Revista e-TECH: Tecnologias para Competitividade Industrial

DOI: https://doi.org/10.18624/etech.v15i3.1198

Detecção de fraudes no segmento de crédito financeiro utilizando aprendizado de máquina: uma revisão da literatura

Fraud detection in the financial credit segment using machine learning: a literature review

Emerson Martins[1*], Napoleão Verardi Glegale [2].

[1] Centro Paula Souza, São Paulo, Brasil

[2] PUC SP, São Paulo, Brasil

[*]Correspondente: emerson.emtech@gmail.com

Resumo

A fraude financeira é uma ameaça cada vez maior, com consequências negativas tanto para o setor financeiro quanto para própria sociedade. Embora a mineração de dados tenha se mostrado como uma ferramenta útil na detecção de fraude de cartão de crédito, ela também tem enfrentado desafios, pois os perfis de comportamentos normais e fraudulentos mudam constantemente, em que os tipos das transações fraudulentas se aproximam muito das legítimas. Esta pesquisa tem como objetivo avaliar quais tipos de algoritmos são mais utilizados na detecção de fraudes no uso de cartão de crédito. Para tanto, foi empregado o método de revisão sistemática da literatura, com base no protocolo PRISMA-P. Foram identificados como algoritmos mais utilizados o NN (*Neural Network Feed-Forward*), NB (*Naive Bayes*), RF (*Random Florest*) e o SVM (*Support Vector Machines Based*). O presente estudo fornece um guia quanto aos métodos que têm alto potencial para atingir a detecção de fraude com cartão de pagamento.

Palavras-chave: aprendizado de máquina; crime financeiro; detecção de fraude.

Abstract

Financial fraud is a growing threat with negative consequences, both for the financial sector and for society itself. While data mining has proven to be a useful tool in detecting credit card fraud, it has also faced challenges as the profiles of normal and fraudulent behavior are constantly changing, where the types of fraudulent transactions closely approximate legitimate ones. This research aims to evaluate which types of algorithms are most used in the detection of fraud in the use of credit cards. For this purpose, a systematic review of the literature (SRL) based on the PRISMA-P protocol was used. The most used algorithms were NN "Neural Network Feed-Forward", NB "Naive Bayes", RF "Random Forest" and SVM "Support Vector Machines Based". The present study provides a guide as to methods that have high potential to achieve payment card fraud detection.

Keywords: machine learning; financial crime;fraud detection.

1. INTRODUÇÃO

Com o aumento do comércio eletrônico na última década, o uso de cartões de pagamento aumentou drasticamente (SMADI *et al.,* 2021) e, para evitar perdas com fraudes, dois mecanismos podem ser usados: prevenção e detecção de fraude. A prevenção de fraude é um método proativo, que impede que a fraude aconteça em primeiro lugar. Por outro lado, a detecção de fraude é necessária quando uma transação fraudulenta já está em andamento através de um fraudador. A fraude está relacionada ao uso ilegal das informações de crédito de um titular, na realização de compras sem o seu consentimento. As transações com cartão de pagamento podem ser realizadas fisicamente ou digitalmente (ADEWUMI *et al.,* 2017). Dentro das transações físicas, o cartão de pagamento está envolvido durante a transação; já nas transações digitais, isso pode acontecer via telefone ou internet. Nesses casos, o titular fornece o número, mês e ano de validade, nome completo e número de verificação contidos no cartão.

Os fraudadores são favorecidos pelo uso da internet, pois sua identidade e localização ficam ocultas.

O aumento com a fraude na utilização de cartão de pagamento tem um grande impacto no setor financeiro. A fraude global de cartão de crédito em 2015 atingiu um nível impressionante de US $ 21,84 bilhões (PYMNTS, 2016).

A contribuição deste artigo é avaliar uma variedade de algoritmos de aprendizado de máquina utilizados para detectar transações fraudulentas com a utilização de cartão de pagamento, com base no histórico de compras do próprio cartão e seu titular. Para tal, este artigo foi dividido nas seguintes seções: 2 – Aprendizado de máquina e sua aplicabilidade no segmento financeiro; 3 – Metodologia de pesquisa aplicada; 4 – Revisão qualitativa da literatura; 5 – Resultados da revisão de literatura; 6 – Conclusão.

2. APRENDIZADO DE MÁQUINA E SUA APLICABILIDADE NO SEGMENTO FINANCEIRO

O aprendizado de máquina (ML – *Machine Learning)* é o estudo de algoritmos de computador que se aprimoram automaticamente com a experiência. É tratado com uma subárea da inteligência artificial (IA). Algoritmos de aprendizado de máquina constroem um modelo baseado em dados de amostra, conhecidos como "dados de treinamento", a fim de fazer previsões ou decisões sem serem explicitamente programados para isso. Os algoritmos de aprendizado de máquina são usados em uma ampla variedade de aplicações, como filtragem de e-mail e visão computacional, em que é difícil ou inviável desenvolver algoritmos convencionais para realizar as tarefas necessárias (MAXWELL *et al*., 2015).

Conforme Randhawa *et al*. (2018), a demanda por sistemas financeiros mais eficientes e confiáveis tem levado a um crescente aprimoramento dos sistemas de análise de dados utilizados pelos agentes financeiros. Essa sofisticação vem impulsionando o desenvolvimento de sistemas computacionais capazes de analisar grandes quantidades de dados rapidamente e de gerar relatórios para auxiliar na tomada de decisões. A maioria das principais instituições financeiras, tanto internacionais quanto nacionais, está utilizando algoritmos de ML para a análise de seus dados. Em várias aplicações na área de finanças, o uso de ML tem possibilitado ganhos financeiros expressivos. Entre os principais problemas na área financeira em

que ML tem sido utilizada com sucesso, podem ser citados: I) análise de risco de crédito; II) previsão de falências; III) previsão de cotações de moedas e de ações; IV) segmentação de mercados; V) detecção de fraudes.

Análise de risco de crédito é possivelmente o problema financeiro em que ML tem sido mais popular. Esse problema pode estar associado tanto à pessoa jurídica quanto à pessoa física. A análise de risco de crédito para pessoa física é geralmente utilizada quando uma pessoa solicita um cartão de crédito, cheque especial, financiamento ou crediário. Nesse caso, a instituição que faz a análise de crédito geralmente utiliza um conjunto de dados de aplicações passadas, que possuem, para identificar o perfil de risco do cliente e classificá-lo. A classificação pode ser tanto uma pontuação dada por um analista de crédito, que pode informar se deve ou não ser dado o crédito, quanto uma análise do histórico financeiro do cliente. O histórico financeiro pode conter dados como o número de vezes que o cliente se tornou inadimplente. Observe que, no primeiro caso, o entendimento dos dados é subjetivo. Diferentes analistas de crédito podem atribuir pontuações diferentes à aplicação. Quando ML é utilizada nesses problemas, um algoritmo supervisionado para classificação utiliza esse conjunto de dados para induzir um modelo ou hipótese, que será depois utilizado para os novos clientes (AWOYEMI *et al*., 2017).

