Detecção de fraudes online usando Algoritmos One-class e Autoencoder

line 1: 1st Evelyn Torezone
line 2: *dept.de controle e automação*
line 3: *IFES*
line 4: Serra, Brasil
line 5: etorezone@gmail.com

line 1: 2nd Daniel Cavalieri
line 2: *dept. controle e automação*
line 3: *IFES*
line 4: Serra, Brasil
line 5: daniel.cavalieri@ifes.edu.br

***Resumo*— Este artigo apresenta uma abordagem com técnicas combinadas para detecção de fraudes online, utilizando algoritmos de classe única combinados com autoencoder. A eficácia da abordagem foi testada e os resultados apresentaram melhorias em relação aos métodos tradicionais. Essa abordagem procura uma melhora na detecção de fraudes online, visando maior precisão na identificação de transações fraudulentas, e buscando reduzir o impacto financeiro causado pela detecção equivocada de fraudes.**

Keywords—Fraude online, autoencoder, One-Class, Anomalia.

I. INTRODUÇÃO

A pandemia acelerou significativamente a evolução digital, impulsionando o crescimento do comércio eletrônico em todo o mundo, com destaque para o Brasil, que ocupa a 2ª posição no ranking global[1]. Essa expansão do e-commerce tem trazido benefícios para a economia, no entanto, também surgem preocupações em relação à veracidade das transações. Pesquisas recentes indicam que o Brasil registrou um alarmante aumento de 58% nas tentativas de fraude no setor de e-commerce[2].

O aumento das atividades fraudulentas no comércio eletrônico é uma consequência direta do aumento do volume de transações online. Os criminosos aproveitam-se da crescente demanda e do ambiente digital para tentar obter ganhos ilícitos, prejudicando tanto os consumidores quanto as empresas. Esse cenário destaca a importância de estratégias eficazes para combater e prevenir fraudes nesse setor.

As empresas do setor de e-commerce têm buscado constantemente soluções para enfrentar esse desafio. O investimento em sistemas de segurança avançados, como a verificação em duas etapas, autenticação multifatorial e análise de dados em tempo real, tem se mostrado fundamental para mitigar as tentativas de fraude. Além disso, é essencial educar e conscientizar os consumidores sobre práticas seguras de compra online, como a verificação de certificados de segurança, a utilização de senhas robustas e a verificação de reputação do vendedor.

Outra estratégia eficaz é o uso de algoritmos de detecção de fraudes, que analisam padrões e comportamentos suspeitos nas transações. Esses algoritmos utilizam técnicas avançadas de aprendizado de máquina e inteligência artificial para identificar transações fraudulentas com maior precisão. Ao combinar a análise de dados em tempo real com modelos preditivos, é possível detectar e bloquear atividades fraudulentas de forma proativa, minimizando os danos tanto para as empresas quanto para os consumidores.

Para Van Vlasselaer[3], a fraude é um crime incomum, bem pensado, imperceptivelmente escondido, que evolui no tempo, muitas vezes cuidadosamente organizado e que aparece em muitos tipos e formas. Essa definição evidencia a intenção do fraudador de agir naturalmente e se misturar com o grande número de pessoas que compram online.

A tentativa de compra em fraude disfarçada de transação legítima é uma estratégia enganosa que visa ludibriar vendedores e instituições financeiras. Nesse tipo de golpe, os fraudadores empregam táticas elaboradas para parecerem clientes genuínos, dificultando a identificação de suas intenções maliciosas.

Um dos métodos comuns utilizados por golpistas é a utilização de informações de cartões de crédito roubados ou clonados. Eles podem obter esses dados por meio de violações de segurança de empresas, ataques cibernéticos ou até mesmo comprá-los em mercados ilegais na dark web. Ao realizar a compra, eles se certificam de fornecer informações precisas do titular do cartão, como nome, número, data de validade e código de segurança, para que a transação inicialmente pareça legítima [4].

