hoặc rút tiền.
- Sử dụng thẻ tại ATM. Ai đó có thể đã cài đặt skimmer thẻ tín dụng để đánh cắp thông tin tài khoản của người dùng. Skimmer thẻ tín dụng là một thiết bị nhỏ mà kẻ trộm có thể cài đặt ở bất cứ nơi nào người sử dụng có thể quẹt thẻ. Skimming đã được chứng minh là một cách hiệu quả để những tên trộm đánh cắp thông tin thẻ tín dụng.
- Đôi khi thông tin thẻ tín dụng của người dùng bị đánh cắp không có lỗi của họ. Số thẻ tín dụng của người dùng có thể bị đánh cắp do dữ liệu của các đơn vị thanh toán mà người dùng hay sử dụng bị đánh cắp mua bán hoặc chia sẻ. Kẻ xấu sau đó có thể sử dụng thông tin này để trả các khoản phí trực tuyến với số tài khoản thẻ tín dụng đã chiếm đoạt được.
- Kẻ tấn công thường mua số thẻ tín dụng bị đánh cắp trên web đen, một phần của web mà chỉ truy cập thông qua phần mềm đặc biệt. Số thẻ

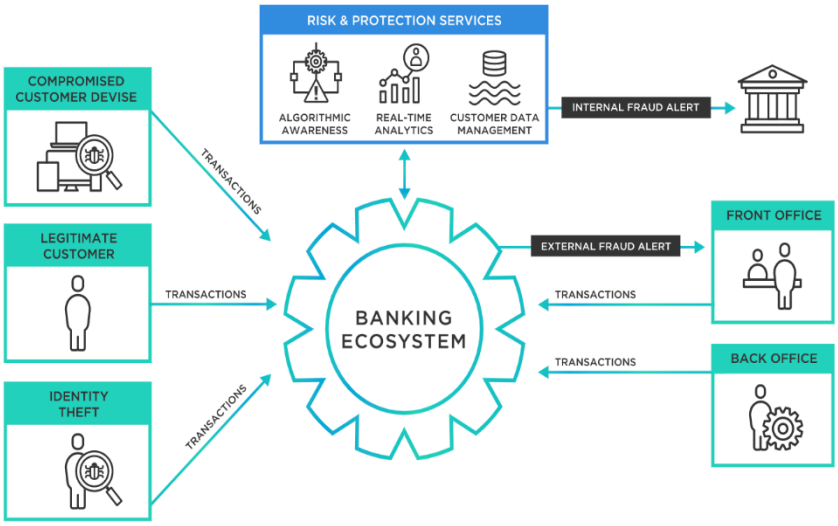
tín dụng có giá trị đối với kẻ tấn công và chúng sử dụng các web đen bất hợp pháp để có che giấu hành vi phạm tội và rửa tiền sau khi chiếm đoạt được tiền từ thẻ tín dụng.

- Thông tin bị chiếm đoạt do người quen của người sử dụng có thể truy cập hoặc kỹ thuật viên dịch vụ - có thể quản lý để truy cập thẻ tín dụng của khách hàng và sử dụng thông tin thẻ tín dụng của người dùng một cách bất hợp pháp mà chưa có sự đồng ý từ chủ sở hữu.

1.3. Hệ thống gian lận tài chính

1.3.1. Tổng quan hệ thống

Hệ thống phát hiện bất thường trên hệ sinh thái ngân hàng dựa trên nền tảng là các service chấm điểm rủi ro, phân tích dữ liệu từ thông tin giao dịch của khách hàng để đưa ra cảnh báo về giao dịch đáng ngờ, phát hiện tài khoản bị đánh cắp (Hình 1.1).



Hình 1. 1. Hệ thống phát hiện giao dịch bất thường của ngân hàng

Việc phát hiện nhầm giao dịch gian lận xảy ra khi hệ thống phát hiện gian lận đọc sai các giao dịch thực và gắn cờ chúng là gian lận, giao dịch bị từ

chối. Điều này có thể dẫn đến làm tổn hại đến mối quan hệ giữa khách hàng chủ tài khoản và ngân hàng. Có thể dẫn đến việc người bán bị mất doanh số do giao dịch bị từ chối . Nếu hệ thống không được hiệu chỉnh để giảm thiểu thông tin sai lệch, ngân hàng có nguy cơ mất khách hàng khi phân loại sai các giao dịch hợp pháp là gian lận. Nếu ngân hàng hủy thẻ tín dụng trong trường hợp như vậy, thì ngân hàng phải tự trả chi phí hoạt động như in thẻ mới và gửi chúng cho khách hàng. Điều này có thể dẫn đến mất lòng tin và gia tăng sự rời bỏ của khách hàng. Do đó, các ngân hàng phải càng chính xác càng tốt trong việc phân biệt giữa giao dịch thật và giao dịch gian lận.

1.3.2. Hệ thống phát hiện gian lận của Alipay

Theo số liệu thống kê năm 2017, số lượng và khối lượng giao dịch trực tuyến lần lượt đạt 48 tỷ giao dịch và 2,075 nghìn tỷ nhân dân tệ [1]. Công ty con Ant Financial hay còn được gọi là Alipay, chiếm khoảng 58% Giao dịch thanh toán trực tuyến bên thứ ba của Trung Quốc. Để thu thập và phân tích số lượng giao dịch như vậy yêu cầu một cơ sở dữ liệu mạnh mẽ để lưu trữ và quản lý. Hơn nữa, yêu cầu hệ thống tính toán phân tán quy mô lớn để chạy các thuật toán. Để đáp ứng các yêu cầu về độ trễ thấp cho phục vụ trực tuyến, dự đoán trực tuyến với truy cập dữ liệu hiệu quả có ý nghĩa rất quan trọng. Các phương pháp dựa trên quy luật đã được nghiên cứu rộng rãi cho vấn đề phát hiện gian lận. Tuy nhiên, các cách thức gian lận thay đổi nhanh chóng theo thời gian, làm giảm đáng kể độ chính xác của việc sử dụng luật. Sau đó, nhiều phương pháp dựa trên khai thác dữ liệu đã được nghiên cứu.

Dữ liệu giao dịch thường có hai đặc điểm:

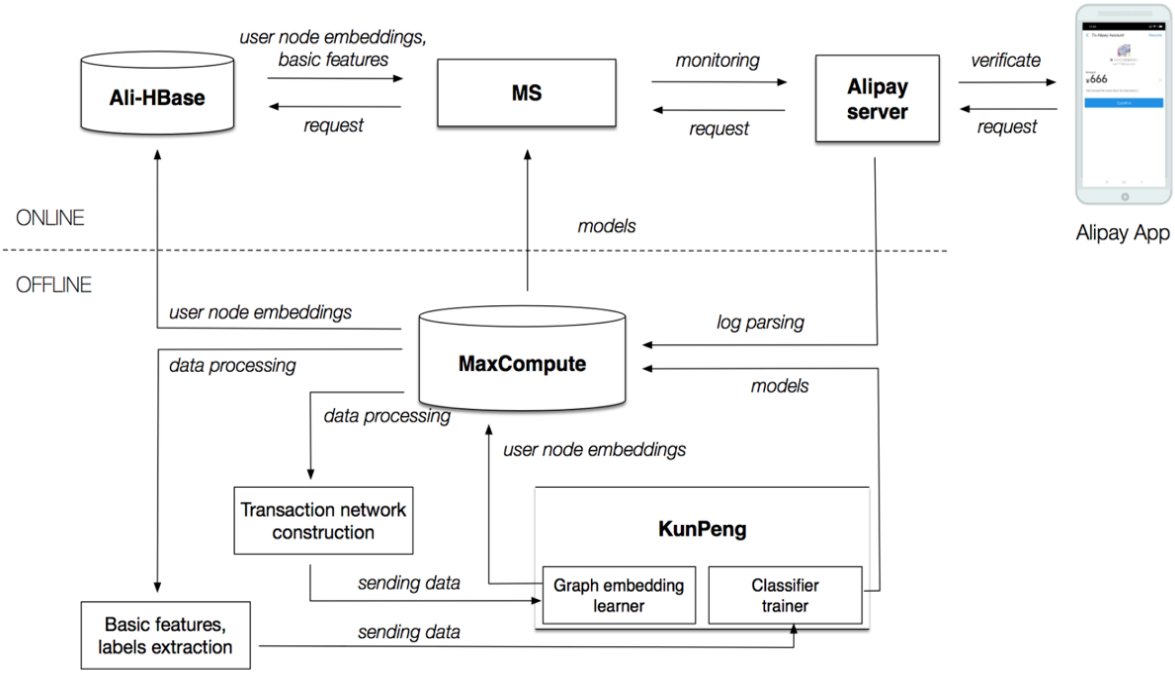
- Mất cân bằng nhãn tức là phần lớn các giao dịch là giao dịch bình thường chỉ số lượng nhỏ là giao dịch gian lận.

- So với phân tích cá nhân hồ sơ giao dịch, dữ liệu tổng hợp thường cung cấp nhiều thông tin phong phú hơn để xác định các mô hình gian lận.

Để giải quyết vấn đề này có nhiều hướng giải quyết như một số phương pháp học không giám sát được đề xuất hoặc một số chiến lược tổng hợp dữ liệu hiện có được áp dụng để phát hiện gian lận tuy nhiên hầu hết các phương pháp trước đây khó có thể nắm bắt được sự phức tạp cách thức gian lận của các giao dịch trực tuyến. Các gian lận giao dịch trực tuyến có thể được phân loại thành hai loại khác nhau là: rõ ràng trong việc nhận biết các hành vi bất thường và ngược lại là không rõ ràng. Trường hợp, người dùng nhận thức được gian lận sau khi giao dịch hoàn tất có thể gửi báo cáo gian lận và tải lên các bằng chứng về việc đó. Dựa trên các chi tiết giao dịch, hồ sơ và bằng chứng, tính xác thực của giao dịch gian lận sẽ bị xác thực. Nếu người dùng này thực sự gian lận, những kẻ gian lận sẽ bị xử lý bằng việc đối mặt với luật pháp hoặc các chế tài xử lý của tổ chức, chẳng hạn như hạn chế hành động hoặc khóa tài khoản nhưng nó có thể thu hồi thiệt hại theo quy định của pháp luật. Đó là trường hợp loại hình gian lận rõ ràng, còn trong một trường giao dịch gian lận không rõ ràng, điều chúng ta quan tâm là lấy hành động chủ động để ngăn chặn sự kiện gian lận tiềm ẩn giao dịch, tức là chủ động phát hiện gian lận giao dịch trực tuyến và thực hiện các bước ngay lập tức để ngăn chặn các giao dịch đáng ngờ. Trái ngược với gian lận rõ ràng, giao dịch gian lận không rõ ràng tiết lộ ít hơn thông tin và yêu cầu dự đoán thời gian thực của hệ thống. Hệ thống Alipay có khoảng 50 đặc trưng được thiết kế. Các đặc trưng đó là các đặc trưng cơ bản cũng được coi là quy tắc hoặc thuộc tính. Đối với mỗi người dùng sẽ có các đặc trưng tổng hợp dưới dạng thông tin bổ sung từ các bản ghi giao dịch tổng hợp. Các đặc trưng cơ bản và các đặc trưng tổng hợp sau đó được nối với nhau. Các nhãn được thu thập từ các báo cáo gian lận của người dùng do đó không thể lấy được

trong thời gian thực. Để tìm ra gian lận một cách chính xác cần điều tra rộng rãi và xác thực các phương pháp dựa trên quy tắc, phương pháp phát hiện bất thường và mô hình phân loại.

Các phương pháp dựa trên quy tắc được sử dụng rộng rãi trong nhiều ứng dụng phát hiện gian lận. Trong đó Iterative Dichotomiser là một cách tiếp cận truyền thống dựa trên học cây quyết định trong khi là phiên bản sửa đổi để trích xuất các mẫu thông tin từ dữ liệu với độ chính xác cao hơn. Các đặc trưng được coi như quy tắc và thông tin nhãn được sử dụng để tinh chỉnh. Còn Isolation Forest (IF) là một công cụ phát hiện bất thường cổ điển, trong đó phương pháp này được sử dụng rộng rãi do tính hiệu quả của nó. Chúng ta coi các đặc trưng là thuộc tính và dự đoán trực tiếp các giao dịch gian lận vì nó không yêu cầu bất kỳ thông tin nhãn nào. Phát hiện gian lận giao dịch tương tự như việc tìm ra các giao dịch bất thường, tức là tìm ra các ngoại lệ có những đặc điểm rất khác so với các dữ liệu bình thường. Công ty Alipay sử dụng hệ thống phát hiện gian lận TitAnt có kiến trúc như hình 1.2.



Hình 1. 2. Hệ thống phát hiện giao dịch bất thường của Alipay(TitAnt)

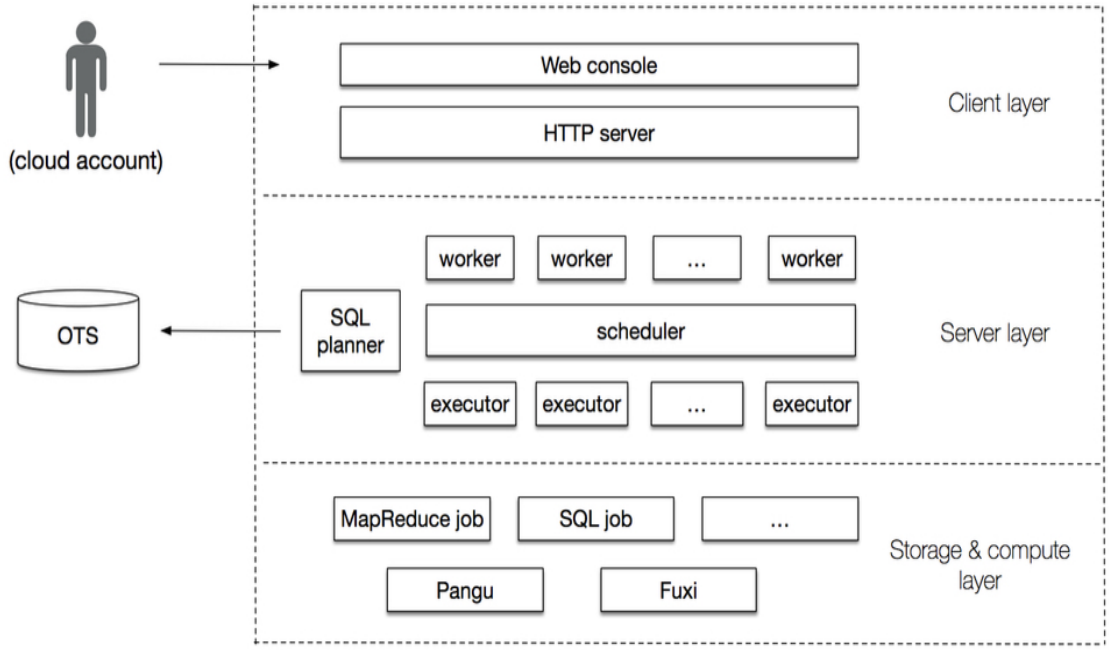
Hệ thống TitAnt

Để đảm bảo phản hồi kịp thời đối với các yêu cầu phát hiện gian lận, công cụ dự đoán độ trễ thấp, nền tảng lưu trữ cơ sở dữ liệu mạnh mẽ và các thuật toán phân tán phải được thiết kế cẩn thận. Việc huấn luyện AI ngoại tuyến nơi các mô hình được đào tạo trên cơ sở thời gian cố định và các tệp mô hình được tải lên công cụ dự đoán trực tuyến để theo dõi giao dịch theo thời gian thực. Sau khi người dùng bắt đầu yêu cầu giao dịch trong Alipay, nhật ký giao dịch sẽ được gửi định kỳ tới MaxCompute cho tính toán. MaxCompute hỗ trợ SQL và MapReduce để trích xuất các tính năng/nhãn cơ bản và xây dựng mạng giao dịch. Đồng thời KunPeng hỗ trợ huấn luyện mô hình phân loại phân tán quy mô lớn. Các mô hình phân loại và node embedding người dùng đã học sẽ được lưu trữ trong MaxCompute. Dự đoán trực tuyến được thực thi tại Model Server (MS) nơi các tệp mô hình được cập nhật định kỳ. Khi một giao dịch được tạo bởi người dùng trong ứng dụng Alipay, máy chủ Alipay ngay lập tức gửi yêu cầu tới máy chủ Model Server (MS), MS sau đó lấy dữ liệu liên quan từ Ali-HBase và đưa ra dự đoán theo thời gian thực. Nếu giao dịch bị phát hiện là gian lận giao dịch đang diễn ra sẽ bị gián đoạn và người chuyển tiền sẽ được thông báo. Các thành phần trong hệ thống TitAnt sẽ được trình bày chi tiết như sau:

MaxCompute

MaxCompute dùng để quản lý tác vụ tính toán cho hệ thống TitAnt. Trước đây MaxCompute được gọi là dịch vụ xử lý dữ liệu mở, một nền tảng quản lý và lưu trữ cơ sở dữ liệu. Nó có ba lớp logic: lớp máy khách, lớp máy chủ và lớp lưu trữ & tính toán. Nhà phát triển có thể đăng nhập bằng tài khoản

cloud của họ và gửi công việc bằng bảng điều khiển web ở lớp máy khách, nơi máy chủ HTTP nhận lệnh và gửi thông báo đến lớp tiếp theo.