Embora esse seja geralmente um problema de classificação binária, ele pode apresentar mais de duas classes, que podem, inclusive, formar um ranking de classes para diferentes perfis de clientes. Uma dificuldade geralmente encontrada é que os conjuntos de dados em geral têm muito mais dados de clientes que foram aprovados (ou têm um bom histórico financeiro) que o contrário. Para lidar com essa limitação, técnicas de pré-processamento para dados desbalanceados podem ser utilizadas (AWOYEMI *et al*., 2017).

Observe que, nessas aplicações, os erros podem ter custos diferentes. Pode ser menos custoso rejeitar o crédito a um cliente que não ficaria inadimplente que o contrário. Algoritmos de agrupamento de dados também podem ser empregados para encontrar diferentes perfis de clientes, que seriam os grupos encontrados pelo algoritmo. Diferentes produtos poderiam ser oferecidos a diferentes grupos de clientes (RANDHAWA *et al*., 2018).

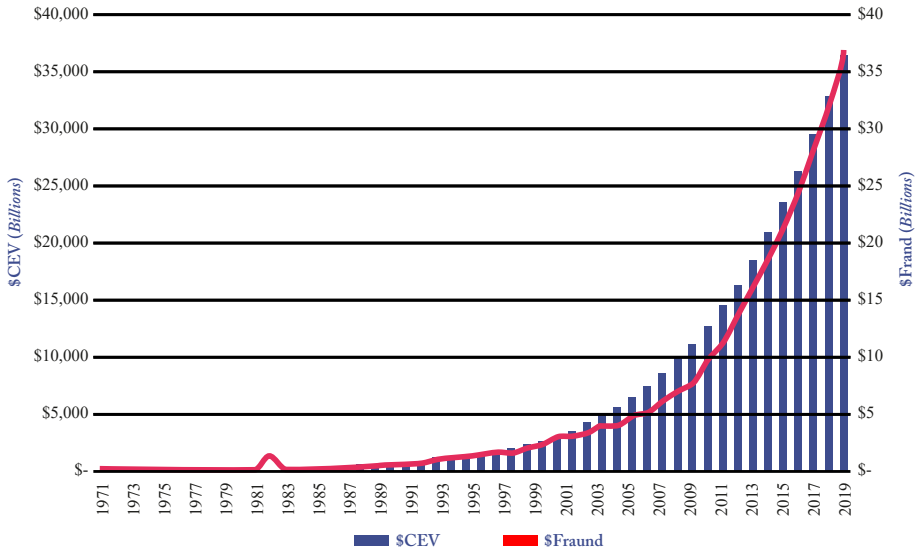
Na detecção de fraudes, o objetivo é detectar transações atípicas. Fraudes ocorrem não apenas em transações financeiras, mas também no uso de energia, na compra de produtos, na utilização de recursos sociais, no acesso a redes de computadores ou, até mesmo, em transações contábeis para manipular os resultados financeiros da empresa. Algoritmos de ML são geralmente utilizados para prever ou classificar quando uma dada transação é uma fraude – trata-se de um problema de classificação binária. Assim como os dados de análise de crédito, os conjuntos de dados de detecção de fraudes apresentam muito mais dados de transações corretas do que dados que representam fraudes, o que dificulta a análise. Outra dificuldade é que novos tipos de fraude são constantemente criados, o que torna necessária a adaptação contínua dos modelos. Dessa forma, a detecção de fraudes também pode ser vista como um problema de fluxo contínuo de dados (XUAN *et al*., 2018).

Enquanto dados de análise de crédito estão disponíveis em vários repositórios públicos, como o "Portal Brasileiro De Dados Abertos", o mesmo não ocorre com dados de crédito. A razão é simples: o portador desses dados não quer liberá-los para que suas vulnerabilidades passadas (e talvez presentes) não sejam conhecidas.

Infelizmente, a sociedade em geral entende a fraude de cartão de pagamento como um crime menor, em que seus efeitos são mitigados pelo reembolso de seu emissor; o impacto individual para a vítima de fraude é suavizado. Há uma crença comum de que a fraude com pagamento afeta apenas bancos, grandes empresas e governo, e que a fraude seja realizada por indivíduos e normalmente por ‘‘hackers” (CASTLE, 2008).

Em 2017, foram realizadas 349 bilhões de transações com cartão de pagamento com um “Volume de Despesas do Cartão” ($ CEV – *Card Expenditure Volume*) em $ 26,3 trilhões e perda direta com fraude ($ *fraud*) de $ 24 bilhões. A mesma tecnologia que permitiu pagamentos sem dinheiro está alimentando o crescimento exponencial da fraude com cartões de pagamento. A Fig. 1 demonstra como o volume de pagamento com cartão de crédito vem crescendo em todo o mundo e como a fraude acompanha esse crescimento (RYMAN-TUBB *et al*., 2018).

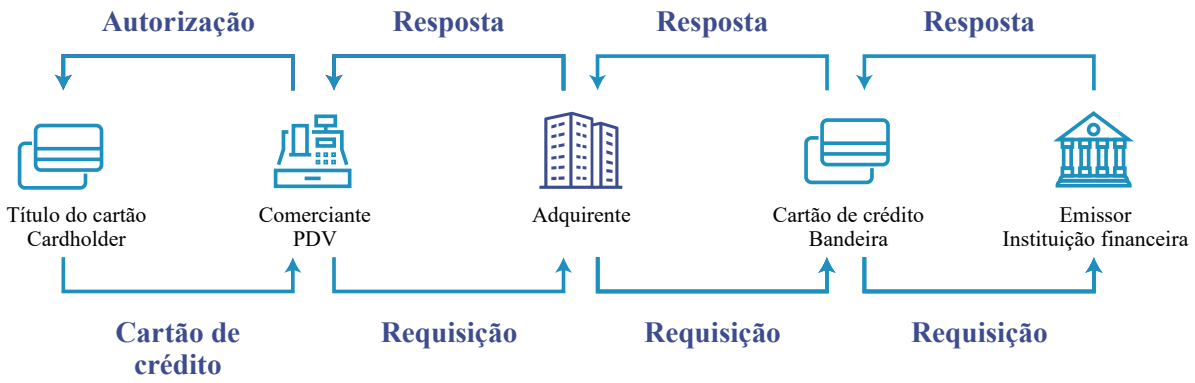
Figura 1 - A evolução da fraude por valor