A fraude de pagamento online envolve o uso de cartões de crédito ou débito roubados, a utilização de informações falsas para realizar compras ou ações de phishing para obter informações de login e senha. Já a fraude de pagamento não online pode incluir a emissão de cheques sem fundos, a clonagem de cartões em estabelecimentos comerciais ou a utilização de moeda falsa[5].

As fraudes online podem ser divididas em dois grupos[6]: fraude de aplicação e fraude comportamental. Na fraude de aplicação, o fraudador obtém informações de terceiros e solicita um novo cartão com dados falsos. Depois de obtê-lo, ele dispara uma série de transações para gastar o máximo possível em um curto espaço de tempo, antes de ser bloqueado. Já na fraude comportamental, os dados do cartão são legítimos, mas foram obtidos de forma inautêntica, e o fraudador os utiliza em compras online. Nesse caso, a tendência é que o fraudador tenha mais sucesso se tentar em locais diferentes e com intervalo de tempo entre as tentativas, já que o cartão é conhecido pelas ferramentas de análise de fraude.

Independentemente do grupo, a ocorrência de fraudes online causa prejuízos tanto para a pessoa portadora do cartão quanto para as entidades financeiras envolvidas em um pagamento online (Banco emissor, adquirente, subadquirente). O banco emissor é responsável por emitir o cartão de crédito ou débito ao cliente. O adquirente é

responsável por adquirir a transação financeira realizada pelo estabelecimento comercial e repassar o valor ao banco emissor. Já o subadquirente é uma empresa intermediária entre o adquirente e o estabelecimento comercial, realizando a venda ou aluguel de máquinas de cartão e oferecendo serviços de suporte e gerenciamento de transações[7]. De acordo com dados levantados em 2021, o crime cibernético resultou em perdas de US$ 6,9 bilhões, 64% a mais do que no ano anterior, sendo grande parte relacionada a serviços e pagamentos online[8].

Para evitar esses danos, é necessário ter um processo para detecção e impedimento dessas operações fraudulentas. Embora os mecanismos de identificação com base em aprendizado automático sejam bastante empregados, recentemente, com o avanço da inteligência artificial novos métodos estão sendo testados para uma nova abordagem de detecção dessas anomalias[9].

As técnicas de aprendizado usadas na detecção de fraudes enfrentam desafios significativos. A necessidade de tomar decisões rápidas é crucial, considerando que as transações ocorrem em tempo real e é essencial identificar e agir prontamente em casos de possíveis fraudes. Além disso, o desbalanceamento dos dados, com as transações fraudulentas representando uma pequena proporção em relação às transações legítimas, requer abordagens especializadas para garantir uma detecção eficaz em ambos os casos.

A detecção de anomalias é um desafio adicional. Os fraudadores estão constantemente aprimorando suas táticas para evitar detecção, tornando essencial que as técnicas de aprendizado possam identificar padrões incomuns e anômalos nos dados, a fim de detectar possíveis fraudes.

Outro aspecto importante é a necessidade de reaprendizado contínuo. À medida que novos tipos de fraudes surgem, os modelos de detecção devem ser atualizados regularmente para se adaptarem a essas mudanças e garantir uma alta taxa de precisão na identificação de fraudes.

Dado o crescimento progressivo das fraudes, elas se tornaram um dos principais problemas enfrentados pelo mercado de pagamentos online. Portanto, é essencial aprimorar constantemente os métodos de detecção, empregando abordagens avançadas de aprendizado de máquina e aproveitando técnicas específicas para lidar com os desafios mencionados.

Superar esses desafios requer uma abordagem integrada, combinando algoritmos eficientes de classificação, técnicas de amostragem para lidar com desbalanceamento de dados, métodos de detecção de anomalias e sistemas de atualização contínua do modelo. Somente dessa forma poderemos enfrentar efetivamente o problema das fraudes no mercado de pagamentos online.

Nos últimos anos, a abordagem de detecção de fraudes online tem sido cada vez mais relevante. Uma delas consiste em utilizar algoritmos de classe única (do inglês, one-class) em conjunto com outros métodos como o autoencoder.