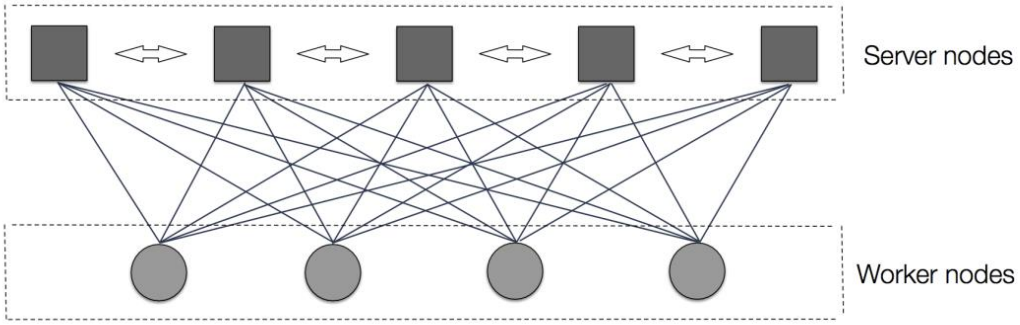


Hình 1. 3. Kiến trúc MaxCompute

Lớp máy chủ (Server layer) bao gồm worker (chứa các nhiệm vụ cần thực thi), executor (làm nhiệm vụ thực thi lệnh) và Scheduler (chứa lịch theo dõi executor). Ngoài ra, còn có các xử lý bất đồng nhất, chẳng hạn như mapreduce, SQL, v.v., có thể được nhận dạng và vận hành trong lớp lưu trữ và tính toán dựa trên Pangu và Fuxi, trong đó Pangu là mô-đun lưu trữ và Fuxi là mô-đun lập lịch tài nguyên. Khi một lệnh SQL được gửi bởi bảng điều khiển web thông báo sẽ được gửi đến máy chủ HTTP yêu cầu xác minh thông tin tài khoản cloud. Nếu xác thực thành công, công việc sẽ được giao cho worker và các công việc tương ứng sẽ được gửi đến bộ lập lịch. Sau đó bộ lập lịch đăng ký phiên bản trong Open Table Service (OTS) thông qua công cụ lập kế hoạch SQL và trạng thái của nó được đặt đồng thời là "đang chạy". OTS duy trì trạng thái của tất cả các phiên bản. Cuối cùng, bộ lập lịch thêm phiên bản vào hàng đợi và ID phiên bản tương ứng sẽ được tạo. Sau đó, bộ lập lịch sẽ chia nhiệm

vụ của thể hiện công việc thành nhiều nhiệm vụ con, các nhiệm vụ này được sắp xếp thành nhiệm vụ nhóm theo thứ tự ưu tiên. Sau đó, lịch trình tiếp tục chờ đợi cho các tài nguyên có sẵn cho máy tính. Ngay sau khi điều kiện tài nguyên được thỏa mãn các nhiệm vụ con được gửi thực thi yêu cầu Fuxi kích hoạt tài nguyên máy tính trong lớp tính toán. Khi tất cả các nhiệm vụ con được kết thúc, người thi hành cập nhật trạng thái riêng biệt như "chấm dứt" trong OTS.

KunPeng



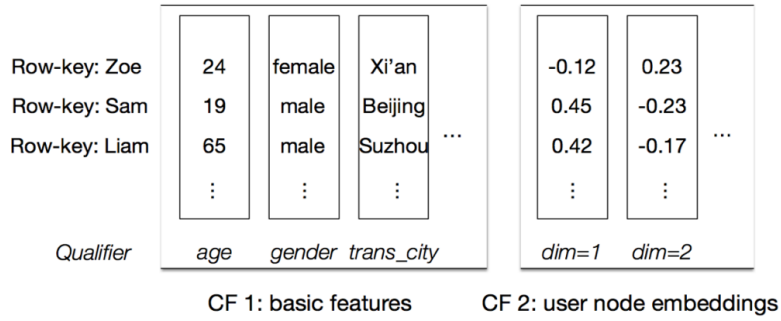
Hình 1. 4. Kiến trúc hệ thống của KunPeng

Trên thực tế có rất nhiều giao dịch cần phân tích mỗi ngày, do đó đòi hỏi có một nền tảng tính toán phân tán . Các tiêu chuẩn truyền thống như MPI không có khả năng chịu lỗi tốt. Máy chủ hỗ trợ lỗi trong trường hợp lỗi có thể được tự động khởi động lại và khôi phục về trạng thái trước đó trong khi các phiên bản khác không bị ảnh hưởng. Cân bằng hệ thống phát triển dựa trên Máy chủ tham số framework nơi có nhiều thuật toán học máy khác nhau chạy đồng thời. KunPeng hỗ trợ song song thực hiện cả dữ liệu và mô hình.

Như minh họa trong hình 1.4 gồm các nút đại diện máy chủ lưu trữ các tham số mô hình trong khi các nút chứa nhiệm vụ (worker) chịu trách nhiệm đào tạo. Dựa trên KunPeng ta thiết kế lại NLR(Nonlinear Logistic Regression) và phân loại thuật toán. Chẳng hạn như DW(Deep Walk), S2V(Structure2Vec),

LR(Logistic Regression) và GBDT(Gradient Boosting Decision Tree). Từ một node nhận được chuỗi đương đi node đó đến các node khác bằng thuật toán Random walk. Đối với mỗi lần lặp lại, trước tiên mỗi node đọc một loạt dữ liệu trình tự và tạo danh sách các node. Các nút embedding sau đó được kéo từ các nút máy chủ và được cập nhật theo độ dốc của hàm tối ưu gốc. Sau đó chúng được embedding và cập nhật được tải lên máy chủ. Mặt khác, các nút máy chủ chịu trách nhiệm cho giao tiếp với các nút nhiệm vụ con để trao đổi dữ liệu embedding. Đầu tiên, các node máy chủ khởi tạo ngẫu nhiên các phần embedding và chờ các yêu cầu đẩy từ các node nhiệm vụ. Sau khi nhận được yêu cầu đẩy các thông tin tương ứng sẽ được gửi. Sau khi cập nhật từng nút các máy chủ sẽ nhận các phần nút embedding mới và tổng hợp chúng bằng cách thực thi hoạt động trung bình của mô hình.

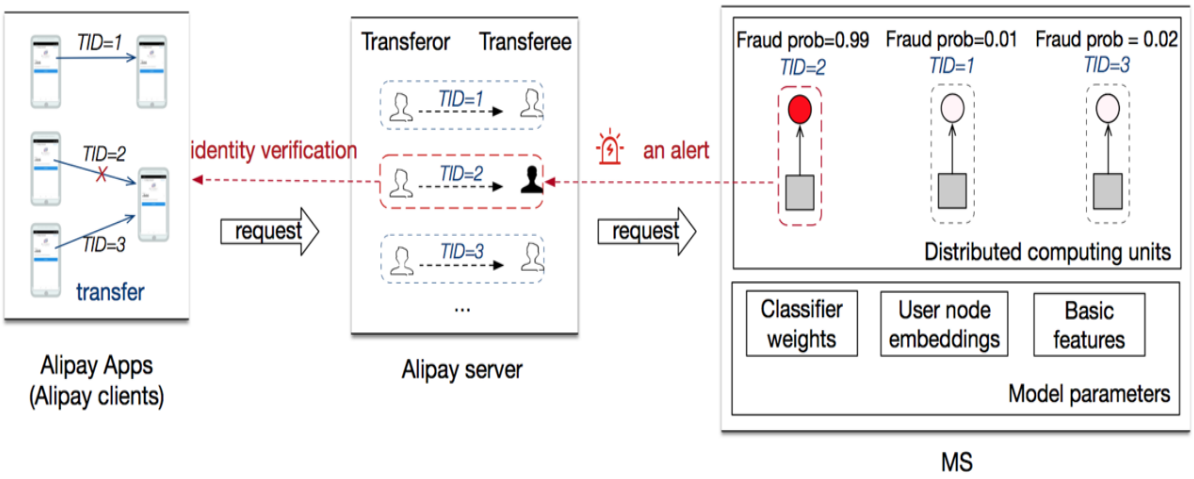
MS and Ali-HBase



Hình 1. 5. Kiến trúc hệ thống của Ali-HBase

Khi phần huấn luyện ngoại tuyến kết thúc, dự đoán thời gian thực trực tuyến sẽ hoạt động. Hình 1.5 cho thấy một ví dụ minh họa về toàn bộ quá trình dự đoán thời gian thực. Khi người dùng chuyển tiền trong ứng dụng Alipay yêu cầu chuyển khoản được gửi đến máy chủ Alipay sau đó MS giám sát gian lận. MS được phân phối để đáp ứng độ trễ thấp và dịch vụ tải tốc độ cao. Như thể hiện trong Hình 1.5, giao dịch TID=2 có thể là gian lận với xác suất gian lận dự đoán là 99%, do đó MS gửi cảnh báo đến máy chủ Alipay điều này sẽ tiếp

tục làm gián đoạn giao dịch đang diễn ra tương ứng. Ali-HBase dựa trên HBase. HBase đầu tiên được đề xuất là Bigtable một giải pháp phân tán có thể mở rộng kho dữ liệu, phù hợp với việc truy cập dữ liệu thời gian thực. Như được hiển thị trong hình 1.5, Column Family (CF) đầu tiên là các thông tin cơ bản bao gồm tuổi, giới tính và thành phố là đặc trưng cơ bản. Column Family (CF) tiếp theo là thông tin người dùng sau khi đã được số hóa thành node embedding.



Hình 1. 6. Kiến trúc hệ thống của MS và sự tương tác với các thành phần khác

Trong hình 1.6, người dùng như Zoe, Sam và Liam là các hàng để lập chỉ mục dữ liệu tương ứng. Mỗi lần huấn luyện mô hình ngoại tuyến hoàn thành dữ liệu được tải lên Ali-HBase theo phiên bản được định dạng bởi thời gian là ngày giờ.

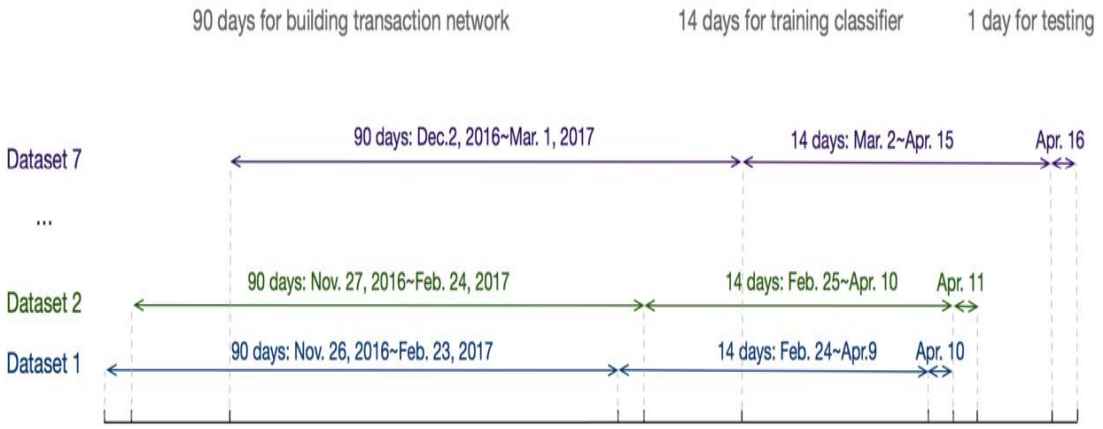
Các vấn đề triển khai và xây dựng mạng lưới giao dịch

Đầu tiên, hệ thống có các yêu cầu phục vụ khắt khe như thời gian ngắn chỉ tính bằng mili giây để phát hiện trực tuyến bao gồm cả chi phí tính toán và liên lạc. Tuy nhiên, việc đánh nhãn thường bị trễ vì chúng được thu thập thông qua phản hồi của người dùng. Trong đó huấn luyện mô hình trực tuyến là không thực tế. Vì vậy áp dụng định kỳ đào tạo ngoại tuyến và dự đoán thời gian thực trong hệ thống là phương án hợp lý được đề xuất.

Thứ hai,trong hệ thống chỉ chứng minh tính hữu ích của việc embedding nút người dùng học được từ giao dịch mạng. Ta có thể có được thông tin tổng hợp khác, chẳng hạn như thông tin về thiết bị và IP. Đó là một câu hỏi thú vị để xây dựng một mạng không đồng nhất. Để trả lời câu hỏi như vậy từ dữ liệu người dùng sẽ được bổ sung để đưa ra những hướng đi trong tương lai.

Thiết lập thử nghiệm

Hệ thống áp dụng chế độ "T+1" để cập nhật mô hình, điều đó có nghĩa là một mô hình sẽ được đào tạo và triển khai theo cách ngoại tuyến hàng ngày và sẽ được sử dụng để dự đoán cho ngày hôm sau trên cơ sở thời gian thực. Để chứng minh hiệu quả của hệ thống. Alipay đã tiến hành một số thử nghiệm và báo cáo hiệu suất của mỗi ngày trong một tuần liên tục. Tổng cộng có bảy bộ dữ liệu. Trong đó mỗi cái được cắt thành ba tập hợp con: mỗi người đại diện cho 1 node, một node khác để huấn luyện mô hình phân loại và cuối cùng để thử nghiệm. Thu thập hồ sơ giao dịch trong 90 ngày để xây dựng mạng lưới giao dịch. 14 ngày tiếp theo, các bản ghi được dán nhãn được coi là tập huấn luyện và ngày cuối cùng của các bản ghi được dán nhãn được sử dụng cho bộ kiểm tra.



Hình 1. 7. Chia dataset huấn luyện và thử nghiệm mô hình dự báo

Ví dụ: trong Bộ dữ liệu 1 (được minh họa trong Hình 1.7), các bản ghi giao dịch của ngày 10 tháng 4 năm 2017 được chọn làm bộ kiểm tra, 14 hồ sơ của ngày trước tập kiểm tra được sử dụng như đào tạo và 90 ngày trước đó của hồ sơ được sử dụng để xây dựng mạng lưới giao dịch. Khác với các lĩnh vực khác chẳng hạn như thương mại điện tử. Thử nghiệm trực tuyến khó để đạt được vì các nhãn không thu được theo thời gian thực. Trong thí nghiệm một trong những mục tiêu là điều tra hiệu quả của các tính năng cơ bản và nút người dùng đã học dựa trên mạng giao dịch. So sánh unsupervised DeepWalk và supervised Structure2Vec, các mô hình xử lý dữ liệu không cân bằng nhãn. Để so sánh khách quan các mô hình, kích thước của các embedding đã huấn luyện được đặt thành 32 và được nối với các đặc trưng cơ bản. Ngoài ra, đối với các phương pháp phát hiện, kiểm tra tính hợp lệ của ID3 dựa trên quy tắc và C5.0, Isolation Forest dựa trên phát hiện bất thường và dựa trên phân loại Logistic Regression và Gradient Boosting Decision Tree. Đối với DeepWalk, đặt độ dài của bước đi ngẫu nhiên là 50, trong đó mỗi nút được lấy mẫu làm nút đầu tiên của chuỗi 100 lần, tức là số lần lấy mẫu là 100. Mất khoảng 1,5 giờ để embedding và xử lý với khoảng 8 triệu bản ghi giao dịch được chọn ngẫu nhiên với 20 máy có 10 luồng. Ngoài mạng giao dịch cũng cung cấp cho Structure2Vec với sự thật gian lận có căn cứ như nhãn cạnh. Ngoài ra còn có tổng cộng 52 tính năng cơ bản được trích xuất cẩn thận. Alipay đặt 100 trường hợp cho Isolation Forest và các tính năng cơ bản thô được cung cấp như thuộc tính. Vì ID3 và C5.0 dựa trên quy tắc không thể hỗ trợ các giá trị liên tục tốt và rời rạc hóa dữ liệu thành các giá trị khác nhau thường sử dụng chuẩn hóa L1 và gán trọng số của nó là 0,1 cho Logistic Regression và đặt 300 lần lặp làm tiêu chí dừng huấn luyện. Đối với Gradient Boosting Decision Tree, Alipay tạo 400 trees với độ sâu từ 3 để ensemble kết quả với mean square error. Tốc độ lấy

mẫu con của các mẫu và đặc trưng được đặt là 0,4 để tránh tình trạng overfitting khi huấn luyện.

1.4. Kết luận chương

Nội dung chương này em đã trình bày tổng quan về các hành vi gian lận tài chính trực tuyến. Thông qua các hành vi gian lận và đặc điểm các hành vi gian lận phổ biến ta hiểu được những nguyên nhân và cách thức thực hiện các hành vi đó. Với xu hướng về chuyển đổi số của thế giới nói chúng và Việt Nam nói riêng đã tạo điều kiện thuận lợi để các công nghệ mới xuất hiện mang đến nhiều sự tiện lợi cho người sử dụng và các cơ quan tổ chức quản lý. Nhưng bên cạnh đó cũng tiềm ẩn nhiều nguy cơ, rủi ro về các hình thức tấn công, lừa đảo cũng ngày càng gia tăng. Thực tế các hành vi này ngày càng gia tăng và trở nên tinh vi khó kiểm soát hơn. Do đó vấn đề đảm bảo an toàn cho các giao dịch tài chính trở nên vô cùng cấp thiết để đảm bảo niềm tin, tài sản và giúp phát triển chuyển đổi số trở nên hiệu quả và an toàn hơn. Ngoài ra trong chương này đã đi giới thiệu về cấu phần, cách thức hoạt động của một hệ thống phát hiện gian lận tài chính trực tuyến phổ biến trên thực tế. Thực trạng hoạt động của các hệ thống phát hiện bất thường và các hành vi gian lận tài chính trước đây. Các phương thức phát hiện gian lận đang được sử dụng tại các hệ thống phát hiện gian lận tài chính đang được sử dụng tại một số công ty. Khả năng đáp ứng nhu cầu trong xu thế chuyển đổi số mới của xã hội về mặt an toàn, bảo mật và mặt trải nghiệm của người sử dụng.