Fonte: Ryman-tubb (2018)

Existem vários participantes que estão envolvidos quando ocorre uma transação com cartão de crédito (Fig. 2). Quando o comerciante recebe pagamento de um cliente através de cartão de crédito, os detalhes da transação são enviados ao adquirente do comerciante. O adquirente, então, solicita autorização do emissor do cartão obtendo a resposta se a transação foi aprovada ou reprovada. Essa resposta é então retornada ao comerciante para concluir a transação. Se a transação for autorizada, então a venda é concluída e as mercadorias são despachadas. Podemos exemplificar essa operação da seguinte forma: Portador (passa o cartão) □ Adquirente recebe transação (p. ex.: Cielo) e valida a bandeira emissora (p. ex.: Visa) □ Instituição emissora recebe a transação (Bancos), valida limite do cliente, retornando à autorização ou não através do fluxo inverso.

Figura 2 - Processo de autorização na utilização de cartão de crédito



Fonte: Dos autores (2022)

Para determinar se uma transação de cartão de pagamento é autorizada, vários processos são realizados, entre eles está o *Fraud Management System* (FMS). O FMS recebe a transação, toma uma decisão usando alguma forma de classificador e retorna isso como parte do processo de autorização. Se a transação for considerada suspeita, normalmente é bloqueada ou recusada, e um tíquete de fraude é criado. Esse bilhete de fraude contém informações suficientes para um revisor humano entender a transação e, em seguida, tomar uma decisão. Na maioria das organizações, uma equipe de revisores verifica tíquetes de fraude, e é realizada uma investigação que pode incluir o contato com o titular do cartão ou comerciante (HAND *et al*., 2008).

Por conta de enormes prejuízos com fraudes, os principais *players* de mercado – Visa, MasterCard, American Express, Discover Financial Services, JCB International – criaram, em 2004, um padrão global para proteger informações confidenciais de cartões de pagamento contra roubo. O padrão PCI DSS (*Payment Card Industry Data Security Standards* – Conselho de padrões de segurança do setor de cartões de pagamento), mencionado por Morse *et al. (*2008), possui um conjunto de regras para gerar maior proteção nas transações via internet e em lojas físicas, em que tais transações devem ser sempre realizadas em ambiente seguro, acompanhadas com certificações digitais SSL. Tal padrão deve ser aplicado por todas as entidades que processam dados com cartões de pagamento.

3. METODOLOGIA

Para esta pesquisa do tipo descritiva e qualitativa, foi realizada revisão sistemática da literatura sobre a utilização de ML na detecção de fraudes no segmento de crédito financeiro.

Para realizar a revisão sistemática da literatura sobre a utilização de ML para detecção de fraudes no segmento de crédito financeiro, foi utilizado o protocolo PRISMA-P. Tal protocolo tem como objetivo apoiar os pesquisadores a melhorar o relato de revisões sistemáticas e meta-análises, filtrando o número de publicações com maior relevância ao tema pesquisado (MOHER, 2015).

Na etapa de identificação, foi realizada busca das publicações nas bases de dados Google Scholar, Microsoft Academic e Scopus, com as seguintes palavras-chaves: "*Machine Learning*", "*Financial Crime*" e "*Fraud Detection*", em que foram retornadas 1.128 publicações. Para a pesquisa, foi utilizado o período de 2017 a 16/04/2021, sem qualquer outro tipo de filtro.

Na etapa de triagem, foram removidas 28 publicações em duplicidade, 105 Livros, 23 Citações, 236 Artigos de Jornais e 21 publicações classificadas como "Outros" pelas bases pesquisadas, totalizando 413 registros excluídos; dessa forma, resultaram 715 registros selecionados para a próxima etapa.

Na etapa de elegibilidade, foram excluídos 613 artigos sem H-Index, 4 artigos cujo acesso não é público e 80 artigos com baixa aderência ao tema da pesquisa, pois tratavam do tema, porém não apresentaram conjunto de dados financeiros, nem a utilização de parâmetros quantitativos para comprovar a eficiência na detecção de fraude através dos algoritmos apresentados.

Foram, então, selecionadas 19 publicações para seguirem para a última etapa, sendo que, após a leitura dos mesmos, foram mantidas somente 9 publicações para análise qualitativa, uma vez que os demais artigos não apresentaram métricas de desempenho equivalentes dos algoritmos apresentados.

4. REVISÃO QUALITATIVA DA LITERATURA

Pesquisadores relataram que a identificação de um melhor algoritmo na detecção de fraude na utilização de cartão de crédito é severamente limitada (Adewumi *at el*., 2017), devido a questões de segurança e privacidade, especialmente por questões relacionadas à violação de dados. Mesmo quando os conjuntos de dados estão disponíveis na indústria, os resultados são censurados, dificultando a avaliação do trabalho como um todo. Alguns pesquisadores nessa pesquisa tiveram que usar conjuntos de dados sintéticos que tentam replicar dados do mundo real, por exemplo, o artigo [5] (SADGALI *at el.,* 2019). Como os perfis de comportamento genuínos e fraudulentos mudam com o tempo, os dados sintéticos podem ser insuficientes. Portanto, os resultados relatados nessa pesquisa podem não ser confiáveis quando dimensionados para conjuntos de dados maiores. Isso destaca a incapacidade da comunidade acadêmica de demonstrar o impacto de forma realista para a indústria.