Nesse contexto, os dados são tratados e testados com múltiplos métodos de detecção de classe única, buscando encontrar o melhor resultado de verdadeiros positivos (Fraude) com o menor impacto possível de falsos positivos. Isso é especialmente importante porque falsos positivos podem causar perdas financeiras tanto para o lojista quanto para a empresa de pagamentos[10].

II. REVISÃO BIBLIOGRÁFICA

A detecção de fraudes em comércio eletrônico é um problema crescente e cada vez mais relevante para as entidades financeiras[11]. Nesse contexto, algoritmos de classe única combinados com mecanismos de atenção autoencoder podem ser uma abordagem promissora para a detecção de anomalias. Para explicar o modelo proposto de detecção de fraudes online usando algoritmos de classe única e mecanismos de atenção, é necessário algumas características dos modelos.

Os algoritmos de detecção de anomalias de classe única trabalham com uma única categoria de dados, sendo treinados para detectar anomalias nessa categoria. Esses algoritmos baseiam-se na premissa de que as anomalias são eventos raros, que representam apenas uma pequena porção dos dados, tornando a detecção mais eficiente. O principal objetivo desses algoritmos é aprender uma fronteira de decisão que permita distinguir exemplos normais daqueles que representam anomalias[12].

Alguns exemplos de algoritmos de classe única incluem o One-class Support Vector Machines (OCSVM), Copula-BasedOutlierDetection (COPOD), Clustering Based Local Outlier Factor (CBLOF), Histogram-based Outlier Detection (HBOS) e Isolation Forest. Pode-se entender conceitualmente um pouco mais do seu funcionamento:

- One-Class SVM: Este algoritmo é eficiente em detectar anomalias em dados uni-variados e pode lidar com conjuntos de dados desbalanceados. Ele utiliza uma técnica de separação de hiperplano para encontrar uma fronteira de decisão que possa separar exemplos normais de anomalias. No entanto, pode ser menos eficaz em dados multidimensionais e com muitas variáveis e pode ser sensível à escolha de parâmetros, o que pode afetar sua precisão [13].
- COPOD: Este algoritmo combina técnicas de clusterização e detecção de anomalias para aumentar a capacidade de detecção em dados multidimensionais. Ele usa um método baseado em densidade para identificar regiões de baixa densidade que possam conter anomalias. Embora seja eficaz em lidar com dados que apresentam variações na densidade, pode ser computacionalmente mais intensivo em comparação com outros algoritmos e precisa de uma boa escolha de parâmetros para uma detecção precisa [14].
- CBLOF: Este algoritmo é eficaz em encontrar anomalias que estão em grupos de baixa densidade e é eficiente em conjuntos de dados grandes. Ele usa um método de agrupamento para identificar esses grupos e avalia a densidade local dos pontos de dados para detectar anomalias. No entanto, pode ser menos eficaz em conjuntos de dados com alta dimensionalidade e pode não funcionar bem em dados com distribuição não uniforme [15].
- HBOS: Este algoritmo é computacionalmente eficiente e eficaz em detectar anomalias em dados com distribuições uniformes. Ele é baseado em histogramas e busca identificar anomalias por meio de

uma contagem de quantas vezes um ponto de dados aparece em um determinado intervalo de histograma. No entanto, pode não funcionar bem em dados com distribuição não uniforme e pode ser menos eficaz em dados com alta dimensionalidade [16].

- Isolation Forest: Este algoritmo é especialmente eficaz em detectar anomalias em dados de alta dimensionalidade e pode lidar com conjuntos de dados desbalanceados. Ele usa árvores de decisão aleatórias para isolar anomalias, dividindo o conjunto de dados em subconjuntos aleatórios. No entanto, pode ser menos eficaz em detectar anomalias que estão em grupos densos e pode ser sensível à escolha de parâmetros [17].

Já o Autoencoder é uma ferramenta poderosa de aprendizado não supervisionado que permite aprender representações compactas e significativas de conjuntos de dados de alta dimensionalidade. Este artigo tem como objetivo explorar o Autoencoder, com ênfase no Autoencoder de Classe Única, uma variante especializada na detecção de anomalias e no aprendizado de dados não convencionais [18].