Chương tiếp theo em sẽ trình bày các nội dung liên quan tới các kỹ thuật trong AI (Artificial Intelligent) được sử dụng trong hệ thống phát hiện gian lận tài chính, các ưu điểm nhược điểm của cũng như lý do lựa chọn kỹ thuật này trong nghiên cứu. Lý do em lựa chọn thuật toán của mình để tích hợp vào hệ

thống phát hiện gian lận tài chính. Điểm mạnh của thuật toán của em sử dụng và cơ sở lý thuyết của thuật toán.

CHƯƠNG 2: ỨNG DỤNG THUẬT TOÁN TRÍ TUỆ NHÂN TẠO PHÁT HIỆN GIAN LẬN TÀI CHÍNH

2.1. Giới thiệu các thuật toán trí tuệ nhân tạo

Như chúng ta đã biết, các hệ thống bảo mật trong hệ thống tài chính ngày càng cần phải xử lý các giao dịch gian lận nhanh hơn và chính xác hơn. Một trong các phương pháp được áp dụng phổ biến trong các hệ thống này hiện nay là các ứng dụng của thuật toán học máy. Vì vậy, trong chương này em sẽ đi giới thiệu chi tiết hơn về các thuật toán học máy theo 4 nhóm cơ bản: Supervised learning, Unsupervised learning, Semi-supervised learning và Reinforcement learning.

2.1.1. Supervised learning (học có giám sát)

Supervised learning là thuật toán dự đoán nhãn đầu ra (output) của một dữ liệu mới (new input) dựa trên các cặp (input, output) đã biết từ trước. Cặp dữ liệu này còn được gọi là (data, label), tức (dữ liệu, nhãn). Supervised learning là nhóm phổ biến nhất trong các thuật toán Machine Learning. Về mặt toán học, Supervised learning là khi chúng ra có một tập hợp biến đầu vào $X=\{x_1,x_2,...,x_N\}$ và tập hợp các nhãn tương ứng với các biến đó $Y=\{y_1,y_2,...,y_N\}$, trong đó x_i, y_i là các vector. Các cặp dữ liệu biết trước $(x_i,y_i) \in X \times Y$ được gọi là tập training data (dữ liệu huấn luyện). Từ tập training data này, chúng ta cần tạo ra một hàm số ánh xạ mỗi phần tử từ tập X sang một phần tử (xấp xỉ) tương ứng của tập Y: $y_i \approx f(x_i)$, $\forall_i=_{1,2,...,N}$ Mục đích là xấp xỉ hàm số f thật tốt để khi có một dữ liệu x mới, chúng ta có thể tính được nhãn tương ứng của nó $y=f(x)$.

Học máy có giám sát trong phát hiện gian lận tài chính là một phương pháp sử dụng các thuật toán học máy để phát hiện các hoạt động gian lận tài

chính dựa trên các dữ liệu tài chính được gắn nhãn. Mô hình học máy được huấn luyện trên một tập dữ liệu tài chính đã được gắn nhãn, với các đặc trưng của giao dịch tài chính như tỷ lệ nợ tín dụng, tỷ suất lợi nhuận, thu nhập và các giao dịch tài chính khác. Việc sử dụng học máy có giám sát trong phát hiện gian lận tài chính có nhiều lợi ích. Nó cho phép tổ chức tài chính và ngân hàng phát hiện các hoạt động gian lận tài chính nhanh chóng và hiệu quả hơn, giúp giảm thiểu rủi ro và tổn thất tài chính.

Ưu điểm khi sử dụng học máy có giám sát trong phát hiện gian lận tài chính, bao gồm:

- Tăng tính hiệu quả và độ chính xác: Phương pháp này sử dụng các thuật toán học máy để tự động phát hiện gian lận tài chính, giúp tăng tính hiệu quả và độ chính xác trong quá trình phát hiện so với các phương pháp thủ công truyền thống.
- Phát hiện gian lận nhanh chóng: Học máy có giám sát giúp phát hiện gian lận tài chính nhanh chóng và hiệu quả hơn, giúp tổ chức tài chính và ngân hàng giảm thiểu rủi ro và tổn thất tài chính.
- Tính linh hoạt: Phương pháp này có thể được áp dụng trên nhiều loại dữ liệu khác nhau, bao gồm dữ liệu tài chính và dữ liệu giao dịch, giúp tăng tính linh hoạt trong việc phát hiện gian lận tài chính.
- Khả năng mở rộng: Học máy có giám sát có thể được mở rộng để bao gồm nhiều loại hoạt động gian lận khác nhau, giúp tổ chức tài chính và ngân hàng nâng cao khả năng phát hiện các hoạt động gian lận.
- Độ tin cậy cao: Phương pháp này giúp nâng cao độ tin cậy trong việc phát hiện gian lận tài chính, đồng thời giảm thiểu các sai sót và mức độ nhầm lẫn trong quá trình phát hiện.

Mặc dù học máy có giám sát là một phương pháp hiệu quả trong phát hiện gian lận tài chính, nhưng nó cũng có một số nhược điểm sau:

- Phụ thuộc vào chất lượng dữ liệu: Để đảm bảo tính chính xác và hiệu quả của phương pháp này cần có dữ liệu đầy đủ và chất lượng cao. Nếu dữ liệu không đầy đủ hoặc bị nhiễu sẽ ảnh hưởng đến kết quả phân loại và dẫn đến sai sót.
- Phải có sự can thiệp của con người: Mặc dù học máy có giám sát được sử dụng để tự động phát hiện gian lận nhưng việc xác định các tập dữ liệu huấn luyện và chọn thuật toán và thiết lập ngưỡng phân loại đòi hỏi sự can thiệp của con người. Việc can thiệp này đòi hỏi sự am hiểu và kinh nghiệm trong lĩnh vực này để đạt được kết quả tối ưu.
- Rủi ro của quyết định sai: Phương pháp này có thể dẫn đến những quyết định sai lầm nếu các thuật toán học máy không được cấu hình đúng hoặc không được huấn luyện đầy đủ. Khi đó, các giao dịch hợp lệ có thể bị coi là gian lận hoặc các giao dịch gian lận có thể bị xác định là hợp lệ, dẫn đến rủi ro cho tổ chức tài chính và ngân hàng.
- Chi phí và thời gian: Việc triển khai và cấu hình các thuật toán học máy có giám sát đòi hỏi chi phí và thời gian đầu tư khá lớn. Cần phải có sự đầu tư về hạ tầng, công nghệ, đội ngũ chuyên gia, phân tích dữ liệu và huấn luyện mô hình để đạt được hiệu quả tối ưu trong việc phát hiện gian lận tài chính.

Vì vậy, các thuật toán học máy có giám sát muốn phát huy hiệu quả của mình cũng cần được đánh giá và xử lý một cách hợp lý.

Các bước xây dựng học máy có giám sát trong phát hiện gian lận tài chính bao gồm:

- Xác định mục tiêu và phạm vi của dự án: Trước khi triển khai học máy có giám sát, cần xác định rõ mục tiêu và phạm vi của dự án, đồng thời xác định các nguồn dữ liệu và bộ dữ liệu cần sử dụng để huấn luyện và đánh giá mô hình.

- Chuẩn bị dữ liệu: Sau khi đã xác định các nguồn dữ liệu, cần tiến hành thu thập, xử lý và chuẩn bị dữ liệu trước khi huấn luyện mô hình. Các bước này bao gồm: tiền xử lý dữ liệu, loại bỏ nhiễu, chuẩn hóa dữ liệu, chọn các tính năng quan trọng, tạo các biến phụ thuộc và biến độc lập, và phân chia dữ liệu thành các tập huấn luyện và kiểm tra.
- Chọn và huấn luyện mô hình: Sau khi đã chuẩn bị dữ liệu, cần chọn thuật toán học máy phù hợp để huấn luyện mô hình. Các thuật toán phổ biến bao gồm: Random Forest, Logistic Regression, Decision Tree, Neural Networks, và Support Vector Machines. Sau khi huấn luyện môn hình trên bộ dữ liệu đã chuẩn bị chúng ta tiến hành và đánh giá hiệu suất của mô hình trên tập kiểm tra để đánh giá độ chính xác của mô hình.
- Kiểm tra và cải thiện mô hình: Dựa trên kết quả việc đánh giá mô hình, chúng ta tiến hành các bước cải thiện mô hình bao gồm: đánh giá hiệu suất của mô hình, phân tích các sai sót của mô hình, tinh chỉnh các tham số của mô hình để cải thiện độ chính xác của mô hình.
- Triển khai và giám sát mô hình: Sau khi đã đánh giá và cải thiện mô hình cần triển khai mô hình trên hệ thống sản xuất và giám sát mô hình để đảm bảo rằng mô hình hoạt động tốt và phát hiện được gian lận tài chính. Các bước này bao gồm: tích hợp mô hình vào hệ thống, đánh giá hiệu suất của mô hình trên dữ liệu thực tế và cập nhật mô hình để đảm bảo tính chính xác và khách quan của mô hình.

Học máy có giám sát có 2 dạng bài toán chính là :

- Classification là bài toán đánh nhãn của dữ liệu đầu vào chia thành một số hữu hạn nhóm. Ví dụ: Gmail xác định xem một email có phải là spam hay không; các hãng tín dụng xác định xem một khách hàng có khả năng thanh toán nợ hay không. Mô hình Classification trong

phát hiện gian lận tài chính là một mô hình học máy có giám sát được sử dụng để phân loại các giao dịch tài chính thành hai loại: giao dịch hợp lệ và giao dịch gian lận. Mô hình sử dụng các đặc trưng của các giao dịch, như số tiền, địa chỉ IP, quốc gia, thời gian giao dịch, loại thẻ và các thông tin khác để dự đoán xác suất của một giao dịch là gian lận.

- Regression là một mô hình học máy có giám sát được sử dụng để dự đoán giá trị liên tục của các đặc trưng của một giao dịch tài chính. Mô hình sử dụng các đặc trưng của các giao dịch, như số tiền, địa chỉ IP, quốc gia, thời gian giao dịch, loại thẻ và các thông tin khác để dự đoán giá trị liên tục của các đặc trưng đó. Sau đó, một ngưỡng được thiết lập để phân loại giao dịch là hợp lệ hoặc gian lận. Mô hình Regression trong phát hiện gian lận tài chính có thể được sử dụng để dự đoán giá trị của các đặc trưng liên tục của một giao dịch, giúp tăng độ chính xác của mô hình phát hiện gian lận.

Các bước để xây dựng một mô hình Regression trong phát hiện gian lận tài chính bao gồm:

- Chuẩn bị dữ liệu: thu thập, xử lý và chuẩn bị dữ liệu cho huấn luyện và đánh giá mô hình.
- Xác định các đặc trưng: Chọn các đặc trưng quan trọng để dự đoán giá trị liên tục của các đặc trưng đó.
- Phân chia dữ liệu: Phân chia dữ liệu thành hai tập: tập huấn luyện và tập kiểm tra.
- Huấn luyện mô hình: Sử dụng tập huấn luyện để huấn luyện mô hình Regression. Các thuật toán phổ biến bao gồm: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression và

Decision Tree Regression, Decision Tree, Random Forest, Naive Bayes, Logistic Regression và Support Vector Machines, ..

- Đánh giá mô hình: Sử dụng tập kiểm tra để đánh giá hiệu suất của mô hình Regression. Các độ đo đánh giá thường được sử dụng là: Mean Squared Error (MSE), Mean Absolute Error (MAE), R-squared, Accuracy, Precision, Recall, F1-score, AUC-ROC.
- Tinh chỉnh mô hình: Tinh chỉnh các tham số của mô hình để cải thiện độ chính xác của mô hình.
- Triển khai và giám sát mô hình: Triển khai mô hình vào hệ thống thực tế và giám sát mô hình để đảm bảo tính ổn định và độ chính xác của mô hình.

2.1.2. Unsupervised Learning (Học không giám sát)

Đối với dạng thuật toán này, chúng ta không biết trước nhãn hãy giá trị dự đoán mà chỉ có dữ liệu đầu vào. Thuật toán unsupervised learning sẽ dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó, ví dụ như phân nhóm (clustering) hoặc giảm số chiều của dữ liệu (dimension reduction) để thuận tiện trong việc lưu trữ và tính toán. Unsupervised learning là khi chúng ta chỉ có dữ liệu vào X mà không biết nhãn Y tương ứng. Unsupervised learning vì không giống như Supervised learning,chúng ta không biết câu trả lời chính xác cho mỗi dữ liệu đầu vào. Giống như khi ta học không có thầy cô giáo nào chỉ cho ta biết đó là chữ A hay chữ B. Cụm không giám sát được đặt tên theo nghĩa này. Học máy không giám sát (unsupervised learning) trong phát hiện gian lận tài chính là một phương pháp học máy không cần đến dữ liệu đã được gán nhãn trước đó (không giám sát). Thay vào đó, phương pháp này sử dụng các thuật toán để khám phá các mẫu và cấu trúc tự nhiên trong dữ liệu.

Trong phát hiện gian lận tài chính, học máy không giám sát được sử dụng để tìm kiếm các giao dịch bất thường mà không cần phải định nghĩa trước các đặc trưng của giao dịch gian lận. Các phương pháp phổ biến của học máy không giám sát bao gồm:

- Clustering: Phân loại các giao dịch thành các nhóm dựa trên các đặc trưng tương tự, sau đó xác định các nhóm có chứa các giao dịch bất thường.
- Dimensionality reduction: Giảm số chiều của dữ liệu bằng cách chuyển đổi các đặc trưng của giao dịch thành một tập hợp các đặc trưng mới. Sau đó, các phương pháp phân tích thống kê được sử dụng để xác định các giao dịch bất thường dựa trên sự khác biệt giữa các đặc trưng mới.
- Anomaly detection: Phát hiện các giao dịch bất thường bằng cách xác định các điểm dữ liệu nằm ngoài phạm vi của phân phối dữ liệu hoặc có sự khác biệt đáng kể so với các điểm dữ liệu khác.

Ưu điểm của học máy không giám sát (unsupervised learning) trong phát hiện gian lận tài chính:

- Khả năng phát hiện các giao dịch gian lận mới: Với học máy không giám sát, không cần phải định nghĩa trước các đặc trưng của giao dịch gian lận. Thay vào đó, các thuật toán có thể tự động phát hiện các đặc trưng mới và tìm kiếm các giao dịch bất thường mà không cần phải dựa vào các giao dịch gian lận đã biết trước đó.
- Khả năng xử lý dữ liệu lớn: Với học máy không giám sát, không cần phải đánh giá các điểm dữ liệu riêng lẻ, điều này giúp cho việc phát hiện gian lận có thể được thực hiện trên một lượng lớn dữ liệu trong thời gian ngắn.

- Khả năng giảm thiểu sự can thiệp của con người: Với học máy không giám sát, các quyết định phát hiện gian lận có thể được tự động hóa, giảm thiểu sự can thiệp của con người trong quá trình phát hiện và giảm thiểu các sai sót do con người gây ra.
- Khả năng tìm kiếm các mối liên hệ tiềm năng: Với học máy không giám sát, các thuật toán có thể tìm ra các mối liên hệ phức tạp giữa các giao dịch và khách hàng mà có thể dẫn đến các hành vi lừa đảo.

Nhược điểm của học máy không giám sát (unsupervised learning) trong phát hiện gian lận tài chính:

- Khó khăn trong việc xác định đối tượng gian lận: Do không có nhãn dữ liệu để phân loại các điểm dữ liệu, việc xác định đối tượng gian lận sẽ trở nên khó khăn hơn. Một số điểm dữ liệu có thể bị sai lệch hoặc bất thường do các lý do khác như lỗi nhập liệu hoặc thay đổi về môi trường.
- Khả năng tìm ra các giao dịch gian lận có thể bị hạn chế: Với học máy không giám sát, các thuật toán phát hiện gian lận có thể bỏ sót các giao dịch gian lận vì chúng không phải là các điểm dữ liệu bất thường hoặc không thuộc vào phân phối dữ liệu của tập huấn luyện.
- Khó khăn trong việc giải thích kết quả: Khi các thuật toán học máy không giám sát phát hiện gian lận, chúng ta không biết chính xác các đặc trưng và quy luật mà thuật toán đã sử dụng để phát hiện gian lận. Điều này làm cho việc giải thích kết quả trở nên khó khăn hơn.
- Sự ổn định của thuật toán: Các thuật toán học máy không giám sát có thể không ổn định trong khi phát hiện gian lận vì chúng dựa trên các mô hình phân phối dữ liệu của tập huấn luyện, điều này dẫn đến sự khác biệt giữa các mô hình được tạo ra từ các tập dữ liệu khác nhau.