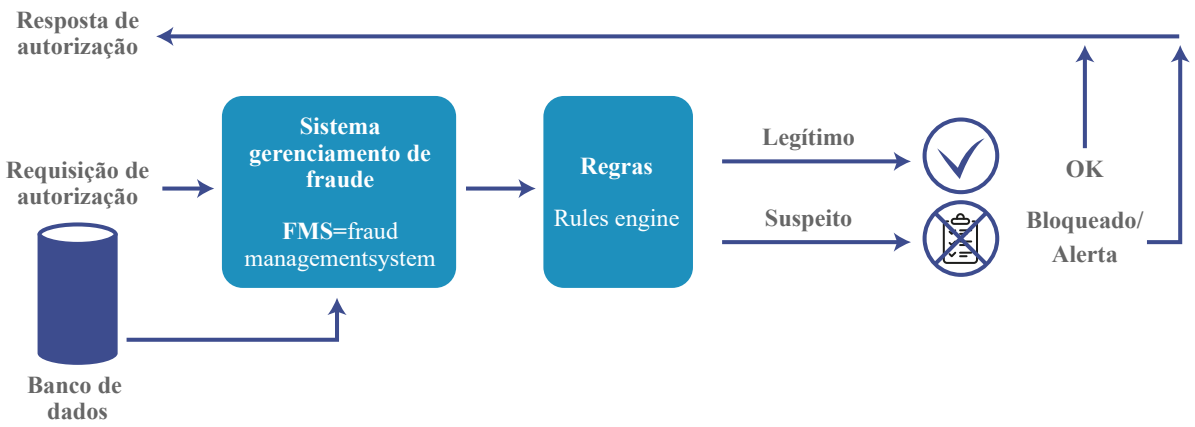
Os dados mantidos em cada transação, incluindo o CHD (*Cardholder Data*) – o titular do cartão e o comerciante –, são sensíveis. É simples usar esses dados para cometer fraude. Isso torna difícil para os processadores de pagamento fornecer dados para os pesquisadores avaliarem novos métodos de detecção. Métodos de ofuscação podem ser usados nos dados, mantendo suas relações, mas esse processo requer do detentor dos dados a certeza de que os dados originais não podem ser recriados ou alterados (SHOKRI, 2015). Existem leis em diferentes jurisdições que proíbem tais dados de deixar suas fronteiras, como a GDPR (Regulamento Geral de Proteção de Dados – União Europeia, 2016) e a LGPD (Lei Geral de Proteção de Dados – Brasil 2018).

É necessário entender que os dados disponíveis para um FMS dependem de qual participante de pagamento iniciou a transação. Um comerciante só tem dados sobre as transações que ocorreram em seu estabelecimento e não tem informações sobre outras transações que tenham sido realizadas pelo titular do cartão. O emissor só tem dados das transações que foram realizadas pelo cartão emitido ao titular do cartão e não tem informações sobre quaisquer outras transações que foram realizadas por outros meios de pagamentos por seus clientes. O adquirente normalmente tem apenas informações do comerciante juntamente com as informações que ele mantém, como os dados do aplicativo original e estatísticas sobre suas transações durante um período, ou seja, os dados são espalhados entre muitos sistemas de computador interconectados. Esse é um desafio considerável para comunidade de pesquisa.

Os vetores de fraude são dinâmicos à medida que os criminosos alteram o seu *modus operandi*, portando, é argumentado que o desvio de conceito dentro dos dados disponíveis é significativo e que as abordagens FMS que não levam isso em consideração se tornarão menos eficazes e, portanto, as perdas e os custos operacionais aumentarão significativamente. Há um evento perturbador na indústria de pagamentos que está criando vetores de fraude desconhecidos, que estão mudando a uma taxa mais rápida do que tem sido observado desde a introdução dos cartões de pagamento (CHOO *et al.*, 2007). Esse evento é devido ao crescimento exponencial das novas tecnologias e dispositivos como: smartphones, e-commerce, pagamentos sem contato, crimes cibernéticos – incluindo grandes violações de dados, computação em nuvem e moedas virtuais. À medida que o crime migra devido a essas tecnologias, isso ocorrerá mais rapidamente do que no passado, devido às novas tecnologias envolvidas. As formas tradicionais de fraude estão dando lugar a criminosos altamente versados em informática, os quais vivem em uma época de alta comunicação de tecnologia, com um estilo de vida voltado para a tecnologia e com uso intensivo das redes sociais. Os vetores de fraude mais sofisticados e sutis estão emergindo, como criminosos que começam usar a inteligência artificial e o próprio aprendizado de máquina para fins ofensivos (DVORSKY, 2017).

A perda por fraude é incorrida no momento da transação para emissores e comerciantes. Portanto, para a proteção ser eficaz, a fraude precisa ser detectada em tempo real. Um FMS em tempo real é ilustrado na (Fig. 3). Ele recebe uma transação e, em seguida, toma uma decisão como parte do fluxo de autorização e retorna essa decisão de aceitar / bloquear / recusar / alertar a transação, como uma mensagem de resposta. A funcionalidade em tempo real é particularmente importante, em que uma transação de cartão pode ser interrompida durante a autorização, com base na saída de um processo de decisão de fraude. Uma transação ocorre em um espaço de tempo e faz parte de uma sequência e, portanto, pode ser considerada um fluxo de dados. A natureza temporal e sequencial das transações é conhecida pelos revisores por conter informações importantes para a detecção de fraude.

Figura 3 - Real-time FMS



Fonte: Dos autores (2022)

O objetivo da pesquisa é comparar e classificar os algoritmos aplicados no processo de detecção de fraude com a utilização de cartão para pagamento, identificando na indústria quais são algoritmos mais utilizados atualmente.

O artigo [1], "*Credit card fraud detection using machine learning techniques: A comparative analysis",* investiga o desempenho de NB (*Naive Bayes*), KNN (*K-Nearest Neighbor*) e LR (*Logistic Regression*) em um conjunto de dados de fraude de cartão de crédito altamente distorcido. Tal conjunto de dados é proveniente de portadores de cartões europeus contendo 284.807 transações. Uma técnica híbrida de subamostragem e sobre amostragem é realizada nos dados distorcidos. As três técnicas são aplicadas no estado bruto e nos dados pré-processados. O trabalho é implementado em Python. O desempenho das técnicas é avaliado com base na acurácia, sensibilidade, especificidade, precisão, coeficiente de correlação de Matthews MCC (*Matthews correlation coefficient*) e taxa de classificação equilibrada. Os resultados mostram uma ótima precisão para NB, KNN e LR, em que foram apresentados 97,92%, 97,69% e 54,86%, respectivamente, porém, os resultados comparativos mostram que KNN tem melhor desempenho do que NB e LR.

No artigo [2], "*Credit card fraud detection using AdaBoost and majority voting"*, o autor utilizou o coeficiente de correlação de Matthews (MCC) para medir a performance dos algoritmos quanto à sua capacidade de identificar transações verdadeiramente fraudulentas e transações falso-positivas, ou seja, transações que inicialmente parecem ser fraudulentas, mas não são. O MCC mede a qualidade de um problema de duas classes, que leva em consideração verdadeiro- e falso-positivos e negativos. É uma medida equilibrada, mesmo quando as classes são de tamanhos diferentes.

A rede neural NN (*Network Neural Feed--Forward)* em conjunto com a NB alcançaram o melhor resultado MCC, com um coeficiente de 0,823.