O Autoencoder consiste em duas partes principais: o encoder (codificador) e o decoder (decodificador). O encoder mapeia a entrada de alta dimensão para um espaço latente de menor dimensão, enquanto o decoder reconstrói a entrada a partir dessa representação latente. A arquitetura mais comum é a do Autoencoder Simples, com uma única camada escondida, mas também existem arquiteturas mais complexas, como os Autoencoders Convolucionais (CAEs) e os Autoencoders Recorrentes (RAEs), adequados para diferentes tipos de dados, como imagens e sequências temporais [19].

O Autoencoder de Classe Única é uma extensão do Autoencoder tradicional, projetada para aprender a representação de uma única classe, geralmente a classe majoritária. Ao treinar o Autoencoder apenas com exemplos dessa classe, ele busca aprender uma representação compacta específica para essa classe. Posteriormente, quando apresentado a exemplos de outras classes ou anomalias, o Autoencoder de Classe Única tende a gerar uma maior discrepância entre a entrada e a saída, permitindo a detecção de padrões anômalos [20].

O Autoencoder de Classe Única demonstra eficácia em diversas aplicações. É frequentemente utilizado na detecção de anomalias em dados, como fraudes em transações financeiras, monitoramento de sistemas de segurança e diagnóstico médico. Além disso, também é aplicado em tarefas de detecção de outliers, filtragem de ruídos e compressão de dados [21].

Com sua arquitetura flexível e capacidade de aprendizado não supervisionado, o Autoencoder de Classe Única destaca-se como uma técnica promissora para a detecção de anomalias e aprendizado de representações compactas. Suas aplicações práticas têm demonstrado resultados significativos na detecção de padrões anômalos e no aprimoramento do desempenho de sistemas de detecção de anomalias [22].

No modelo proposto, é utilizada uma combinação de algoritmos de classe única e Autoencoder. A abordagem envolve o uso de um Autoencoder, um tipo de rede neural que aprende uma representação compacta dos dados e, em seguida, tenta reconstruir os dados originais a partir dessa representação. O Autoencoder é treinado com dados normais e, em seguida, é utilizado para detectar anomalias, ou seja, dados que não se enquadram na distribuição normal dos dados.

O Autoencoder enfatiza as partes mais importantes dos dados durante a reconstrução. Isso permite que o modelo se concentre em partes específicas dos dados que são mais relevantes para a detecção de anomalias. Além disso, o modelo é treinado com diferentes algoritmos de classe única para encontrar o melhor resultado.

Essa abordagem tem se mostrado eficaz na detecção de fraudes online, pois permite que o modelo se adapte a diferentes tipos de dados e padrões de fraude[23]. Além disso, a utilização de mecanismos de atenção ajuda a explicar as decisões do modelo, tornando-o mais transparente e confiável para os usuários.

Em resumo, o modelo proposto de detecção de fraudes online combina algoritmos de classe única e Autoencoder para aprender a fronteira de decisão que permite distinguir anomalias dos dados normais, tornando a detecção mais eficiente e precisa. A abordagem tem mostrado resultados promissores e pode ser uma solução viável para um dos principais problemas do mercado de pagamentos atual.

III. METODOLOGIA

A metodologia utilizada neste estudo consiste em utilizar cinco algoritmos de detecção de anomalias, a saber: OCSVM, COPOD, CBLOF, HBOS e IForest, sozinhos e combinados com Autoencoder.

Os dados empregados neste estudo foram adquiridos de uma empresa brasileira do setor de pagamentos, compreendendo informações de 25 atributos de transações de cartão de crédito, todos submetidos a medidas de anonimização. É relevante enfatizar que a base de dados foi previamente submetida a um conjunto de tratamentos nas variáveis preditoras, tais como técnicas de fuzzyficação, binarização, normalização dos dados e redução da dimensionalidade das variáveis.