- Khả năng phát hiện gian lận giả: Các kẻ gian lận có thể tìm cách đánh lừa các thuật toán học máy không giám sát bằng cách thực hiện các hành vi giả để làm cho các giao dịch của họ trông giống như các giao dịch bình thường.

Các bước sử dụng học máy không giám sát trong phát hiện gian lận tài:

- Chuẩn bị dữ liệu: Thu thập và chuẩn bị các tập dữ liệu về tài chính, bao gồm các thuộc tính như số tiền giao dịch, thời gian giao dịch, loại giao dịch, thông tin người thực hiện giao dịch, v.v.
- Tiền xử lý dữ liệu: Tiền xử lý dữ liệu để loại bỏ nhiễu, xử lý các giá trị bị thiếu, thay thế giá trị bất thường và chuyển đổi các thuộc tính dạng văn bản sang số hóa để có thể đưa vào mô hình học máy.
- Phân tích đặc trưng: Tìm kiếm và lựa chọn các đặc trưng quan trọng, có thể ảnh hưởng đến gian lận tài chính như số tiền giao dịch lớn, tần suất giao dịch, địa điểm giao dịch, v.v.
- Huấn luyện mô hình: Sử dụng thuật toán học máy không giám sát để huấn luyện mô hình trên tập dữ liệu đã được tiền xử lý và phân tích đặc trưng.
- Đánh giá mô hình: Đánh giá hiệu suất của mô hình bằng cách sử dụng các phương pháp như ma trận nhầm lẫn, độ chính xác, độ phủ, độ F1, ..
- Điều chỉnh mô hình: Tinh chỉnh các tham số của mô hình để cải thiện hiệu suất, bao gồm việc thay đổi số lượng và kiểu các đặc trưng, lựa chọn các thuật toán khác nhau và điều chỉnh các tham số của thuật toán.
- Triển khai mô hình: Triển khai mô hình trên hệ thống thực tế và thực hiện giám sát thường xuyên để đảm bảo rằng mô hình vẫn có thể phát hiện các hành vi gian lận mới.

2.1.3. Semi-Supervised Learning (Học bán giám sát):

Các bài toán khi chúng ta có một lượng lớn dữ liệu nhưng chỉ một phần trong chúng được gán nhãn được gọi là Semi-Supervised Learning. Những bài toán thuộc nhóm này nằm giữa hai nhóm Supervised và Unsupervised. Mô hình Semi-Supervised Learning là một trong những phương pháp học máy được sử dụng trong phát hiện gian lận tài chính. Điểm mạnh của phương pháp này là nó có thể sử dụng các dữ liệu không được gán nhãn để cải thiện khả năng dự đoán và tăng cường tính linh hoạt của mô hình. Trong phương pháp Semi-Supervised Learning, mô hình được huấn luyện trên hai loại dữ liệu khác nhau: dữ liệu được gán nhãn và dữ liệu không được gán nhãn. Dữ liệu được gán nhãn là các dữ liệu đã được đánh dấu là gian lận hoặc không gian lận, trong khi dữ liệu không được gán nhãn là các dữ liệu chưa được phân loại. Phương pháp Semi-Supervised Learning sử dụng các kỹ thuật phân tích dữ liệu để tìm ra các đặc trưng quan trọng trong dữ liệu và xác định các mẫu gian lận tiềm năng. Sau đó, mô hình sẽ được huấn luyện trên cả dữ liệu được gán nhãn và dữ liệu không được gán nhãn để cải thiện khả năng dự đoán của mô hình. Một trong những ứng dụng của mô hình Semi-Supervised Learning trong phát hiện gian lận tài chính là phát hiện gian lận trong các giao dịch tài chính. Mô hình sẽ được huấn luyện trên các dữ liệu được gán nhãn và không được gán nhãn để xác định các giao dịch gian lận tiềm năng. Khi có các giao dịch mới, mô hình sẽ dự đoán xác suất giao dịch đó có phải là gian lận hay không. Nếu xác suất đó vượt quá ngưỡng được thiết lập trước đó, hệ thống sẽ đưa ra cảnh báo cho nhân viên quản lý tài chính để xác nhận lại thông tin và đưa ra quyết định.

Các bước để xây dựng mô hình Semi-Supervised Learning trong phát hiện gian lận tài chính bao gồm:

- Chuẩn bị dữ liệu: Thu thập và chuẩn bị dữ liệu tài chính để sử dụng trong mô hình. Dữ liệu có thể bao gồm thông tin về giao dịch, tài khoản, người dùng, v.v.
- Xác định mục tiêu: Xác định mục tiêu của mô hình, tức là các mẫu hoặc quy tắc cần tìm kiếm trong dữ liệu để phát hiện gian lận tài chính.
- Phân tích dữ liệu: Sử dụng các kỹ thuật phân tích dữ liệu để tìm ra các tính năng quan trọng và giảm chiều dữ liệu nếu cần thiết.
- Xây dựng mô hình: Áp dụng mô hình Semi-Supervised Learning để huấn luyện mô hình trên tập dữ liệu đã được gán nhãn và không được gán nhãn. Các mẫu không được gán nhãn sẽ được sử dụng để tăng cường khả năng dự đoán của mô hình.
- Đánh giá mô hình: Đánh giá kết quả của mô hình và đưa ra cảnh báo về các hoạt động không bình thường.
- Tối ưu hóa mô hình: Tối ưu hóa mô hình bằng cách thay đổi các thông số và thử nghiệm trên các bộ dữ liệu khác nhau để cải thiện tính chính xác và đáng tin cậy của mô hình.
- Triển khai mô hình: Triển khai mô hình vào hệ thống phát hiện gian lận tài chính để theo dõi các hoạt động không bình thường và đưa ra cảnh báo cho người quản lý tài chính.

Ưu điểm trong phát hiện gian lận tài chính:

- Tận dụng dữ liệu chưa được gán nhãn: Mô hình Semi-Supervised Learning có thể sử dụng dữ liệu không được gán nhãn để cải thiện khả năng dự đoán và tăng cường tính linh hoạt của mô hình. Điều này giúp cải thiện khả năng phát hiện gian lận và giảm thiểu sai sót phân loại.
- Giảm chi phí: Mô hình Semi-Supervised Learning giúp giảm chi phí phát hiện gian lận bằng cách sử dụng dữ liệu không được gán nhãn.

Việc sử dụng dữ liệu này giúp giảm số lượng dữ liệu cần được gán nhãn, giảm thời gian và chi phí cho việc thu thập và gán nhãn dữ liệu.

- Tăng cường tính linh hoạt: Mô hình Semi-Supervised Learning có tính linh hoạt cao, có thể được sử dụng trong nhiều loại dữ liệu khác nhau, từ dữ liệu tài chính đến dữ liệu y tế và dữ liệu thương mại điện tử. Nó cũng có thể được sử dụng để phát hiện nhiều loại gian lận khác nhau, bao gồm gian lận thẻ tín dụng, gian lận định giá chứng khoán, và gian lận bảo hiểm.
- Độ chính xác cao: Mô hình Semi-Supervised Learning có khả năng dự đoán chính xác cao hơn so với các mô hình truyền thống chỉ sử dụng dữ liệu được gán nhãn. Điều này giúp giảm thiểu số lượng các trường hợp gian lận bị bỏ sót hoặc bị xác định sai.

Nhược điểm trong phát hiện gian lận tài chính:

- Phụ thuộc vào chất lượng dữ liệu: Mô hình Semi-Supervised Learning phụ thuộc vào chất lượng của dữ liệu được sử dụng để huấn luyện mô hình. Nếu dữ liệu không được chuẩn hoá, không đủ đại diện hoặc có nhiều nhiễu, kết quả dự đoán của mô hình có thể bị sai lệch.
- Đòi hỏi khả năng kết hợp dữ liệu: Mô hình Semi-Supervised Learning đòi hỏi khả năng kết hợp dữ liệu được gán nhãn và không được gán nhãn để huấn luyện mô hình. Điều này đòi hỏi kỹ năng và kinh nghiệm để có thể thiết kế một quá trình huấn luyện hiệu quả.
- Có thể gặp vấn đề khi có quá nhiều dữ liệu không được gán nhãn: Mô hình Semi-Supervised Learning có thể gặp vấn đề khi có quá nhiều dữ liệu không được gán nhãn, đặc biệt là khi mô hình không thể phân tích và hiểu được những đặc trưng quan trọng của dữ liệu.
- Chưa phù hợp với một số loại dữ liệu: Mô hình Semi-Supervised Learning có thể không phù hợp với một số loại dữ liệu nhất định, như

dữ liệu có tính độc lập cao hoặc dữ liệu không phân bố đều. Điều này có thể ảnh hưởng đến hiệu quả của mô hình.

2.1.4. Reinforcement Learning (Học Củng Cố/Tăng cường):

Reinforcement Learning (RL) là một phương pháp học máy trong đó một đối tượng (agent) tương tác với môi trường để đạt được một mục tiêu cụ thể. RL đã được sử dụng trong một số ứng dụng phát hiện gian lận tài chính.
Trong phát hiện gian lận tài chính, RL có thể được sử dụng để giám sát và phát hiện hành vi gian lận của các đối tượng tài chính. Mô hình RL được huấn luyện để tự động tìm ra các hành động (actions) tối ưu để giảm thiểu gian lận. Các hành động này có thể là việc kiểm tra lại các giao dịch, hạn chế truy cập vào các tài khoản nhạy cảm, hoặc xác nhận thông tin với các bên liên quan trước khi thực hiện các giao dịch.

Một trong những ứng dụng cụ thể của RL trong phát hiện gian lận tài chính là trong các giao dịch chứng khoán. Mô hình RL có thể được huấn luyện để tự động phát hiện các giao dịch bất thường và hành vi gian lận bằng cách theo dõi các mẫu giao dịch và áp dụng các kỹ thuật học tăng cường để tối đa hóa lợi ích. Tuy nhiên, một trong những nhược điểm của RL trong phát hiện gian lận tài chính là nó có thể bị ảnh hưởng bởi các tác nhân gian lận thông minh. Các tác nhân này có thể cố gắng đánh lừa mô hình RL bằng cách tạo ra các hành động mà có vẻ hợp lệ nhưng thực chất là gian lận. Để giải quyết vấn đề này, cần phải sử dụng các kỹ thuật mới để xác định và ngăn chặn các hành vi gian lận của các tác nhân này.

Trong trường hợp phát hiện gian lận tài chính, RL có thể được sử dụng để tìm ra cách tối đa hóa việc phát hiện gian lận và giảm thiểu tỷ lệ báo động sai. Dưới đây là một số bước để xây dựng một hệ thống RL cho phát hiện gian lận tài chính:

- Xác định mục tiêu của RL: Mục tiêu của RL trong phát hiện gian lận tài chính là tối đa hóa khả năng phát hiện gian lận và giảm thiểu tỷ lệ báo động sai.
- Xác định môi trường: Môi trường là tập hợp các dữ liệu tài chính. Chúng ta cần đảm bảo rằng dữ liệu được xử lý và định dạng đúng để có thể được sử dụng bởi RL.
- Xác định các action space: Action space là tập hợp các hành động mà agent có thể thực hiện trên môi trường. Ví dụ, trong phát hiện gian lận tài chính, các hành động có thể bao gồm kiểm tra các thông tin tài chính và phát hiện các giao dịch gian lận.
- Xác định các state space: State space là tập hợp các trạng thái của môi trường. Ví dụ, trong phát hiện gian lận tài chính, các trạng thái có thể bao gồm các thông tin về tài khoản, số tiền trong tài khoản, lịch sử giao dịch, v.v.
- Xây dựng hàm phần thưởng (reward function): Hàm phần thưởng sẽ được sử dụng để đánh giá các hành động của agent. Ví dụ: hàm phần thưởng có thể đánh giá các giao dịch và phát hiện các giao dịch gian lận để tăng điểm phần thưởng.
- Xây dựng mô hình học tăng cường (RL model): RL model sẽ được sử dụng để tìm ra các hành động tốt nhất để tối đa hóa phần thưởng. Model này có thể được xây dựng bằng các thuật toán như Q-learning, Actor-Critic, v.v.
- Huấn luyện RL model: RL model sẽ được huấn luyện với dữ liệu tài chính
- Đánh giá mô hình: Đánh giá kết quả của mô hình và đưa ra cảnh báo về các hoạt động không bình thường.

- Tối ưu hóa mô hình: Tối ưu hóa mô hình bằng cách thay đổi các thông số và thử nghiệm trên các bộ dữ liệu khác nhau để cải thiện tính chính xác và đáng tin cậy của mô hình.
- Triển khai mô hình: Triển khai mô hình vào hệ thống phát hiện gian lận tài chính để theo dõi các hoạt động không bình thường và đưa ra cảnh báo cho người quản lý tài chính.

Ưu điểm trong phát hiện gian lận tài chính:

- Học tập liên tục: RL có thể học từ dữ liệu mới, tự động cập nhật chính sách và điều chỉnh chiến lược phát hiện gian lận dựa trên dữ liệu mới, giúp cho hệ thống phát hiện gian lận ngày càng chính xác hơn.
- Khả năng tự động hóa: RL có khả năng tự động hóa quyết định về cách xử lý các trường hợp gian lận phức tạp và không định trước được, giúp cho hệ thống phát hiện gian lận tự động hơn và tránh được sai sót do con người.
- Xử lý dữ liệu phi cấu trúc: RL có khả năng xử lý dữ liệu phi cấu trúc và tìm ra các mẫu gian lận không được biết trước đó. Điều này rất hữu ích trong việc phát hiện các hành vi gian lận mới mà không cần phải dựa trên mô hình đã được huấn luyện.
- Tính linh hoạt: RL có thể được sử dụng để phát hiện các loại gian lận khác nhau, từ nhỏ đến lớn và từ đơn giản đến phức tạp. Hơn nữa, RL có thể được tùy chỉnh để phù hợp với nhiều ngành công nghiệp và ứng dụng khác nhau.

Nhược điểm trong phát hiện gian lận tài chính:

- Đòi hỏi dữ liệu lớn: RL đòi hỏi một lượng lớn dữ liệu để huấn luyện mô hình hiệu quả. Trong phát hiện gian lận tài chính, việc có được số lượng dữ liệu đủ lớn và đa dạng để huấn luyện mô hình RL có thể là một thách thức.

- Khó khăn trong việc xác định thưởng: Trong RL, mô hình học từ các phản hồi hoặc "thưởng" từ môi trường để xác định những hành động tốt và xấu. Tuy nhiên, trong phát hiện gian lận tài chính, việc xác định thưởng là rất khó, do đó, việc huấn luyện mô hình RL để phát hiện gian lận có thể rất phức tạp.
- Cấu trúc dữ liệu phức tạp: Trong phát hiện gian lận tài chính, dữ liệu có thể được tổ chức theo cấu trúc phức tạp, bao gồm nhiều loại dữ liệu khác nhau như hình ảnh, âm thanh, văn bản và số liệu tài chính. Việc huấn luyện mô hình RL để xử lý các loại dữ liệu này đòi hỏi nhiều kỹ năng và kinh nghiệm.
- Khó khăn trong việc giải thích kết quả: Một trong những thách thức lớn khi sử dụng RL trong phát hiện gian lận tài chính là khó khăn trong việc giải thích tại sao một hành động được lựa chọn và vì sao nó được coi là gian lận. Điều này có thể làm giảm độ tin cậy của mô hình RL trong phát hiện gian lận và khó khăn trong việc đưa ra quyết định thực tế.

2.2. Ứng dụng AI phát hiện gian lận tài chính

2.2.1. Hiện trạng các thuật toán AI trong phát hiện gian lận tài chính

Các thuật toán AI sẽ phù hợp với từng bài toán về phát hiện giao dịch bất thường khác nhau tuy nhiên điểm chung là chúng ta phải đánh đổi về hiệu suất và độ chính xác, khó để quan sát toàn diện các thông tin của người dùng để đưa ra các dự đoán sát nhất với mức độ biến đổi hành vi của người dùng như thay đổi thói quen chi tiết, công việc, do dịp lễ tết ngày nghỉ ...Với những thuật toán về học không giám sát độ chính xác thường không cao và khó tận dụng được những thông tin đã được xác minh từ các bộ phận nghiệp vụ hay từ thông tin có được từ phản hồi của khách hàng độ phủ thông tin không đủ rộng nên thường

được sử dụng trong các trường hợp theo dõi mức độ biến động của các giao dịch kết hợp với phương pháp time series để dự đoán được xu hướng giao dịch của người dùng nhưng lại không thể tận dụng thông tin liên quan đến quan hệ, công việc, tiềm năng tài chính của khách hàng dẫn đến khi khách hàng thay đổi công việc hay tài chính thay đổi dễ gây cảnh báo nhầm dẫn đến khó chịu cho khách hàng. Điểm mạnh của Semi-Supervised là giúp khắc phục phần lớn các nhược điểm của các mô hình khác và khả năng kết hợp thêm các mô hình khác để làm giảm nhược điểm các mô hình.

2.2.2. Mô hình Semi-Supervised learning phát hiện gian lận tài chính

Mục tiêu của việc phát hiện gian lận là để dự đoán liệu một thực thể, có thể là người dùng hoặc thiết bị hoặc nhiều hơn, sẽ tham gia vào lừa đảo hay không trong tương lai. Bài toán có thể được xây dựng như một bài toán phân loại. Trên thực tế có những tương tác phong phú trong các kịch bản tài chính.