O artigo [3], "Random forest for credit card fraud detection", lida com a detecção de fraude com dois tipos de algoritmos de floresta aleatória: RF (Random Forest Random-tree-based e Random Forest Classification-and-Regression-Tree CART-based). Os dados utilizados são de uma empresa de comércio eletrônico da China. O conjunto original de dados contém mais de 30.000.000 transações individuais. Cada registro de transação consiste em 62 valores de atributos, como tempo de transação, local e quantidade. Cada registro é rotulado por Fraude ou Legal. Como exigido pela empresa, os detalhes dos atributos do conjunto de dados não são divulgados. No conjunto de dados, apenas cerca de 82.000 transações foram rotuladas como fraude, o que significa que a taxa de fraude de 0,27%, sendo que problema de desequilíbrio do conjunto de dados deve ser levado em consideração. O Quadro 1 mostra os resultados produzidos pela Random--tree-based e CART-based. Embora a precisão da CART-based seja um pouco pior, a acurácia, sensibilidade e medida F (F-Measure) é muito melhor. Obviamente, o desempenho abrangente de CART-based é muito mais adequado para aplicação nesse subconjunto de experimentos.

Uma única taxa de acurácia, *Accuracy*, não é suficiente para medir o desempenho de uma modelo de floresta aleatória, quando os dados estão significativamente desequilibrados, portanto, faz-se necessário considerar outras medidas, as quais são listadas no Quadro 2, onde o **Positivo** corresponde a instâncias de fraude e o **Negativo** corresponde a instâncias normais, ou seja, legítimas. A taxa de precisão, *Precision*, é uma medida de resultado de predição, já a taxa de sensibilidade, *Recall*, mede a taxa de detecção de todos os casos de fraude. *F-Measure* é a média harmônica de *Recall* e *Accuracy.*

Quadro 1 - Resultados dos dois tipos de florestas aleatórias

MEASURE MODELOS	ACCURACY	PRECISION	RECALL	F-MEASURE
Random-tree-based	91,96%	90,27%	67,89%	0,7811
CART-Based	96,77%	89,46%	95,27%	0,9601

Fonte: XUAN *et al.* (2018)

Quadro 2 - Possíveis combinações na classificação da transação

REAL PREDICT	POSITIVE	NEGATIVE
Positive	True Positive	False Positive
Negative	False Negative	True Negative

Fonte: XUAN *et al.* (2018)

Através do Quadro 3, é observado melhor desempenho no algoritmo *CART-based*. No artigo [4], *"A survey of machine-learning and nature-inspired based credit card fraud detection techniques"*, as técnicas pesquisadas revelam que vários algoritmos de ML e NI (*Nature Inspired*) têm sido usados para lidar com detecção de fraudes ao se utilizar cartão de pagamento. É indicado que HMM, NN, SVM "*Support Vector Machines Based Techniques"*, AIS (*Artificial Immune System Based Techniques*) e GA (*Genetic Algorithm Based Techniques*) são as técnicas mais usadas no domínio de detecção de fraude de cartão de crédito. Além disso, entre esses algoritmos amplamente utilizados, os tipos HMM e NN ganharam mais atenção e eles têm sido usados consistentemente entres os anos de 2012 a 2015. Esses algoritmos são usados sozinhos ou em combinação com outras técnicas, como meta-aprendizagem ou técnicas de conjunto. HMM é simples de implementar, remove a complexidade da classificação e pode ser usado para produzir modelos de classificação simples.

No artigo [5], "*Performance of machine learning techniques in the detection of financial frauds*", o autor constata que as técnicas híbridas de detecção de fraudes são as mais utilizadas, combinando os pontos fortes de cada algoritmo. A detecção de fraudes em cartões de pagamento utiliza diversas técnicas de ML, combinadas com técnicas de otimização como a agregação, sendo que, para a detecção de fraudes de demonstrações financeiras, são utilizadas, principalmente, técnicas de processamento de texto.

NB e SVM forneceram bons resultados com o conjunto de dados sintético NSL-KDD – *Network Security Laboratory - Knowledge Discovery and Data Mining*, com 99,02% e 98,8%, respectivamente, na identificação de fraude com cartão de crédito. Também foram realizadas as seguintes descobertas:

Tradicionalmente, muitos problemas de classificação tentam resolver a situação de duas ou várias classes. O objetivo da aplicação de aprendizado de máquina é distinguir os dados de teste entre várias classes, usando dados de treinamento. Mas e se você só tiver dados de uma classe e o objetivo for testar novos dados e descobrir se eles são semelhantes ou não aos dados de treinamento? Um método para essa tarefa, é a Máquina de Vetores de Suporte de Uma Classe (OCSVM) – *One Class Support Vetor Machine.*

No artigo [6], "*Real Time Data-Driven Approaches for Credit Card Fraud Detection*", foi utilizado um grande conjunto de transações de comércio eletrônico online de titulares de

cartões de crédito europeus, contendo 284.807 transações coletadas durante dois dias em setembro de 2013; os dados sofreram um pré-processamento através do PCA – *Principal Component Analysis.* Essa transformação gerou variáveis numéricas no conjunto de dados para manter a confidencialidade de informações sensíveis. Com base nesse conjunto de dados, foram conduzidas simulações para gerar os dados de transações fraudulentas. Então, foram usadas 284.000 transações para a fase de treinamento, 200 transações legítimas e 200 transações fraudulentas para testar os algoritmos. Foram utilizadas duas abordagens para detecção de fraude na utilização de cartão de crédito, sendo uma o OCSVM, com a seleção de parâmetros *Kernel function* otimizada e o controle de gráfico T2 (TRACY *et al*., 1992). Os resultados da pesquisa demostram que o OCSVM supera o controle de gráfico T2 em todos os campos de comparação, com uma precisão de 96,6%.

No artigo [7], "*Real-time Credit Card Fraud Detection Using Machine Learning*", após revisão da literatura, o autor identifica quatro algoritmos mais utilizados na detecção de fraude com cartão de crédito: NB, LR, KNN e SVM. A partir de uma base de dados com 917.781 registros com transações legítimas e 200 registros com transações fraudulentas, foi desenvolvida uma API para tratar a detecção de quatro tipos de fraudes em tempo real. Nesse conjunto de 200 registros, as fraudes foram classificadas em quatro grupos: I) Fraudes que ocorrem devido ao risco MCC; II) Transações maiores que $ 100; III) Transações com código de resposta ISO de risco (Google Standard Payments); IV) Transações com endereços da web desconhecidos.