Essas características incluem informações de tempo entre as transações, dados envolvidos nas compras (por exemplo, e-mail e telefone) e tipo de pagamento utilizado (por exemplo, pagamento à vista ou parcelado). Todos esses dados são utilizados para desenvolver modelos que permitem a detecção de fraudes em transações de cartão de crédito. Para a realização dos testes, o conjunto de dados foi dividido em dois subconjuntos, um contendo apenas observações normais e outro contendo as observações anômalas. Em seguida, foram realizados os seguintes passos:

1. Dividimos os dados em "normais" e "anomalias". Os dados "normais" foram divididos em treino e teste, enquanto os dados "anomalos" foram mantidos em um único conjunto. Todos os dados foram normalizados antes dessa divisão.
2. Definimos o Autoencoder com 0,1 de contaminação com uma rede neural composta por sete camadas, contendo 128, 64, 32, 16, 32, 64 e 128 neurônios, respectivamente, o número de épocas foi definido como 50, o tamanho do lote como 32 e o tamanho de validação como 0,1. Todos os valores foram definidos de maneira empírica.

3. Treinamento do Autoencoder: foi realizado o treinamento do Autoencoder utilizando os dados de treino normalizados.
4. Gerar os dados latentes: após o treinamento do Autoencoder, foram gerados os dados latentes das observações de treino normalizadas.
5. Na etapa cinco, foi realizado o ajuste do algoritmo de detecção de anomalias utilizando os dados latentes gerados na etapa anterior. Esse processo permitiu que o algoritmo fosse capaz de identificar padrões anômalos nos dados de teste, baseado na análise dos dados latentes gerados a partir dos dados normais de treinamento.
6. Na etapa seis, foram gerados os dados latentes a partir das observações de teste, tanto as normais quanto as anormais, utilizando o modelo que foi treinado anteriormente. Vale ressaltar que esses dados foram normalizados para garantir que estivessem na mesma escala dos dados utilizados no treinamento do modelo.
7. Detectar anomalias: Com o algoritmo de detecção de anomalias de classe única ajustado, as observações de teste foram avaliadas. As previsões foram geradas e um vetor binário foi retornado, contendo valores 1 para anomalias e 0 para observações normais.

Os mesmos passos foram repetidos para todos os algoritmos de classificação única.

IV. RESULTADOS

Para avaliar os resultados, consideramos tanto o impacto financeiro dos verdadeiros positivos (TP) quanto dos falsos positivos (FP). No nosso modelo, uma transação erroneamente classificada como fraude possui um impacto financeiro maior do que o impacto de uma transação fraudulenta não detectada.

É importante ressaltar que a proporção do impacto pode variar entre diferentes negócios. No entanto, para o nosso projeto específico, estabelecemos que o impacto de uma transação legítima indevidamente barrada é equivalente a seis fraudes não detectadas.

Essa abordagem nos permite considerar a importância de identificar corretamente as transações fraudulentas, bem como a necessidade de evitar bloqueios indevidos de transações legítimas. Dessa forma, estamos comprometidos em maximizar o desempenho do sistema, equilibrando de maneira adequada os riscos e prejuízos envolvidos no processo de detecção de fraudes.

Para avaliar os resultados, utilizaremos métricas específicas, como a área sob a curva ROC (AUC-ROC), que integra as medidas de especificidade e sensibilidade.

A AUC-ROC é amplamente utilizada na avaliação de modelos de detecção de fraudes. Essa métrica mede a capacidade do modelo de classificar corretamente as transações fraudulentas e legítimas, considerando a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR). Quanto maior o valor da AUC-ROC, melhor o desempenho do modelo em distinguir fraudes de transações legítimas [24].

A especificidade e a sensibilidade são métricas fundamentais na avaliação de modelos de detecção de fraudes. A especificidade avalia a capacidade do modelo de identificar corretamente as transações legítimas, ou seja, a taxa de verdadeiros negativos (TNR) [25]. Já a sensibilidade mede a capacidade do modelo de detectar corretamente as transações fraudulentas, representando a taxa de verdadeiros positivos (TPR) [26].