Trên thực tế người dùng có các mối quan hệ xã hội như bạn bè, bạn cùng lớp và các mối quan hệ họ hàng với nhau. Người dùng có thể có giao dịch với người bán hoặc người dùng khác. Người dùng phải đăng nhập vào một số ứng dụng để đạt được các giao dịch tài chính. Tất cả các mối quan hệ này có thể có lợi cho vấn đề phát hiện gian lận. Sau đó một số phương pháp sau bắt đầu sử dụng embedding feature đồ thị để kết hợp các tương tác của người dùng. Tuy nhiên để phát hiện gian lận và rất ít dữ liệu được gắn nhãn thường có một số lượng lớn dữ liệu chưa được gắn nhãn chưa được khai thác hết trong các phương pháp dựa trên đồ thị hiện có. Hơn nữa, các mô hình và kết quả có thể diễn giải thường được ưu tiên trong các kịch bản tài chính nhưng các phương pháp embedding đồ thị hiện tại thường là mô hình hộp đen.

Xem xét những hạn chế của các phương pháp hiện có đề án sử dụng phương pháp có thể sử dụng cả dữ liệu nhiều lần được gắn nhãn và không được

gắn nhãn để phát hiện gian lận. Tuy nhiên nó phải đối mặt với những thách thức. Như làm thế nào để kết nối dữ liệu được gắn nhãn với dữ liệu không được gắn nhãn. Do đó chỉ mô hình hóa dữ liệu được gắn nhãn là khó có thể đạt được hiệu suất. Để kết hợp dữ liệu chưa được gắn nhãn, cách kết nối các mối quan hệ giữa thông tin không được giám sát và thông tin được giám sát .Vấn đề về thông tin là thách thức đầu tiên. Làm thế nào để lập mô hình cho sự không đồng nhất của dữ liệu? Việc sử dụng dữ liệu đa chế độ có thể cung cấp thông tin toàn diện hơn về nhiệm vụ. Tuy nhiên dữ liệu đa chế độ như quan hệ xã hội và thuộc tính người dùng có các phần tử thống kê khác nhau. Sự không đồng nhất như vậy đặt ra một thách thức lớn để tích hợp dữ liệu nhiều chế độ xem. Làm thế nào để học một mô hình có thể diễn giải? Kết quả của việc phát hiện gian lận thường được phục vụ cho việc kiểm soát rủi ro tài chính. Không có vấn đề gì đối với các nhà cung cấp dịch vụ tài chính hay đối với những người giám sát. Tất cả đều yêu cầu các mô hình có thể diễn giải để có kiến thức tốt hơn về các kết quả dự đoán.

Để giải quyết thách thức trên, đề án sử dụng mô hình Semi Supervised Graph Neural. Ý tưởng cơ bản của SemiGNN là nâng cao tính đại diện của người dùng bằng cách khai thác triệt để dữ liệu quan hệ và dữ liệu thuộc tính của cả dữ liệu có nhãn và dữ liệu không có nhãn. Cụ thể, em kết nối những người dùng được gắn nhãn và chưa được gắn nhãn thông qua các mối quan hệ xã hội của họ. Và đối với mỗi người dùng, em cũng sử dụng các thuộc tính của họ để xây dựng mạng thuộc tính. Các mối quan hệ xã hội và các thuộc tính kết hợp với nhau tạo thành một mạng lưới đa phương diện rộng lớn. Mạng nơ-ron đồ thị bán giám sát để lập mô hình đồng thời thông tin đa chế độ trong mạng để phát hiện gian lận. Mô hình này có một số ưu điểm như nó có thể khai thác triệt để thông tin được giám sát và thông tin cấu trúc không được giám sát để phát hiện gian lận. Mô hình của em có thể tích hợp dữ liệu nhiều lần để có được

kết quả toàn diện cho việc phát hiện gian lận. Xây dựng mạng thuộc tính thay vì sử dụng các thuộc tính vì các đối tượng địa lý dày đặc có thể nâng cao khả năng biểu diễn của thông tin thuộc tính**.** Để nắm bắt mối quan hệ giữa các giao dịch thẻ tín dụng liên quan đến thông tin thời gian, ta sử dụng đồ thị giao dịch thời gian để mô hình hóa các mẫu liên quan đến thời gian. Chỉ một tỷ lệ rất nhỏ (ít hơn 10%) giao dịch được dán nhãn trong hàng tỷ giao dịch thực tế, trong đó chứa đựng nhiều hình thức lừa đảo chưa bị phát hiện. Vì vậy, việc khai thác các đặc điểm tự nhiên là rất quan trọng từ dữ liệu chưa được gắn nhãn.

2.3. Kết luận chương

Trong chương này đã trình bày các thuật toán AI thường được sử dụng trong hệ thống phát hiện bất thường trong gian lận tài chính được trình bày chi tiết về đặc điểm, ưu điểm, nhược điểm và các trường hợp sử dụng phù hợp với dữ liệu cụ thể. Lý do lựa thuật toán SemiGNN.

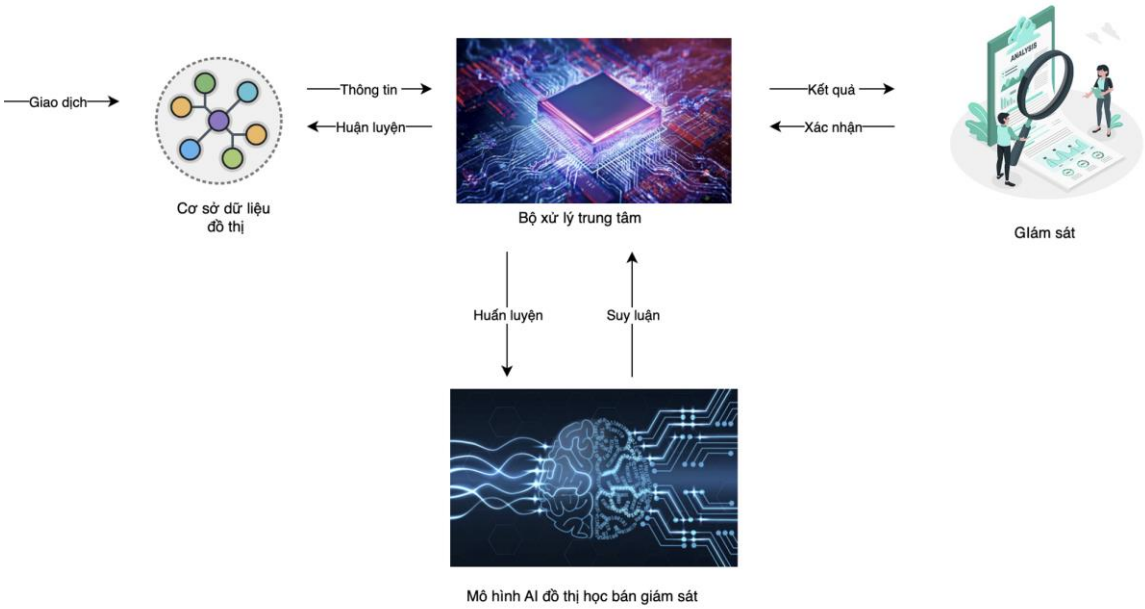
Chương tiếp theo sẽ trình bày về việc huấn luyện, kiểm thử, đánh giá, thiết kế hệ thống tích hợp thuật mô hình GTAN vào hệ thống phù hợp để áp dụng vào thực tế sử dụng.

CHƯƠNG 3: XÂY DỰNG HỆ THỐNG PHÁT HIỆN GIAN LẬN

3.1. TỔNG QUAN HỆ THỐNG

Hệ thống phát hiện gian lận tài chính chia làm 4 phần chính :

- Cơ sở dữ liệu đồ thị
- Mô hình AI đồ thị học bán giám sát
- Giám sát (hệ thống cảnh báo và giám sát viên)
- Bộ xử lý trung tâm



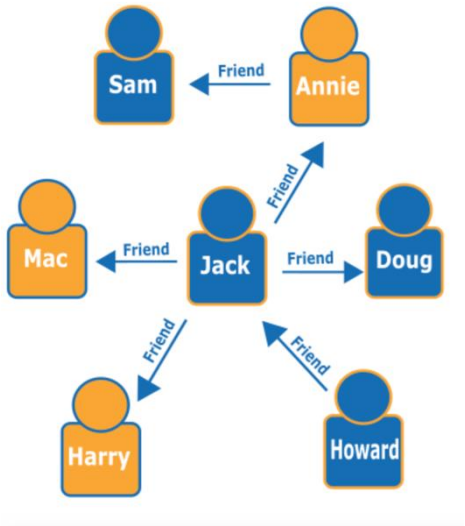
Hình 3 . 1. Kiến trúc hệ thống sử dụng mô hình Semi-Supervised Graph Neural Network

3.1.1. Cơ sở dữ liệu đồ thị

Giới thiệu Graph DataBase

Các ngân hàng và công ty bảo hiểm mất hàng tỷ đô la mỗi năm vì gian lận. Các phương pháp phát hiện gian lận truyền thống đóng một vai trò quan trọng trong việc giảm thiểu những tổn thất này. Tuy nhiên những kẻ lừa đảo ngày càng tinh vi đã phát triển nhiều cách khác nhau để tránh bị phát hiện, cả

bằng cách hợp tác với nhau và bằng cách tận dụng nhiều phương tiện khác để tạo danh tính giả. Mặc dù không có biện pháp ngăn chặn gian lận nào là toàn diện nhưng có thể cải thiện đáng kể bằng việc tăng phạm vi thông tin quan sát bao quát hơn từ các điểm dữ liệu riêng lẻ và các kết nối liên kết chúng. Những mối liên hệ này là những manh mối vô cùng quan trọng để phát hiện sớm các đặc trưng gian lận làm rõ những đặc trưng đó. Cơ sở dữ liệu đồ thị là cơ sở dữ liệu được thiết kế để sử dụng các mối quan hệ giữa các dữ liệu. Nó được thiết kế để lưu giữ dữ liệu và mối quan hệ mà không cần biến nó thành một mô hình được xác định trước. Thay vào đó, dữ liệu được lưu trữ cách riêng lẻ kết nối với nhau hoặc có liên quan với những thực thể khác như hình 3.2.

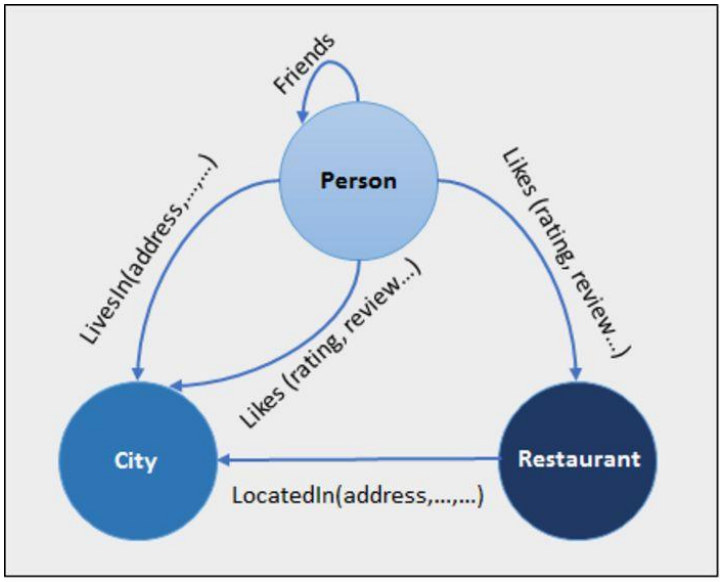


Hình 3 . 2. Đồ thị mối quan hệ mạng xã hội

Chúng ta đang sống trong một thế giới kết nối với nhau không biên giới thông qua các nền tảng mạng xã hội. Không có phần thông tin riêng biệt, mà là các miền thông tin phong phú, được kết nối xung quanh chúng ta. Chỉ một cơ sở dữ liệu bao gồm các mối quan hệ nguyên bản mới có thể lưu trữ, xử lý và truy vấn các kết nối một cách hiệu quả. Trong khi các cơ sở dữ liệu khác tính toán các mối quan hệ tại thời điểm truy vấn thông qua SQL là không hiệu quả thì với cơ sở dữ liệu đồ thị lưu trữ dữ liệu về quan hệ thì vấn đề đã được giải

quyết. Giả sử trong bài toán mạng xã hội thì con người (là các nút) và các mối quan hệ của họ (là các cạnh), có thể tìm ra ai là "bạn của những người bạn" của một người cụ thể. Truy cập các nút và mối quan hệ trong cơ sở dữ liệu đồ thị thực hiện trong thời gian ngắn một cách liên tục và cho phép nhanh chóng duyệt qua hàng triệu kết nối mỗi giây. Không phụ thuộc vào tổng kích thước của tập dữ liệu, cơ sở dữ liệu đồ thị vượt trội trong việc quản lý dữ liệu được kết nối cao và các truy vấn phức tạp. Chỉ với một mẫu và tập hợp các điểm bắt đầu, cơ sở dữ liệu đồ thị có thể khai phá dữ liệu lân cận xung quanh các điểm xuất phát ban đầu, thu thập, tổng hợp thông tin từ hàng triệu nút và mối quan hệ. Cơ sở dữ liệu đồ thị thường đơn giản về cách cấu trúc dữ liệu bao gồm 3 thành phần:

- Nút (node): là các thực thể hay các đối tượng
- Thuộc tính (properties) : là thuộc tính của các node hoặc cạnh
- Cạnh (edge): là các mối quan hệ thực tế giữa các node.



Hình 3 . 3. Ví dụ minh họa node và cạnh trong cơ sở dữ liệu đồ thị

Cơ sở dữ liệu đồ thị có khả năng ngăn chặn gian lận tinh vi. Chẳng hạn như một khách hàng có khả năng là gian lận đang sử dụng cùng một địa chỉ

email và thẻ tín dụng như trong một trường hợp gian lận đã biết. Nhờ sử dụng các mối quan hệ để xử lý các giao dịch tài chính trong thời gian gần sát với thời gian thực tế (near-real time).

Một số cơ sở dữ liệu đồ thị (Graph database) tiêu biểu

Mặc dù cơ sở dữ liệu đồ thị không quá phổ biến như một số cơ sở dữ liệu NoSQL khác, nhưng có một số cơ sở dữ liệu đã trở thành tiêu chuẩn khá tốt khi nói về graph database:

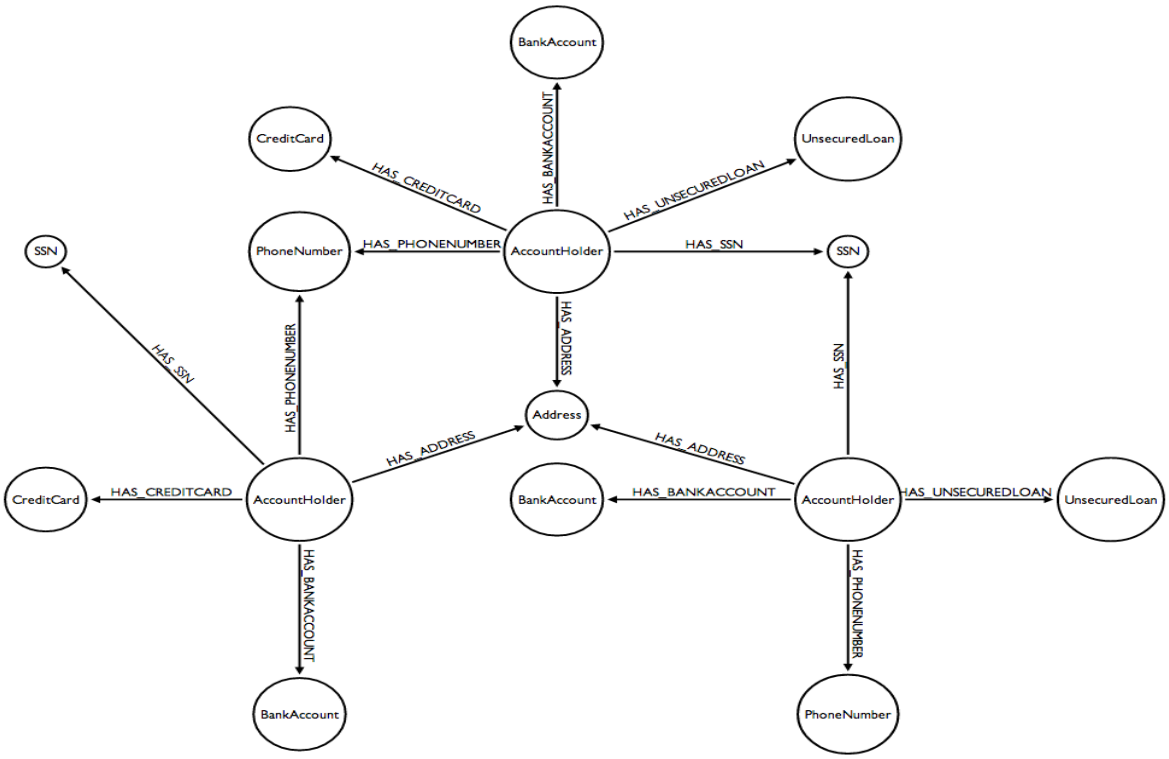
- Neo4j: là cơ sở dữ liệu mã nguồn mở vừa được xây dựng trên Java. Cypher là ngôn ngữ riêng của Neo4j tương tự như ngôn ngữ SQL. Ngoài ra Neo4j cũng hỗ trợ các ngôn ngữ phổ biến như Java, Python, .NET, JavaScript và một số ngôn ngữ khác.
- Redis Graph: là một mô đun đồ thị được tích hợp sẵn trong Redis, bản thân nó là một cơ sở dữ liệu NoSQL có key-value. Redis Graph được tạo ra để có dữ liệu được lưu trữ trong Ram. Nhờ đó mà Redis Graph có hiệu suất cao, với khả năng truy vấn và lập chỉ mục nhanh chóng. Redis Graph cũng sử dụng Cypher, nhờ đó tạo sự linh hoạt hơn.
- OrientDB: là sự kết hợp của nhiều loại mô hình dữ liệu khác nhau. Các mối quan hệ được lưu trữ bằng cách sử dụng mô hình đồ thị sử dụng kết nối trực tiếp giữa các cơ sở dữ liệu. Giống như hai cơ sở dữ liệu đồ thị trước đó, OrientDB cũng là nguồn mở và được viết bằng Java (mặc dù không sử dụng Cypher). Ý tưởng OrientDB là để sử dụng khi yêu cầu nhiều mô hình dữ liệu và do đó được tối ưu hóa để đảm bảo tính nhất quán của dữ liệu, cũng như giảm độ phức tạp của dữ liệu.