As taxas de precisão obtidas com os algoritmos LR, NB, KNN e SVM foram 74%, 83%, 72% e 91%, respectivamente, demonstrando que o SVM teve maior precisão. No artigo [8], "*Ensemble Learning for Credit Card Fraud Detection*", o autor observa que o algoritmo RF possui maior acurácia na detecção de transações legítimas e o algoritmo NN possui maior acurácia na detecção de transações fraudulentas. Dessa forma, o autor optou por utilizar os métodos de forma conjunta, ou seja, baseado em floresta randômica RF e rede neural NN, com isso, segundo o autor, mantendo o melhor dos dois mundos, sendo capaz de prever, com alta acurácia e confiança, transações de novas amostras, ou seja, obtendo um método de identificação generalista, descartando qualquer viés no modelo treinado. Informa também que a validação experimental foi realizada em conjuntos de dados do mundo real.

No artigo [9], "*An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation*", são coletados dados reais de transações financeiras ocorridas na Coréia em 2015, entre os meses de junho a novembro, totalizando 270.000 transações. Na revisão da literatura proposta no artigo, foram pesquisados métodos de detecção de fraude na utilização de cartão de crédito entre os anos de 2016 e 2018, os quais foram separados entre: aprendizado de máquina ML e aprendizado profundo DL (*Deep Learning).* Dentro do aprendizado de máquina, foram utilizados métodos de aprendizado supervisionado e não supervisionado, e, no aprendizado profundo, foram utilizadas as redes neurais artificiais.

Os resultados experimentais mostraram que os métodos baseados em aprendizado de máquina têm maior eficiência na detecção de fraudes do que as redes neurais. Os algoritmos de classificação, *Classification Algorithms*, LR e RF alcançaram uma taxa de acurácia de 0.99971 e 0.99969, respectivamente; já os algoritmos de agrupamento, *Clustering Algorithms*, EM e *Density-Based Clustering* alcançaram uma taxa de acurácia de 0.99862 e 0.98788, respectivamente. A rede neural artificial NN obteve uma taxa de acurácia de 0.7728.

5. RESULTADOS DA REVISÃO DE LITERATURA

Esta pesquisa coletou informações dos algoritmos utilizados em detecção de fraude na utilização de cartão de pagamento com base em um recorte de nove artigos publicados de 2017 a abril de 2021. Com base nos artigos citados, foram evidenciados os algoritmos que tiveram maior desempenho, conforme a metodologia utilizada por cada autor.

Embora esta pesquisa tente fornecer uma referência, devido ao conjunto de dados diferentes usados em cada artigo, variação no tamanho do conjunto de dados, registros de fraudes desiquilibrados com campos diferentes, dimensionalidade e complexidade, eles permanecem difíceis de comparar. Deve-se ter cuidado ao se tirar conclusões sobre a eficácia dos métodos de detecção de fraude. Infelizmente, ainda há uma escassez de trabalhos de pesquisa nesse domínio da indústria, dado o impacto estabelecido da fraude na sociedade. Isso pode ser explicado, em parte, por um legado na indústria de pagamentos, que aceita tacitamente que o custo da fraude como um custo de negócios possui um volume de negócios baixo e aceitável.

A aceitação de cartões de pagamento cresceu e com ele os lucros das instituições financeiras. Dessa forma, os níveis de fraude aumentaram, mas representam uma parcela desproporcionalmente pequena desses lucros (ANGEL *et al.*, 2014). Os bancos consideram os prejuízos por fraude semelhantes à inadimplência e, portanto, como um "custo dos negócios" (GATES; JACOB, 2008)

Apesar da rápida mudança na tecnologia de computação e o crescimento da Internet, os vetores de fraude, até recentemente, evoluíram lentamente e, assim, os métodos de detecção atuais foram considerados adequados por participantes e fornecedores de um FMS. Isso pode ter levado a uma motivação limitada pela indústria para colaborar e financiar pesquisas futuras sobre detecção de fraude com cartões de pagamento, uma vez que o custo da fraude se tornou normativo. Isso teve um impacto significativo na comunidade de pesquisa. Uma observação da pesquisa é que melhorar o desempenho de um classificador geralmente tem sido foco de pesquisa, em vez de uma abordagem sistêmica.

A fraude normalmente é realizada repetidamente usando o mesmo CHD/cartão de pagamento, até que seja bloqueado. Portanto, é importante que essas sequências de fraudes, que ocorrem ao longo de um período, sejam detectadas o mais cedo possível. Existem apenas alguns métodos que descrevem esse problema e estes usam estatísticas que são agregadas ao longo do tempo para melhorar sua atuação. É sugerida a utilização de uma modelagem de série temporal, com abordagens mais avançadas que possam resultar em melhor desempenho no mundo real.

A falta de grandes conjuntos de dados do mundo real no campo da fraude para a comunidade acadêmica dificulta a pesquisa de novas abordagens para a detecção. É sugerido que deve haver um objetivo de facilitar a cooperação entre pesquisadores e o mundo comercial, para disponibilizar esses conjuntos de dados publicamente com as devidas permissões, em que esses dados sejam suficientemente "ofuscados" para manter a segurança, proteção de dados e respeitar a LGPD em vigor.

A utilização apenas dos conjuntos de dados transacionais da conta ou do titular do cartão pode não oferecer informações suficientes para melhorar a classificação adicional. Isso leva ao uso sugerido de dados mais complexos, incluindo dados não estruturados que estão fora dos conjuntos de dados atuais. "*Pode a mídia social ser usada para aprender padrões*

de comportamento para identificar potenciais fraudadores?" Os dados são um ativo crítico na detecção de fraude, mas muitas vezes são mantidos em silos dentro de uma organização. Reunindo esses dados e adicionando novas fontes de dados, como mídia social e informações que são carregadas todos os dias na Internet, podemos traçar o perfil dos cibercriminosos, viabilizando a identificação do criminoso e seu comportamento. Com isso, podemos adicionar uma nova abordagem para interromper o crescimento da fraude cibernética.

Os estudos pesquisados se concentram em classificadores de fraude e como estes podem ser melhorados em relação a outras abordagens. Trata-se de um problema não trivial, dadas as complexidades dos dados do mundo real; no entanto, é sugerido que o classificador de detecção de fraude atingiu um ponto em que há poucos insights práticos a serem obtidos, concentrando-se em sua melhoria de forma isolada. O DL, *Deep Learning,* com redes neurais recentemente recebeu muita atenção de pesquisa, especialmente em aplicações como reconhecimento de imagem e processamento de linguagem natural. No entanto, não é claro se esse método tem alguma vantagem no domínio da detecção de fraude (SALAKHUTDINOV; HINTON, 2009). Essas abordagens podem não ser melhores em relação a métodos menos complexos, mas é uma área desafiadora para pesquisas futuras. A pesquisa indica que a natureza temporal e sequencial das transações é importante, pois os humanos desenvolvem comportamentos habituais, em que padrões de gastos em certos bens, lojas, marcas, valores podem ser observados ao longo de um período. Como o FMS normalmente opera em tempo real em um fluxo de dados, essa é uma área-chave de melhoria, e parece que os pesquisadores estão voltando sua atenção para a questão de reconhecer sequências.