Ao considerar essas métricas em nosso projeto, busca obter uma visão abrangente do desempenho do sistema, garantindo uma detecção eficaz de fraudes e minimizando os impactos de bloqueios indevidos de transações legítimas.

As tabelas 1 e 2 apresentam os resultados obtidos com e sem a utilização do Autoencoder. É importante destacar o papel significativo do Autoencoder na redução do efeito de contaminação, resultando em uma melhoria na curva ROC.

Ao aplicar o Autoencoder, observa-se na tabela 2 uma diminuição na influência das informações contaminadas nos resultados. Isso se reflete na curva ROC, que apresenta um desempenho aprimorado em comparação com a abordagem sem o uso do Autoencoder.

O Autoencoder desempenha um papel crucial ao filtrar e remover os efeitos negativos da contaminação nos dados. Ao reconstruir a representação dos dados originais, o Autoencoder elimina as características indesejadas introduzidas pela contaminação, permitindo que o modelo se concentre nas informações relevantes para a detecção de fraudes ou anomalias.

Consequentemente, a inclusão do Autoencoder no processo resulta em uma curva ROC mais robusta, refletindo uma melhor capacidade de distinguir entre padrões normais e anormais. Essa melhoria é fundamental para a eficácia da detecção de fraudes, pois reduz a ocorrência de falsos positivos e aumenta a taxa de detecção correta de transações fraudulentas.

Em resumo, os resultados evidenciam que a utilização do Autoencoder tem um impacto positivo na atenuação do efeito da contaminação e na melhoria da curva ROC. Essa abordagem aprimorada é crucial para aumentar a precisão e a confiabilidade da detecção de fraudes, fornecendo uma solução mais robusta e eficiente para os desafios enfrentados nesse campo.

Tabela 1: Resultado dos algoritmos de classe única sem Autoencoder.

Modelo	Contaminação	AUC-ROC	Especificidade	Sensibilidade
CBLOF	0,15	0,63	0,85	0,41
CBLOF	0,1	0,62	0,91	0,32
COPOD	**0,1**	**0,79**	**0,90**	**0,67**
COPOD	**0,15**	**0,79**	**0,86**	**0,72**
HBOS	0,1	0,69	0,90	0,47
HBOS	0,15	0,68	0,85	0,51
IForest	0,1	0,77	0,90	0,65
IForest	0,15	0,77	0,85	0,69
OCSVM	**0,1**	**0,78**	**0,90**	**0,66**
OCSVM	**0,15**	**0,78**	**0,85**	**0,70**

Na tabela acima, podemos observar os resultados com as contaminações de 0,1 e 0,15 para os 5 algoritmos One-Class, destacando COPOD e OCSVM com os melhores desempenhos.

Tabela 2: Resultado dos algoritmos de classe única com Autoencoder.

Modelo	Contaminação	AUC-ROC	Especificidade	Sensibilidade
CBLOF	0,1	0,79	0,90	0,68
CBLOF	0,15	0,79	0,85	0,73
COPOD	**0,1**	**0,79**	**0,85**	**0,73**
COPOD	**0,15**	**0,79**	**0,85**	**0,73**
HBOS	0,1	0,75	0,95	0,55
HBOS	0,15	0,79	0,85	0,73
IForest	0,1	0,79	0,85	0,73
IForest	0,15	0,79	0,90	0,68
OCSVM	**0,1**	**0,76**	**0,93**	**0,59**
OCSVM	**0,15**	**0,79**	**0,85**	**0,73**

Figura 1: Curva ROC OCSVM sem autoencoder.