Neo4j

Neo4j cung cấp các phương pháp để phát hiện các gian lận theo nhóm và các hành vi gian lận tinh vi với độ chính xác cao trong thời gian thực. Neo4j cũng cho phép lưu trữ các thực thể và quan hệ giữa các thực thể. Các đối tượng

này còn được gọi là các nút, trong đó có các thuộc tính. Mỗi nút là một thể hiện của một đối tượng. Quan hệ được gọi là các cạnh, có thể có các thuộc tính. Cạnh có ý nghĩa định hướng. Các nút được tổ chức bởi các cạnh. Cấu trúc đồ thị cho phép các dữ liệu được lưu trữ một lần và được giải thích theo nhiều cách khác nhau dựa trên các mối quan hệ. Thông thường, khi chúng ta lưu trữ một cấu trúc đồ thị giống như trong cơ sở dữ liệu quan hệ (RDBMS) việc tăng thêm một mối quan hệ có nghĩa là rất nhiều thay đổi trong khi việc đó rất đơn giản với cơ sở dữ liệu đồ thị.

Neo4J sử dụng ngôn ngữ Cypher đơn giản về ngữ nghĩa để phát hiện các nhóm trong đồ thị và điều hướng các kết nối bộ nhớ trong thời gian thực. Mô hình Graph database hình 3.4 biểu thị các dữ liệu thực sự trông như thế nào trong Graph database và minh họa cách một người có thể tìm thấy các nhóm bằng cách di chuyển trên đồ thị:



Hình 3 . 4. Kiến trúc dữ liệu trong Graph DataBase

Cài đặt Neo4J

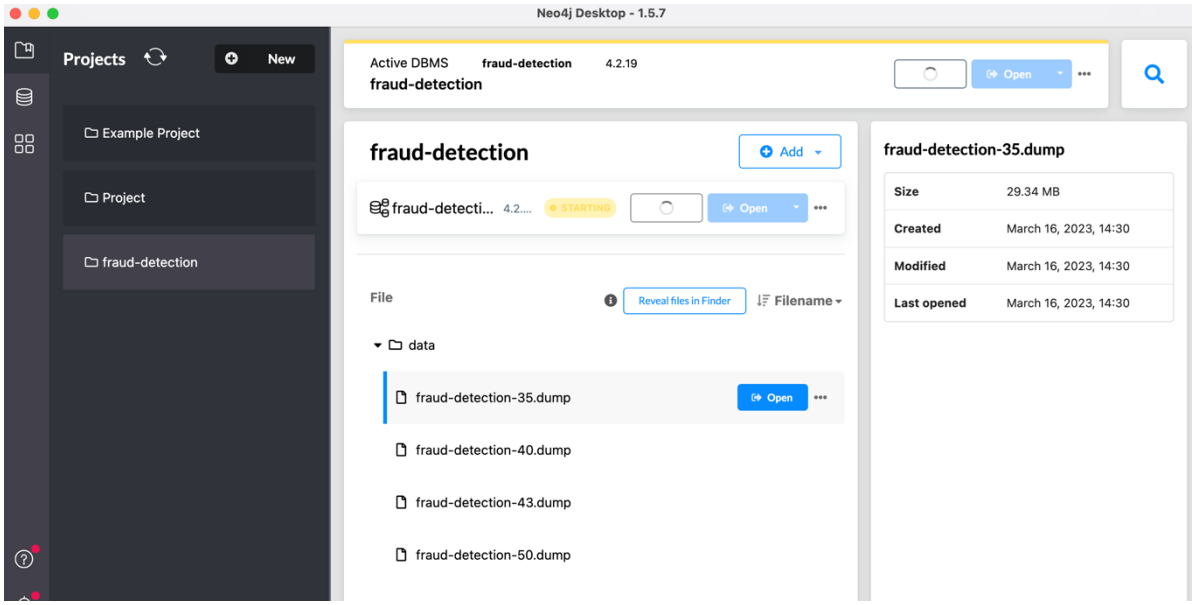
Neo4j có thể được cài đặt trong nhiều môi trường và phạm vi khác nhau, do đó yêu cầu hệ thống phần lớn phụ thuộc vào việc sử dụng phần mềm. Neo4j được hỗ trợ trên các hệ thống có kiến trúc x86_64 và ARM trên các nền tảng vật lý, ảo hoặc containerized platforms.

Điều chỉnh phần cứng phù hợp với kích thước cơ sở dữ liệu sử dụng:

- CPU: Giới hạn tính toán đối với đồ thị phù hợp với bộ nhớ và khả năng tính toán của CPU.
- RAM: Nhiều ram hơn cho phép làm việc với đồ thị lớn hơn nhưng cần được cấu hình đúng cách để tránh hệ thống dọn dẹp tài nguyên thừa gây gián đoạn.
- Ổ cứng: Hiệu suất của ổ cứng ảnh hưởng lớn đến khi lưu trữ

Neo4j có 2 phiên bản Neo4j Browser và Neo4j Desktop:

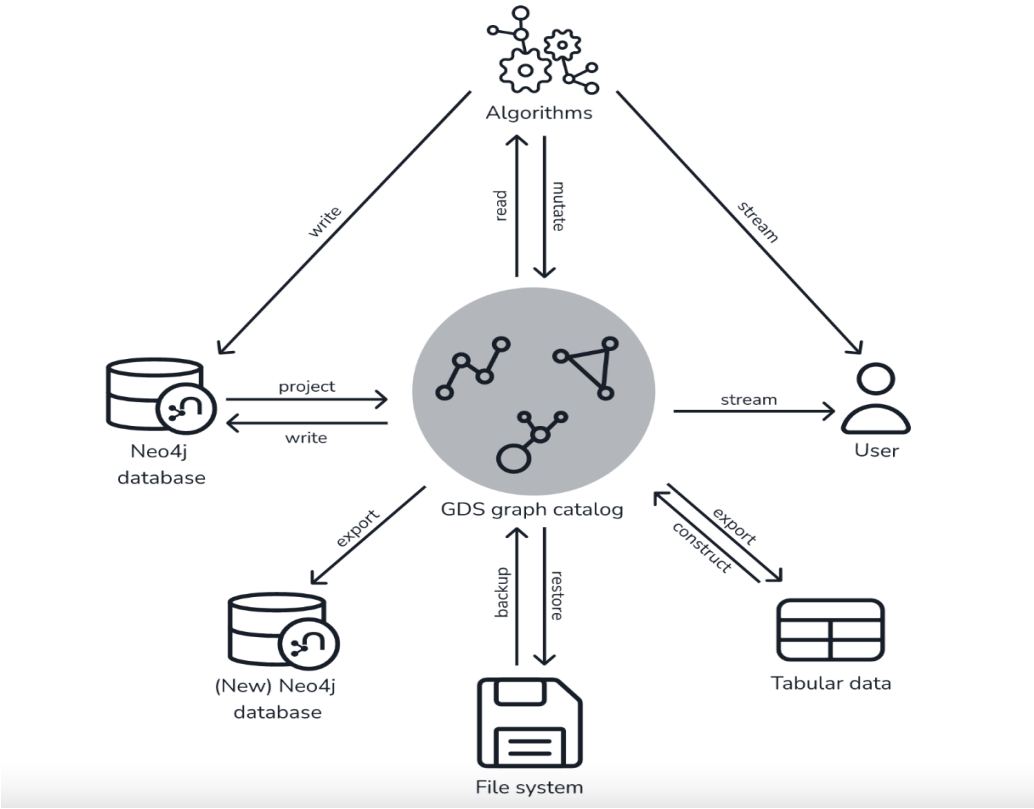
- Neo4j Browser là một công cụ sử dụng trên trình duyệt giúp tương tác với đồ thị. Neo4j Browser có 2 phiên bản là Doanh nghiệp và phiên bản dành cho cộng đồng. Neo4j Browser đi kèm với Neo4j DBMS, bao gồm cả Neo4j Server và Neo4j Desktop. Các browser hỗ trợ bao gồm: Chrome, FireFox, Edge
- Neo4j Desktop là ứng dụng trên máy tính thuận tiện để làm với cơ sở dữ liệu cục bộ.



Hình 3 . 5. Giao diện Neo4j

Graph Data Science

Thư viện GDS thường được chia thành hai giai đoạn: phát triển và sản xuất. Trong giai đoạn phát triển, mục tiêu là thiết lập quy trình làm việc của các thuật toán hữu ích. Để làm được điều này, hệ thống phải được cấu hình, xác định các phép chiếu đồ thị và lựa chọn các thuật toán. Hệ thống được cấu hình phù hợp để chạy thành công các thuật toán mong muốn. Trình tự các hoạt động thông thường sẽ là chiếu một đồ thị, chạy một hoặc nhiều thuật toán và đánh giá kết quả.



Hình 3 . 6. Tổng quan về hoạt động tiêu chuẩn của thư viện GDS

Ngoài ra, GDS cung cấp các chế độ thực thi khác nhau:

- Stream: Chế độ Stream sẽ trả về kết quả tính toán thuật toán dưới dạng các hàng. Tương tự như cách truy vấn Cypher .Dữ liệu được trả về có thể là id nút và giá trị tính toán cho nút hoặc hai id nút và giá trị tính toán cho cặp nút (chẳng hạn như điểm tương đồng về độ tương đồng của nút). Nếu đồ thị rất lớn, kết quả tính toán chế độ stream cũng sẽ rất lớn.
- Stat: Chế độ thống kê trả về kết quả thống kê như số lượng hoặc phân phối. Bản tóm tắt thống kê của phép tính được trả về dưới dạng một hàng kết quả. Chế độ này tạo cơ sở cho các chế độ ghi và thay đổi.
- Mutate: Chế độ ghi kết quả tính toán của thuật toán trở lại đồ thị được chiếu. Nó cho phép chạy nhiều thuật toán trên cùng một đồ thị được

chiếu mà không cần ghi kết quả vào Neo4j ở giữa các lần thực thi thuật toán. Chế độ thực thi này đặc biệt hữu ích trong các trường hợp:

- Các thuật toán có thể phụ thuộc vào kết quả của các thuật toán trước đó mà không cần ghi vào Neo4j.
- Kết quả thuật toán có thể được viết hoàn toàn (xem ghi thuộc tính nút và ghi mối quan hệ).
- Kết quả thuật toán có thể được truy vấn thông qua Cypher mà không cần phải ghi vào Neo4j.

Một bản tóm tắt thống kê của tính toán được trả về tương tự như chế độ thống kê. Dữ liệu bị thay đổi có thể là thuộc tính của nút (chẳng hạn như điểm xếp hạng), mối quan hệ mới (chẳng hạn như điểm tương đồng của nút tương tự) hoặc thuộc tính của mối quan hệ.

- Write: Chế độ ghi sẽ ghi kết quả tính toán thuật toán trở lại cơ sở dữ liệu Neo4j. Tương tự như cách truy vấn viết Cypher. Một bản tóm tắt thống kê của tính toán được trả về tương tự như chế độ thống kê. Đây là chế độ thực thi duy nhất sẽ sửa đổi cơ sở dữ liệu Neo4j. Dữ liệu được viết có thể là thuộc tính của nút, mối quan hệ mới hoặc thuộc tính của mối quan hệ. Chế độ ghi có thể rất hữu ích cho các trường hợp sử dụng mà kết quả thuật toán sẽ được kiểm tra nhiều lần bằng các truy vấn riêng biệt vì kết quả tính toán được thư viện xử lý. Để thuật toán khác sử dụng kết quả từ phép tính chế độ ghi, một đồ thị mới phải được chiếu từ cơ sở dữ liệu Neo4j với đồ thị được cập nhật.
- Logging: Trong thư viện GDS có ba loại ghi nhật ký: ghi nhật ký gỡ lỗi, ghi nhật ký tiến trình và ghi nhật ký gợi ý hoặc cảnh báo.

Ghi nhật ký gỡ lỗi cung cấp thông tin về các sự kiện trong hệ thống. Ví dụ: khi quá trình tính toán thuật toán hoàn tất, dung lượng bộ nhớ đã sử dụng và tổng thời gian chạy có thể được ghi lại. Các sự kiện đặc biệt, khi một hoạt

động không hoàn thành bình thường, cũng được ghi lại. Thông tin nhật ký gỡ lỗi rất hữu ích để hiểu các sự kiện trong hệ thống, đặc biệt là khi khắc phục sự cố.

Ghi nhật ký tiến trình được thực hiện để theo dõi tiến trình của các hoạt động dự kiến sẽ mất nhiều thời gian. Điều này bao gồm các phép chiếu đồ thị, tính toán thuật toán và viết kết quả.

Ghi nhật ký gợi ý hoặc cung cấp cảnh báo cho người dùng liên quan đến truy vấn của họ.

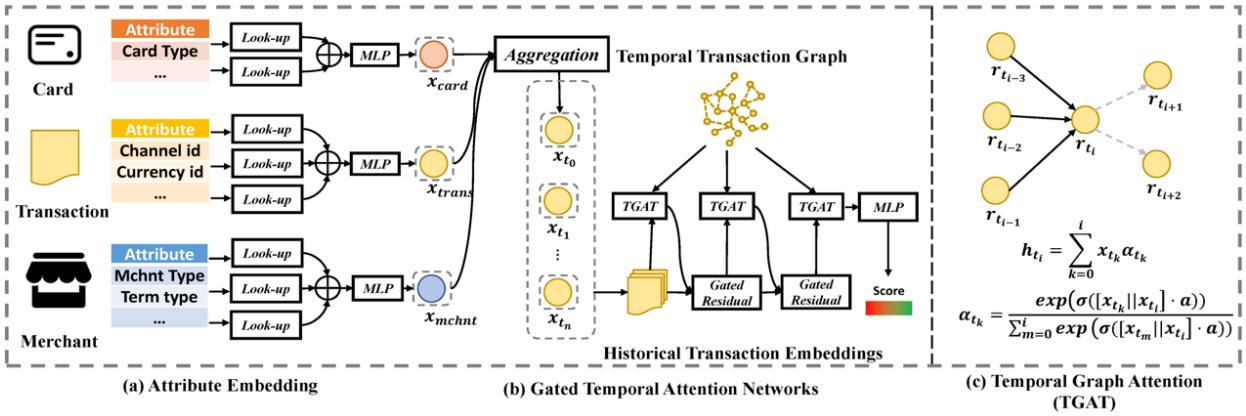
3.1.2. Mô hình AI đồ thị học bán giám sát

a.Mô hình AI

Sử dụng mô hình Gated Temporal Attention Network (GTAN) cho đồ thị giao dịch theo thời gian, có thể trích xuất các mô hình gian lận tạm thời và khai thác cả dữ liệu được dán nhãn và không được gắn nhãn. Ngoài ra, các thuộc tính phân loại rất phổ biến và hữu ích trong các ứng dụng thực tế. Do đó, cần tận dụng thông tin hữu ích thông qua mô hình hướng thuộc tính. Trong đề án sử dụng feature learning để mô hình tự động xác định và tối ưu hóa các mẫu, cấu trúc,đặc điểm từ dữ liệu thô để nâng cao hiệu quả xử lý trước thuộc tính của các giao dịch và thêm thuộc tính mới, để có thể mô hình hóa các hành vi gian lận tốt hơn. Cách tiếp cận của em giải quyết vấn đề bằng cách phát hiện gian lận thông qua mô hình truyền thông tin rủi ro từ các nút lân cận. Việc sử dụng mô hình định semi-supervised graph neural network để tìm ra nhiều kiểu gian lận hơn, giúp cải thiện đáng kể độ chính xác của việc phát hiện gian lận.[3]