Neste artigo, foram revisados os métodos mais recentes utilizados na detecção de fraude financeira usando aprendizado de máquina. Um total de nove artigos recém-publicados, entre os anos de 2017 e 2019, foi analisado, com foco principalmente nos artigos com resultados experimentais usando conjuntos de dados e demostrando a eficiência da detecção por meio do conjunto de dados, onde o método e algoritmo mapeado para o conjunto de dados são especificados e as vantagens e limitações de cada artigo são demonstradas, conforme o Quadro 3.

Quadro 3 - Avaliação dos artigos

ARTIGO	CONJUNTO DE DADOS	MÉTODO APLICADO	MÉTODO DE AVALIAÇÃO	VANTAGEM	LIMITAÇÃO
01	284.807 transações de cartões europeus	Abordagem híbrida NB, KNN e LR	MCC e Taxa de classificação equilibrada	Estudo comparativo usando vários algoritmos	Conjunto de dados altamente distorcido
02	Conjunto de dados de cartão de crédito do mundo real de uma instituição financeira ao longo de 3 meses	NB, NN, RF, SVM, LR, AIRS, GP, DT, CDFM, 10-K, SOM, MLP, KNN, LIR	MCC e benchmark	Utilizado Fuzzy Query 2+ e AdaBoost para melhorar desempenho	Detalhes do conjunto de dados não demonstrado

ARTIGO	CONJUNTO DE DADOS	MÉTODO APLICADO	MÉTODO DE AVALIAÇÃO	VANTAGEM	LIMITAÇÃO
03	30.000.000 transações de uma empresa de comércio eletrônico da China com 62 atributos. Apenas 82.000 transações foram rotuladas como fraude (0,27%)	RF e CART-Based	F-Measure e Matriz de confusão	RF é robusto ao lidar com ruídos e outliers	Desiquilíbrio no conjunto de dados
04	Não fornecido	HMM, NN, SVM, AIS, GA, KNN, DT, SVM	benchmark	Abordagem com métodos híbridos ML e NI "Nature Inspired"	Conjunto de dados para testes e treinamento não fornecido
05	O autor apresenta superficialmente vários "Datasets" classificados como (P) Particular, (S) Standard, (G) Generic.	SOM, OD, AR, LR, DT,	Benchmark	Utilização do conjunto de dados sintético NSL-KDD	Detalhes do conjunto de dados não demonstrado
06	Conjunto de dados reais de transações de comércio eletrônico online de titulares de cartões de crédito europeus, obtidos em setembro de 2013 durante 2 dias. Contendo 284.807 transações legítimas.	OCSVM e T2 Control Chart	DR (Recall or Sensitivity), FPR, Accuracy, F-score	Apresentado método heurístico para selecionar o parâmetro Kernel Ideal. "Optimal Kernel Parameter"	Comparativo limitado a somente dois tipos de métodos
07	Origem não informada, com 917.781 transações legítimas e 200 transações fraudulentas	LR, NB, KNN, SVM	MCC e F-Measure	Fornecer protótipo via API para detecção de fraude em tempo real.	Desiquilíbrio no número de transações legítimas e fraudulentas
08	284.807 transações, sendo 492 transações fraudulentas. Obtidas de titulares de cartões europeus durante 2 dias em setembro de 2013	NN, RF, RL	Matriz de Confusão	Aplicação do PCA para transformação dos dados	O conjunto de dados é altamente desequilibrado, a classe positiva (fraudes) representam 0,172% de todas as transações. Contém apenas números variáveis de entrada que são resultado da transformação PCA
09	Base de dados com transações de pagamento real em ambiente IoT na Coréia em 2016	NB, SVM, RF, RL, OneR, C4.5	F-Measure	Comparativo entre técnicas utilizando redes neurais artificiais "Artificial Neural Networks" e métodos de ML.	Conjunto de dados desiquilibrado. Para encontrar a razão de amostragem adequada foi aplicada a taxa de amostragem de 95:5

Fonte: Dos autores (2022)

6. CONCLUSÃO

Com base neste estudo, foram identificados que os métodos de classificação mais usados são NN (*Neural Network Feed-Forward*), NB (*Naive Bayes*), RF (*Random Florest*) e o método SVM (*Support Vector Machines Based*). Enquanto as redes neurais dominam como classificador, não há evidências suficientes para uma conclusão firme. Conforme discutido, as variações nos diferentes conjuntos de dados provavelmente afetam o desempenho de cada algoritmo. Esse mapeamento fornece um guia quanto aos métodos que têm potencial para atingir um bom nível na detecção de fraude com cartão de pagamento. A partir desta pesquisa, ficou evidenciado que muitos estudos usam pequenos conjunto de dados, sendo que a maioria dos métodos classificadores são sensíveis a grupos e aos respectivos padrões aleatórios presentes em cada subclasse e quando estão próximos do limite de decisão. Esse pode ser o caso, quando há apenas um pequeno volume de exemplos de vetor de fraude para que uma subclasse tenha uma grande fração desses padrões aleatórios. Se uma determinada subclasse é colocada no espaço de pesquisa que está distante de outras subclasses e a dimensionalidade é alta, tornam-se necessários muitos registros de treinamento. Com isso, nesses pequenos conjuntos de dados, o classificador irá generalizar mal – especialmente em dados mais recentes coletados após o modelo ter sido criado (pois isso não refletirá a mudança de comportamentos criminosos). É argumentado que, nesse caso, o método provavelmente se ajustará demais aos padrões aleatórios comum aos membros dessa subclasse, e o classificador resultante não irá capturar adequadamente o domínio do conhecimento de fraude.

As perdas por fraude têm crescido a cada ano desde 1971, e o crime migra cada vez mais rápido para novas tecnologias, pois usam a mesma tecnologia para compartilhar informações. Isso é significativo, pois é estabelecido que há uma necessidade oportuna de pesquisas fundamentais sobre a prevenção eficaz na detecção de fraude de pagamento. Esses novos métodos de pesquisa devem convergir para um aplicativo implantado no mundo real que tenha impacto demonstrável e ser capaz de se integrar com as diversas soluções existentes do setor. Conclui-se que há uma lacuna na indústria em pesquisas para ajudar a reduzir a fraude na utilização de cartão de pagamento.