Cụ thể, chúng ta tận dụng các bản ghi giao dịch để xây dựng đồ thị giao dịch theo thời gian, trong đó bao gồm các giao dịch tạm thời (nút) và tương tác (cạnh) giữa chúng. Sau đó truyền thông tin giữa các nút thông qua mạng GTAN để biểu diễn giao dịch. Sau đó mô hình hóa mô hình hóa hành gian lận thông qua việc lan truyền rủi ro giữa các giao dịch. Các thí nghiệm mở rộng được tiến

hành trên thực tế tập dữ liệu giao dịch phát hiện gian lận được công khai. Kết quả cho thấy phương pháp GTAN vượt trội với kết quả tốt nhất. Các thí nghiệm bán giám sát thực tế chứng minh hiệu quả phát hiện gian lận của mô hình đề xuất chỉ với một tỷ lệ nhỏ dữ liệu được dán nhãn. Các thuộc tính thô của bản ghi giao dịch được học bởi thuộc tính tổng hợp và feature learning, bao gồm việc tổng hợp đặc trưng với lớp multi-layer perception (MLP). Thuộc tính giao dịch bao gồm ID kênh, ID tiền tệ, số tiền giao dịch,v.v. Các thuộc tính người bán bao gồm loại hình kinh doanh, địa điểm kinh doanh, lĩnh vực, tỷ lệ phí, v.v. Sau đó tạo ra một mạng lưới tạm thời có kiểm soát để tổng hợp và tìm hiểu tầm quan trọng của việc embedding giao dịch lịch sử. Sau đó, ta tận dụng MLP hai lớp để tìm hiểu hành vi gian lận xác suất từ những biểu diễn này. Toàn bộ mô hình có thể được tối ưu hóa theo cơ chế end-to-end cùng với thuật toán giảm độ dốc ngẫu nhiên hiện có. [3]



Hình 3 . 7. Kiến trúc mô hình Gated Temporal Attention Network (GTAN)

Attribute Embedding and Feature Learning

Bản ghi giao dịch r = (r_1, r_2, · · ·, r_N), mỗi bản ghi r_i chứa thuộc tính thẻ ngân hàng f_c^i, thuộc tính giao dịch f_r^i và thuộc tính người bán hàng f_m^i như r_i =f_c^i, f_r^i, f_m^i . Trong quá trình tiền xử lý, ta không loại bỏ bất kỳ node của thẻ ngân hàng hay người bán nào có ít hồ sơ giao dịch. Mặc dù số lượng thẻ ngân hàng và đơn vị kinh doanh cần được kiểm tra thủ công lớn hơn nhiều so với đã

kiểm tra, em áp dụng hồ sơ giao dịch đầy đủ của khách hàng để đảm bảo việc phát hiện được hành vi gian lận tiềm ẩn. Từ đó ta xây dựng biểu diễn thuộc tính số của từng bản ghi thành định dạng tensor $X_{num} \in R^{N\times d}$ trong đó N biểu thị số lượng giao dịch và d biểu thị kích thước của các đặc trưng. Bên cạnh đó, ta trích xuất các thuộc tính thẻ, giao dịch và danh mục người bán $X_{cat} \in R^{N\times d}$ riêng biệt thông qua thuộc tính

$$\begin{aligned} e_{\text{attr}} &= \text{onehot}(f_{\text{attr}}) \odot \mathbf{E}_{\text{attr}}, \\ x_{\text{cat},i} &= \text{MLP}_i\left(\sum_{\forall j \in table} e_j\right), i \in \{\text{card,trans, mchnt}\} \end{aligned} \tag{3.1}$$

Trong đó $j \in table_i$ biểu thị cột j trong bảng đầu vào của dữ liệu i, $e_{attr} \in R^{1\times d}$ biểu diễn embedding của thuộc tính, one hot (·) biểu diễn mã hóa one-hot, f_{attr} biểu diễn thuộc tính đơn lẻ của giao dịch và $E_{attr} \in R^{m\times d}$ biểu diễn ma trận embedding của thuộc tính, trong đó m biểu diễn số lượng thuộc tính tối đa. Sau khi có embedding vector của từng thuộc tính của thẻ, giao dịch và bảng người bán, em tổng hợp các phần embedding này để có được phần embedding phân loại của mỗi giao dịch thông qua việc cộng gộp với $x_{cat}^{(u)} = \sum_i x_{cat,i}^{(u)}, i \in$ {card, trans, mchnt}, với $x_{cat}^{(u)} \in R^{1\times d}$ biểu thị embedding vector danh mục của bản ghi giao dịch thứ u. Để giải quyết tính không đồng nhất của các thuộc tính phân loại, feature learning được sử dụng để mô hình hóa tất cả các phân loại thuộc tính và chiếu lên một chiều không gian thống nhất, giúp mô hình graph learning hiệu quả hơn.

Gated Temporal Attention Networks

Để học được những đặc trưng về thời gian giao dịch gian lận, ta tạo đồ thị về giao dịch theo thời gian và tổng hợp thông tin từ đồ thị và cập nhật embedding của mỗi giao dịch. Các cạnh có hướng được tạo ra với các giao dịch trước đó là nguồn và những giao dịch hiện tại làm đích. Sau đó tổng hợp thông

tin thông qua Temporal Graph Attention. Số cạnh thời gian được tạo trên mỗi node là một siêu tham số .

Temporal Graph Attention

Sau khi xử lý đặc trưng và embedding thuộc tính, ta tận dụng các giao dịch embedding $\mathbf{X} = \{x_{t_0}, x_{t_1}, ...x_{t_n}\}$ để tìm hiểu thời gian được embedding ở từng bản ghi giao dịch. Đầu tiên chúng ta kết hợp thuộc tính danh mục và thuộc tính số làm đầu vào của mạng GTAN với $x_{ti} = x_{num}^{(t_i)} + x_{cat}^{(t_i)}$. Lúc đầu lớp TGAT, ta đặt $\mathbf{H_0} = \mathbf{X}$ làm ma trận embedding đầu vào. Sau đó, sử dụng multi-head attention để tính toán riêng độ quan trọng của từng node lân cận và cập nhật embedding như công thức dưới:

$$\mathbf{H} = \text{Concat}(\text{Head}_1, \ldots, \text{Head}_{h_{att}})\mathbf{W}_o, \tag{3.2}$$

Trong đó h_{att} biểu thị số lượng head, Wo ∈ $R^{d\times d}$ ghi các tham số cóđược học, H biểu diễn cho tổng hợp embedding với, và $\mathbf{H} = \{h_{t_0}, h_{t_1}, ..., h_{t_n}\}$ attention head được xây dựng như sau:

$$\begin{aligned}\text{Head} &= \sum_{x_i \in \mathcal{X}} \sigma\left(\sum_{x_t \in \mathcal{N}(x_i)} \alpha_{x_t,x_i} x_t\right) \\ \alpha_{x_t,x_i} &= \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T[x_t \| x_i]\right)\right)}{\sum_{x_j \in \mathcal{N}(x_t)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^T[x_t \| x_j]\right)\right)},\end{aligned} \tag{3.3}$$

Trong đó N (xi) biểu diễn các thời gian các node lân cận, α_{x_t,x_i} biểu diễn tầm quan trọng của cạnh thời gian (x_t,x_i) trong attention head và a ∈ R^{2d} biểu thị trọng số vector của head. Trong thực tế, để tránh quá tải không gian tính toán của hệ thống trong trường hợp tải của hệ thống bị đẩy lên cao(chẳng hạn như các giao dịch xuất hiện tần số lớn trong một thời gian ngắn), ta sử dụng chiến lược lấy mẫu và hạn chế số lượng node lân cận để kiểm soát số lượng

node lân cận |N (xt)| (là số lượng các cạnh thời gian liên quan trên mỗi node) thông qua đó lớp Temporal Graph Attention truyền thông tin . Ngoài ra, để tránh mượn thông tin trong tương lai, các giao dịch lân cận được lấy mẫu cho mỗi giao dịch bắt buộc phải là các giao dịch trong quá khứ của cùng một chủ thẻ để chúng ta có thể mô hình hóa thời gian giao dịch gian lận thông qua cơ chế message passing của Temporal Graph Attention.

Attribute-driven Gated Residual

Để cải thiện hơn nữa tính hiệu quả và khả năng diễn giải của phương pháp, sau khi sử dụng các embedding được tổng hợp, ta tận dụng các embedding và thuộc tính thô để suy ra mức độ quan trọng của các embedding được tổng hợp và các thuộc tính thô sau mỗi lớp TGAT, có thể được xây dựng như sau:

$$\begin{aligned} \text{gate}_{t_i} &= \sigma([x_{cat,t_i} \| x_{num,t_i} \| h_{t_i}]\beta_{t_i}) \\ z_{t_i} &= \text{gate}_{t_i} \cdot h_{t_i} + (1 - \text{gate}_{t_i}) \cdot x_{t_i} \end{aligned} \qquad (3.4)$$

Trong đó $\text{gate}_{t_i} \in [0, 1]$ biểu diễn các biến gate của giao dịch thứ t_i -th, σ biểu diễn hàm sigmoid, $\beta ti \in R^{3d\times1}$ biểu diễn vector gate và z_{ti} biểu thị vector đầu ra của mỗi lớp TGAT, được làm đầu vào cho lớp tiếp theo. Nếu ta xếp chồng một lớp TGAT mới với cơ chế dư gated dựa trên thuộc tính, ta sử dụng đầu ra của cơ chế gating k-th làm đầu vào của k + 1-th TGAT. [3]

Risk Embedding and Propagation

Em sử dụng thông tin từ việc gán nhãn thủ công cho các giao dịch là đặc trưng . Cụ thể, ta lấy nhãn được gán thủ công làm đặc trưng rủi ro của mỗi giao dịch, trong đó danh mục dữ liệu không được gắn nhãn là 'không được gắn nhãn' và danh mục của phần còn lại của dữ liệu là 'gian lận' hoặc 'bình thường'. Sau đó, ta thêm đặc trưng này vào dữ liệu giao dịch như một trong những thuộc tính phân loại đầu vào. Do các vấn đề về khả năng mô hình học đượ các nhãn đã bị che mà thuộc tính này chưa được sử dụng trong các giải pháp phát hiện gian

lận trước đây. Do vậy đề án embedding các thuộc tính rủi ro được quan sát một phần (tức là nhãn) vào cùng một không gian với các đặc trưng của nút khác, bao gồm các vector embedding rủi ro cho các nút được dán nhãn và embedding các vector 0 cho các vector không được gán nhãn. Sau đó, ta thêm các đặc trưng nút và embedding lại với nhau dưới dạng các đặc trưng nút đầu vào với $x_{t_i} = x_{num}^{(t_i)} + x_{cat}^{(t_i)} + \tilde{y}^{(t_1)}\mathbf{W}_r$, trong đó Wr ghi các tham số có thể học embedding đặc trưng rủi ro. Bằng cách ánh xạ $\hat{\mathbf{Y}}$ được gắn nhãn một phần và nút có các đặc trưng X vào cùng một không gian và cộng chúng lại, ta có thể sử dụng một mạng graph neural network để học được thông tin thuộc tính và nhãn. Do đó, mô hình phát hiện gian lận có thể mô hình hóa đặc trưng lận theo thời gian và truyền thông tin khả năng gian lận bằng cách thêm nhãn cho những giao dịch chưa có nhãn làm một trong các thuộc tính phân loại giao dịch..[3]

Dự đoán rủi ro gian lận (Fraud Risk Prediction)

Sau khi tổng hợp embedding của giao dịch, ta tận dụng MLP hai lớp để dự đoán rủi ro gian lận, được xây dựng như sau:

$$\hat{\mathbf{y}} = \sigma(\mathrm{PReLU}(\mathbf{H}\mathbf{W}_0 + \mathbf{b}_0)\mathbf{W}_1 + \mathbf{b}_1) \qquad (3.5)$$

$\hat{\mathbf{y}} \in \mathbb{R}^{N\times 1}$ trong đó yˆ∈ $\mathrm{R}^{\mathrm{N}\times 1}$ biểu diễn kết quả dự đoán rủi ro của tất cả các giao dịch, W và b biểu diễn các tham số có thể học được của MLP. Sau đó, tính toán hàm mục tiêu L thông qua binary cross-entropy, được xây dựng như sau:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=0}^{N}[\mathbf{y}_i \cdot \log p(\hat{\mathbf{y}}_i \mid \mathbf{X}, \mathbf{A}) + (1 - \mathbf{y}_i) \cdot \log(1 - p(\hat{\mathbf{y}}_i \mid \mathbf{X}, \mathbf{A}))] \quad (3.6)$$

Trong đó y biểu diễn nhãn ground-truth của các giao dịch. Mô hình GTAN được sử dụng có thể được tối ưu hóa thông qua các thuật toán dựa trên thuật toán SGD.

Masking to Avoid Label Leakage

Trước đây thông tin rủi ro thường được lấy làm mục tiêu tối ưu hóa để giám sát việc huấn luyện mô hình phát hiện gian lận. Khác với các phương pháp phát hiện gian lận trước đây, mô hình semi-supervise bằng cách truyền các thuộc tính giao dịch và embedding thông tin về rủi ro giữa các giao dịch được gán nhãn và không gán nhãn. Nếu không chủ đích che giấu nhãn cho mô hình sẽ dẫn đến rò rỉ nhãn trong quá trình huấn luyện. Label leakage, hay còn gọi là data leakage (rò rỉ dữ liệu), là vấn đề quan trọng trong lĩnh vực học máy (machine learning) và trí tuệ nhân tạo (AI). Điều này xảy ra khi thông tin từ tập dữ liệu huấn luyện (training data) chứa thông tin không hợp lệ hoặc tương lai mà không nên có trong mô hình dự đoán. Điều này có thể dẫn đến việc mô hình học máy hoạt động tốt bất thường trên tập dữ liệu huấn luyện nhưng lại kém hiệu quả khi áp dụng vào dữ liệu mới hoặc thực tế. Cụ thể, label leakage xảy ra khi thông tin về nhãn (labels) - kết quả mà mô hình đang cố gắng dự đoán - bị lộ vào quá trình huấn luyện mô hình. Đây là một ví dụ phổ biến của data leakage, và có thể khiến mô hình dự đoán chính xác một cách giả tạo, nhưng không thực sự hiểu được mối quan hệ giữa các đặc trưng (features) và nhãn. Trong trường đó, mô hình của sẽ trực tiếp học các nhãn quan sát được và bỏ qua các thông tin về cách hành vi gian lận ẩn phức tạp hơn, điều này dẫn đến nguyên nhân mô hình không thể được khái quát hóa trong việc dự đoán các

hành động gian lận trong tương lai. Do đó, đề án sử dụng mô hình học từ thông tin rủi ro của các giao dịch lân cận mỗi giao dịch thay vì chỉ học từ thông tin duy nhất nhãn của node đó. Cụ thể, chiến lược huấn luyện mô hình phát hiện gian lận với các node bị che được tận dụng. Mỗi bước huấn luyện, ta lấy mẫu ngẫu nhiên các nút theo batch bao gồm các nút trung tâm, cùng với các nút lân cận tương ứng với mỗi nút trung tâm. Sau đó, chuyển đổi nhãn $\hat{\mathbf{Y}}$ được quan sát một phần thành $\tilde{\mathbf{Y}}$ bằng cách che tất cả các nút embedding thông tin rủi ro của node trung tâm thành zero embedding và giữ cho các nút khác không thay đổi. Sau đó, hàm mục tiêu dự đoán $\hat{\mathbf{Y}}$ với X, $\tilde{\mathbf{Y}}$ và **A** đã cho là :

$$\mathcal{L} = -\frac{1}{|V|}\sum_{i=0}^{|V|}\Big[\mathbf{y}_i \cdot \log p\Big(\hat{\mathbf{y}}_i \mid \mathbf{X}, \tilde{\mathbf{Y}}, \mathbf{A}\Big) + (1-\mathbf{y}_i)\cdot \log\Big(1 - p\Big(\hat{\mathbf{y}}_i \mid \mathbf{X}, \tilde{\mathbf{Y}}, \mathbf{A}\Big)\Big)\Big] \tag{3.7}$$

Trong đó |V| đại diện cho số lượng nút trung tâm có nhãn bị che. Bằng cách này, ta có thể huấn luyện mô hình mà không bị rò rỉ thông tin rủi ro và trong quá trình suy luận, ta sử dụng tất cả các nhãn quan sát được $\hat{\mathbf{Y}}$ làm thuộc tính phân loại đầu vào để dự đoán rủi ro của các giao dịch ngoài bộ dữ liệu huấn luyện. Mục tiêu tối ưu hóa của mô hình là mô hình hóa các mẫu gian lận bằng thông tin thuộc tính giao dịch của các nút lân cận và thông tin thuộc tính.

b. Huấn luyện mô hình

Chia làm 2 nhánh xử lý :

- Training : Là quá trình huấn luyện mô hình phân loại các giao dịch là gian lận và giao dịch thông thường . bằng cách trích xuất thông tin về mối quan hệ của giao dịch với các user khác trong một mạng xã hội của giao dịch thông qua graph database . Giúp mô hình có thể học được mối quan hệ của các giao dịch gian lận từ những tài khoản khả nghi từ nó phát hiện được giao dịch gian lận

Graph embeddings học cấu trúc đồ thị để đưa ra biểu diễn bằng số cho mỗi nút. Giả sử có một đồ thị 5 tỷ nút. Nghĩa là ta có một ma trận kề 5 tỷ x 5 tỷ. Graph embeddings sẽ tổng hợp tất cả thông tin đó thành 50 số trên mỗi nút và những số đó có thể được đưa vào mô hình học máy sử dụng với các thuật toán. Điều này thường được sử dụng để đưa ra đề xuất cho các nút khác dựa trên hành vi của nút từ đó tìm được điểm tương đồng giữa các hành vi gian lận tài chính.

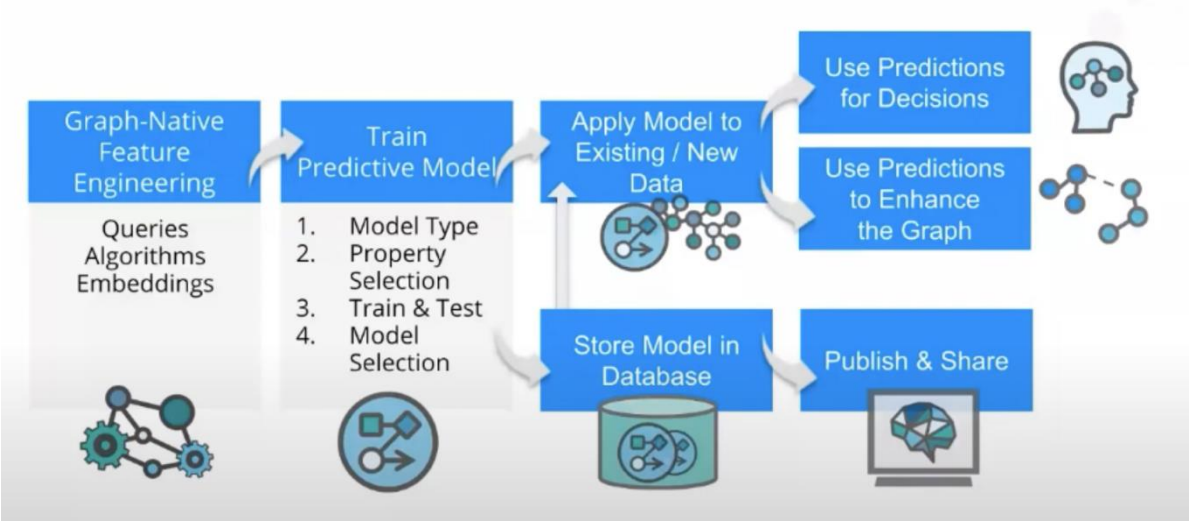
Ví dụ: các nút tương tự trong đồ thị các hành tư thuộc tính tương tự hành vi tấn công của kẻ tấn công trước đó. Chúng ta cũng không cần phải tự tay thiết kế nó. Sau đó đưa vào model Semi-Supervised Graphs Neural Network. Từ sẽ xác định điều gì quan trọng, điều gì bất thường, điều gì được nhóm lại với nhau và sau đó bạn có thể xem xét điều đó và thực hiện hành động.

```
In epoch:000|batch:0090, train_loss:0.483813, train_ap:0.3777, train_acc:0.9000, train_auc:0.7576
In epoch:000|batch:0100, train_loss:0.480433, train_ap:0.2500, train_acc:0.8125, train_auc:0.5071
In epoch:000|batch:0110, train_loss:0.479931, train_ap:0.2181, train_acc:0.8140, train_auc:0.4857
In epoch:000|batch:0120, train_loss:0.479617, train_ap:0.3639, train_acc:0.7308, train_auc:0.6410
In epoch:000|batch:0130, train_loss:0.474258, train_ap:0.2720, train_acc:0.8929, train_auc:0.7233
In epoch:000|batch:0140, train_loss:0.472477, train_ap:0.1882, train_acc:0.7778, train_auc:0.4527
In epoch:000|batch:0150, train_loss:0.469074, train_ap:0.1888, train_acc:0.8824, train_auc:0.6259
In epoch:000|batch:0160, train_loss:0.467906, train_ap:0.5306, train_acc:0.8214, train_auc:0.8109
In epoch:000|batch:0170, train_loss:0.464721, train_ap:0.3171, train_acc:0.8261, train_auc:0.7566
In epoch:000|batch:0180, train_loss:0.462771, train_ap:0.3845, train_acc:0.8431, train_auc:0.8256
In epoch:000|batch:0190, train_loss:0.458146, train_ap:0.2519, train_acc:0.9273, train_auc:0.8529
In epoch:000|batch:0200, train_loss:0.460737, train_ap:0.4758, train_acc:0.7442, train_auc:0.6818
In epoch:000|batch:0210, train_loss:0.460754, train_ap:0.2603, train_acc:0.8163, train_auc:0.5139
In epoch:000|batch:0220, train_loss:0.457072, train_ap:0.4105, train_acc:0.7736, train_auc:0.6646
In epoch:000|batch:0230, train_loss:0.455620, train_ap:0.2605, train_acc:0.8039, train_auc:0.6667
In epoch:000|batch:0240, train_loss:0.455732, train_ap:0.3351, train_acc:0.7407, train_auc:0.6036
In epoch:000|batch:0250, train_loss:0.454319, train_ap:0.2937, train_acc:0.9057, train_auc:0.7833
In epoch:000|batch:0260, train_loss:0.454490, train_ap:0.5516, train_acc:0.8723, train_auc:0.8455
In epoch:000|batch:0270, train_loss:0.453117, train_ap:0.5373, train_acc:0.7949, train_auc:0.6371
In epoch:000|batch:0280, train_loss:0.452304, train_ap:0.4849, train_acc:0.8571, train_auc:0.8605
```

Hình 3 . 8. Huấn luyện model

- Serving : Là quá trình triển khai mô hình sau khi training lên hệ thống để kiểm tra các giao dịch qua thông tin có được trong graph database

Deploy



Hình 3 . 9. Luồng triển khai

Sử API phẳng cho giúp dễ dàng mở rộng khả năng của nền tảng với nhiều trường hợp sử dụng. API sử dụng là RESTful.

3.1.3. Giám sát

Là thành phần hậu kiểm do các con người kiểm soát lắng nghe phản hồi từ phía khách hàng khi có giao dịch gian lận không được phát hiện bởi mô hình cũng như những giao dịch bị bình thường nhưng bị mô hình phân loại thành giao dịch gian lận . Dữ liệu sau khi được hậu kiểm sẽ được đưa vào training lại cho mô hình giúp độ chính xác của mô hình càng ngày càng cao .Với những giao dịch chưa rõ ràng vượt qua ngưỡng về rủi ro của giao dịch thì giao dịch sẽ được giữ lại .

GDS(Graph Data Science) hỗ trợ nhiều người dùng đồng thời làm việc trên cùng một hệ thống. Thông thường, GDS có thể sử dụng nhiều bộ nhớ /hoặc nhiều lõi CPU để thực hiện tính toán . Để biết liệu thời điểm hợp lý để người dùng chạy quy trình GDS, ta nên biết dung lượng hiện tại của hệ thống lưu trữ Neo4j và GDS, cũng như khối lượng công việc GDS hiện tại trên hệ thống. Đồ thị và mô hình không được chia sẻ giữa những người dùng không phải quản trị viên, tuy nhiên, người dùng GDS trên cùng một hệ thống sẽ chia sẻ dung lượng

của nó. Để có thể xem tổng quan về dung lượng hiện tại của hệ thống và khối lượng công việc phân tích của nó, ta có thể sử dụng quy trình gds.alpha.systemMonitor. Nó sẽ cung cấp thông tin về dung lượng của phiên bản JVM của DBMS xét về bộ nhớ và lõi CPU, cũng như tổng quan về các tài nguyên được sử dụng bởi các thủ tục GDS hiện đang chạy trên hệ thống.

3.1.4. Bộ xử lý trung tâm

Có các nhiệm vụ :

- Truy vấn dữ liệu trong graph database phục vụ cho quá trình huấn luyện và dự đoán của mô hình
- Ghi nhận dữ liệu được hậu kiểm từ Monitor ghi lại dữ liệu vào graph database
- Training mô hình và Deploy mô hình AI

3.2. KẾT QUẢ

3.2.1. Bộ dữ liệu

Trong bài nghiên cứu này em sử dụng dữ liệu S-FFSD được thu thập từ dữ liệu đối tác về các công ty tài chính trong bài báo để training, đánh giá kết quả của nghiên cứu thử nghiệm.

Bảng 3. 1. Mô tả dữ liệu S-FFSD

Name	Type	Range	Note
Time	np.int32	0 đến N	N Biểu thị số lượng giao dịch.
Source	string	S_0 đến S_{ns}	ns biểu thị số lượng người giao dịch
Target	string	T_0 đến T_{nt}	nt biểu thị số lượng người nhận giao dịch
Amount	np.float32	0.00 đến np.inf	Số tiền mỗi giao dịch

Location	string	L_0 đến L_{nl}	nl biểu thị số lượng vị trí giao dịch
Type	string	TP_0 đến TP_{np}	np biểu thị số lượng các loại giao dịch khác nhau.
Label	np.int32	0 đến 2	0 là giao dịch bình thường, 1 giao dịch gian lận, 2 là giao dịch chưa được phân loại

Bảng 3. 2. Thống kê dữ liệu S-FFSD

Label	Quantity
0	24387
1	5256
2	48238
Total	77881

Từ dữ liệu ta có quan sát và phân tích sau:

- Dữ liệu phân phối cho thấy dữ liệu phân phối không đều do giao dịch tài chính rất nhiều nên không thể gán nhãn hết được nên tỉ lệ giao dịch không được gán nhãn lớn hơn nhiều so với các giao dịch được gán nhãn
- Dữ liệu ngoài việc giao dịch được gán nhãn ít hơn so với giao dịch được gán nhãn tỉ lệ giao dịch thì tỉ lệ giao dịch gian lận so với giao dịch bình thường có gây nên mất cân bằng nhãn lớn làm giảm hiệu quả các mô hình dự đoán

Xây dựng từ dữ liệu dạng bảng chuyển sang dữ liệu dạng đồ thị phù hợp với mô hình GTAN. Sử dụng thư viện Networkx và DGL để biến đổi từ dữ liệu dạng bảng sang dữ liệu dạng đồ thị

Bảng 3. 3. Thống kê dữ liệu đồ thị S-FFSD

Node	1,820,840
Edge	31,619,440
Fraud	33,858
Legitimate	141,861
Unlabeled	1,645,121

3.2.2. Đánh giá kết quả thực nghiệm

Sử dụng chỉ số Precision, Recall, F1-Score và Accuracy để đánh giá độ hiệu quả của mô hình training.

- Precision là tỷ lệ giữa số sample được tính là True Positive (TP) với tổng số sample được phân loại là Positive (bằng chính TP + FP). Precision càng lớn có nghĩa là độ chính xác của các điểm tìm được càng cao.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{3.8}$$

- Recall là tỷ lệ giữa các điểm positive thực được nhận đúng trên tổng điểm positive thực. Recall càng cao tỉ lệ bỏ sót các sample positive thực thấp.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3.9}$$

- Trong thực tế nếu ta điều chỉnh model để tăng Recall quá mức có thể dẫn đến Precision giảm và ngược lại, có điều chỉnh model để tăng Precision có thể làm giảm Recall. Để cân bằng 2 đại lượng này ta sử dụng F1 Score. F1 càng cao thì càng tốt. Khi lý tưởng nhất thì F1 = 1 (khi Recall = Precision=1).

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} \quad (3.10)$$

- Accuracy là tỷ lệ model dự đoán đúng trên tổng mẫu các dự đoán

$$accuracy = \frac{true\ positives + true\ negatives}{true\ positives + true\ negatives + false\ negatives + false\ positives} \quad (3.11)$$

Chia dữ liệu huấn luyện và dữ liệu kiểm thử với với tỉ lệ 80% train và 20% testing cho model GTAN .

Bảng 3. 4. Kết quả so sánh các mô hình trên nhãn Fraud

	precision	recall	f1-score	support
0	0.89	0.99	0.94	4878
1	0.93	0.43	0.58	1051
accuracy			0.89	5929

So sánh kết quả của model GTAN với các mô hình XGBClassifier, SVM, RandomForestClassifier, KNN

Bảng 3. 5. Kết quả so sánh các mô hình trên nhãn Fraud

Model	precision	recall	f1-score	accuracy
GTAN	**0.93**	**0.43**	**0.58**	**0.89**
SVM	0.73	0.39	0.51	0.87
RandomForestClassifier	0.92	0.25	0.39	0.86
KNN	0.59	0.54	0.57	0.85

Từ Bảng 5 ta có quan sát và phân tích sau:

- Thấy rằng phương pháp GTAN cho kết quả tốt nhất về độ chính xác . Với chỉ số precision ta thấy GTAN dự đoán các giao dịch gian lận ít nhầm lẫn

với các giao dịch bình thường nhất . Recall thì cho thấy GTAN phát hiện được nhiều giao dịch trên tổng số các giao dịch gian lận .

- XGBClassifier không thể học được các đặc trưng các giao dịch gian lận nên không thể phát hiện được cá giao dịch gian lận
- SVM có khả năng phát hiện giao dịch gian lận ở mức trung bình
- RandomForestClassifier dự đoán nhãn gian lận tỉ lệ nhầm tương đối thấp nhưng lại ít phát hiện được giao dịch gian lận trên tổng số giao dịch gian lận
- KNN có tỷ lệ dự đoán nhầm nhãn gian lận và nhãn bình thường tương đối nhiều mặc dù tỷ lệ phát hiện được số giao dịch gian lận tương đối tốt

3.3. Kết luận chương

Nội dung chương này đã trình bày chi tiết thiết kế hệ thống để tích mô hình GTAN. Tích hợp Graph Database (cơ sở dữ liệu đồ thị) để trích xuất, lưu trữ và xử lý dữ liệu để huấn luyện mô hình và đưa ra kết quả dự báo cho hệ thống về giao dịch bất thường tiềm ẩn hành vi gian lận hay không. Huấn luyện và đánh giá các mô hình GTAN so với một số thuật toán AI khác.

KẾT LUẬN

Với sự phát triển mạnh mẽ của các nền tảng số các hoạt động giao dịch, thanh toán trực tuyến phục vụ cho nhu cầu công việc, giải trí, sinh hoạt, .. ngày càng trở nên phổ biến thông dụng với đại bộ phận người dân. Sự thay đổi thói quen hành vi là trên là tất yếu nhất là trong thời đại số và với bước nhảy là đại dịch covid 19 dẫn đến tiến trình này càng được rút ngắn lại. Ngoài những ưu điểm như tiện lợi, nhanh chóng cho cả người sử dụng và các nhà cung cấp dịch vụ nhưng đồng thời cũng phải sinh rất nhiều hành vi lạm dụng, gian lận, lừa đảo nhằm mục đích xấu. Nên hệ thống phát hiện bất thường trong giao dịch là không thể thiếu trong các nền tảng số hiện nay nhất là những nền tảng liên quan đến thanh toán, giao dịch của khách hàng. Nhưng hệ thống này cũng phải đối mặt với vấn đề về cân bằng giữa trải nghiệm của khách hàng, rủi ro phải chấp nhận của nhà cung cấp, độ chính xác của những dự đoán nhất là với độ lệch lớn giữa số lượng giao dịch bình thường và giao dịch bất thường chứa yếu tố gian lận.

Do đó trong nghiên cứu của đề án mô hình mạng chú ý theo thời gian có kiểm soát (GTAN) để phát hiện gian lận. Mô hình liên kết dữ liệu được gắn nhãn và không được gắn nhãn thông qua các mối quan hệ của chúng. Học phân loại nhất quán với các nhãn của dữ liệu đã được gắn nhãn bằng cách đề xuất hàm mất mát phân loại và mặt khác làm cho kết quả phân loại cho các đỉnh trong đồ thị tương tự giống nhau bằng cách đề xuất mất mát thông tin dựa trên đồ thị. Cơ chế attention của mô hình GTAN giúp học đặc trưng đồ thị tốt hơn. Theo thực nghiệm phương pháp này đạt được kết quả tốt hơn so với các thuật toán AI phổ biến khác. Mô hình GTAN có thể cho biết các đặc trưng quan trọng các giao dịch bất thường rủi ro cao. Bằng việc dựa vào mối quan hệ khác nhau của những người dùng có thể mở rộng độ phủ cho việc sự chênh lệch giữa giao

dịch bình thường và giao dịch bất thường nên mô hình cho khả năng phát hiện nhiều giao dịch bất thường gian lận hơn.

DANH MỤC TÀI LIỆU THAM KHẢO

Tiếng Anh

[1] Daixin Wang, Jianbin Lin, Peng Cui , Quanhui Jia , Zhen Wang , Yanming Fang , Quan Yu , Jun Zhou , Shuang Yang , Yuan Qi (2020). "A Semi-supervised Graph Attentive Network for Financial Fraud Detection". *Ant Financial Services Group, China,Department of Computer Science and Technology, Tsinghua University, China*, 1-3

[2] Shaosheng Cao, Xinxing Yang, Cen Chen, Jun Zhou, Xiaolong Li, and Yuan Qi (2019). "TitAnt: Online Real-time Transaction Fraud Detection in Ant Financial", *Cornell University* , 2-7

[3] Sheng Xiang, Mingzhi Zhu, Dawei Cheng, Enxia Li, Ruihui Zhao, Yi Ouyang, Ling Chen, Yefeng Zheng (2023)."Semi-supervised Credit Card Fraud Detection via Attribute-Driven Graph Representation".*Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, Australia Department of Computer Science and Technology, Tongji University, Shanghai, China,Shanghai Artificial Intelligence Laboratory, Shanghai,China, Tencent Jarvis Laboratory, Shenzhen, China*, 2-5