REFERÊNCIAS

ADEWUMI, A. O.; AKINYELU, A. A. A survey of machine-learning and nature-inspired based credit card fraud detection techniques. **International Journal of System Assurance Engineering and Management**, v. 8, pp. 937-953, 2017.

ANGEL, J. J.; MCCABE, D. The Ethics of Payments: Paper, Plastic, or Bitcoin? **Journal of Business Ethics**, v. 132, pp. 603-611, 2014.

AWOYEMI, J. O.; ADETUNMBI, A. O.; OLUWADARE, S. A. Credit card fraud detection using machine learning techniques: A comparative analysis. *In*: **Proceedings of the International Conference on Computing Networking and Informatics** (ICCNI), Canaanland, Nigeria: Covenant University, 2017.

CASTLE, A. Drawing conclusions about financial fraud: crime, development, and international co-operative strategies in China and the West. In: **Proceedings of the International Symposium for the Prevention and Control of Financial Fraud**, 1998, Beijing, China. Published by ICCLR, Vancouver, Canada [1998].

CHOI, D.; LEE, K. An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation. **Security and Communication Networks**, v. 2018, ID 5483472, 15 p., 2018.

CHOO, K.-K. R.; SMITH, R. G.; MCCUSKER, R.; CRIMINOLOGY, A. I. O. Future directions in technology-enabled crime: 2007–09. **Research and Public Policy Series** no. 78. Australian Institute of Criminology, Canberra, Australia, 2007.

DVORSKY, G. Hackers Have Already Started to Weaponize Artificial Intelligence. **Gizmodo Brasil**, 9 nov. 2017. Disponível em: https://gizmodo.com/hackers-have-already-started-to-weaponize-artificial-in-1797688425. Acesso em: 6 jul. 2022.

GATES, T.; JACOB, K. Payments Fraud: Perception Versus Reality – A conference summary. *In*: "Payments Fraud: Perception Versus Reality" Payments Conference. Federal Reserve Bank of Chicago, pp. 7-13, June 2008. **Economic Perspectives**, 2009. Disponível em: http://refhub.elsevier.com/S0952-1976(18)30152-0/sb71. Acesso em: 6 jul. 2022.

GOOGLE STANDARD PAYMENTS. **Google LLC**, Mountain View, CA, 2022. Disponível em: https://developers.google.com/standard-payments/v1/fops/card/response-codes. Acesso em: 6 jul. 2022.

HAND, D.; WHITROW, C.; ADAMS, N.; JUSZCZAK, P.; WESTON, D. Performance criteria for plastic card fraud detection tools. **Journal of the Operational Research Society**, v. 59, pp. 956-962, 2008.

MAXWELL, A. E.; WARNER, T. A.; STRAGER M. P.; CONLEY, J. F.; SHARP, A. L. Assessing machine-learning algorithms and image-and lidar-derived variables for GEOBIA classification of mining and mine reclamation. **International Journal of Remote Sensing**, v. 36, 2015.

MOHER, D. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. **Systematic Reviews**, v. 4, n. 1, 2015.

MORSE, E. A.; RAVAL, V. PCI DSS: Payment card industry data security standards in context. **Computer Law and Security Report**, v. 24, n. 6, pp. 540-554, 2008.

PORTAL BRASILEIRO DE DADOS ABERTOS. Brasil, 2022. Disponível em: https://dados.gov.br/. Acesso em: 6 jul. 2022.

PYMNTS. Credit, Debit Card Fraud Losses Reached $21.84 Billion Last Year. **PYMNTS**, Boston, USA, 26 Oct. 2016. Disponível em: https://www.pymnts.com/news/security-and-risk/2016/credit-and-debit-card-fraud-reaches-21-84-billion/. Acesso em: 6 jul. 2022.

RANDHAWA, K.; LOO, C. K.; SEERA, M.; LIM, C. P.; NANDI, A. K. Credit card fraud detection using AdaBoost and majority voting. **IEEE Access**, v. 6, 2018.

RYMAN-TUBB, N. F.; KRAUSE, P.; GARN, W. How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. **Engineering Applications of Artificial Intelligence**, v. 76, pp. 130-157, 2018.

SADGALI, I.; SAEL, N.; BENABBOU, F. Performance of machine learning techniques in the detection of financial frauds. *In*: **Proceedings of the Second International Conference on Intelligent Computing in Data Sciences** (ICDS 2018), Fez, Morocco: Elsevier, 2019.

SMADI, B. A.; MIN, M. A Critical review of Credit Card Fraud Detection Techniques. *In*: **Proceedings of the 2019 IEEE International Conference on System, Computation, Automation and Networking** (ICSCAN), Ottawa: IEEE Xplore, Carleton University, 2021.

SOHONY, I.; PRATAP, R.; NAMBIAR, U. Ensemble Learning for Credit Card Fraud Detection. *In*: **Proceedings of the ACM India Joint International Conference on Data Science and Management of Data**, CoDS-COMADpp, pp. 289-294, 2018.

SALAKHUTDINOV, R. R.; HINTON, G. E. Deep Boltzmann machines International Conference on Artificial Intelligence and Statistics. *In*: **Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics**, PMLR, v. 5, pp. 448-455, 2009.

SHOKRI, R. Privacy games: Optimal user-centric data obfuscation. **Proc. Priv. Enhanc. Technol.**, v. 2, pp. 1-17, 2015.

THENNAKOON, A.; BHAGYANI, C.; PREMADASA, S.; MIHIRANGA, S.; KURUWITAARACHCHI, N. Real-time credit card fraud detection using machine learning. **Proceedings of the 9th International Conference on Cloud Computing, Data Science & Engineering** (Confluence), Noida, India, 2019.

TRACY, N.; YOUNG, J.; MASON, R. Multivariate Control Charts for Individual Observations. **J. Qual. Technol.**, v. 24, n. 2, pp. 88-95, 1992.

TRAN, P. H.; TRAN, K. P.; HUONG, T. T.; HEUCHENNE, C. Real Time Data-Driven Approaches for Credit Card Fraud Detection. **Proceedings of the 2018 International Conference on E-Business and Applications**, pp. 6-9, 2018.

XUAN, S.; LIU, G.; LI, Z.; ZHENG, L.; WANG, S.; JIANG, C. Random forest for credit card fraud detection. **Proceedings of the 15th International Conference on Networking, Sensing and Control**, IEEE, pp. 1-6, 2018.