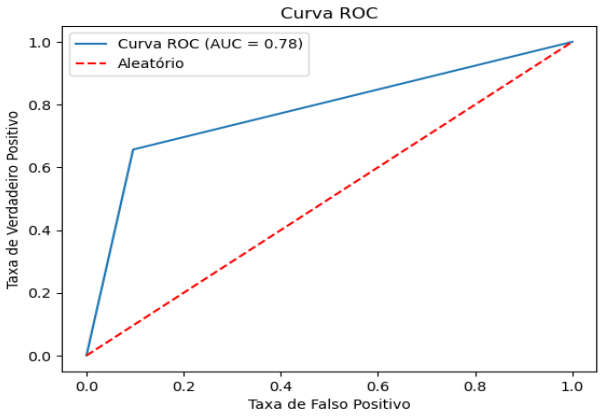
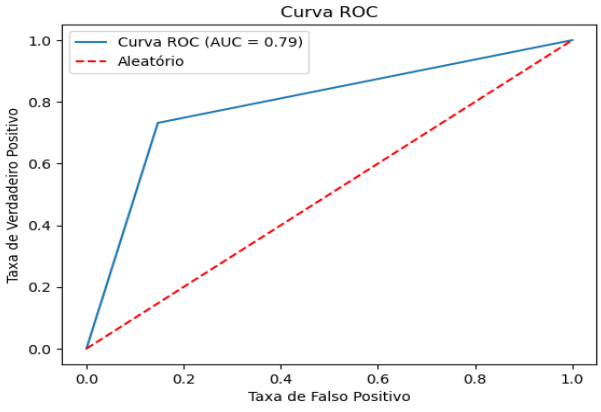


Figura 1: Curva ROC OCSVM com autoencoder.



Um Autoencoder de classe única é um tipo de algoritmo de aprendizado não supervisionado que é projetado para aprender a representação dos dados normais e reconstruí-los de forma precisa. Ele é treinado apenas com exemplos de uma única classe (por exemplo, dados não fraudulentos) e tem como objetivo capturar a estrutura dos dados normais. Ao fazer isso, o Autoencoder pode detectar anomalias que não se ajustam bem a essa estrutura e atribuir a elas um alto valor de erro de reconstrução.

No entanto, o Autoencoder de classe única pode ter dificuldades em detectar anomalias que são muito diferentes dos dados normais ou que estão em regiões pouco representadas pelos exemplos de treinamento. É aí que a combinação com outros algoritmos one class pode ser útil.

Os algoritmos one class, como o Isolation Forest e o One-Class SVM, são projetados especificamente para a tarefa de detecção de anomalias. Eles são capazes de identificar padrões incomuns nos dados que não se enquadram nas características normais. Ao combinar esses algoritmos com o Autoencoder de classe única, podemos aproveitar as vantagens de cada abordagem.

Enquanto o Autoencoder de classe única é eficaz em capturar a estrutura dos dados normais e detectar anomalias que se desviam dessa estrutura, os algoritmos one class podem complementar essa detecção, identificando anomalias que são mais distintas ou estão em regiões menos representadas. Dessa forma, a combinação dos dois métodos pode levar a uma detecção mais abrangente de fraudes ou anomalias, resultando em melhores resultados na curva ROC.

Em resumo, ao combinar um Autoencoder de classe única com outros algoritmos one class, aproveitamos a capacidade do Autoencoder de capturar a estrutura dos dados normais e a capacidade dos algoritmos one class de detectar anomalias mais distintas ou em regiões pouco representadas. Essa combinação pode resultar em melhores resultados na curva ROC, melhorando a detecção de fraudes e anomalias.

V. CONCLUSÕES

Com base nos resultados obtidos e considerando o objetivo do trabalho de encontrar o melhor equilíbrio entre falsos positivos e falsos negativos, a combinação de um Autoencoder de classe única com outros algoritmos one class demonstrou ser uma abordagem promissora para a detecção de anomalias.

Observa-se uma melhora nos resultados ao manter o OCSVM e o CBLOF como os métodos mais promissores. No entanto, é necessário investigar e compreender melhor a relação causal por trás da melhoria obtida pela combinação desses métodos com o Autoencoder.

O Autoencoder de classe única, ao ser treinado apenas com exemplos de uma única classe e aprender a reconstruir de forma precisa os dados normais, mostrou-se eficaz na captura da estrutura desses dados e na detecção de anomalias que não se ajustam bem a essa estrutura. No entanto, pode ter dificuldades em detectar anomalias que são muito diferentes dos dados normais ou que estão em regiões pouco representadas.

Ao combinar o Autoencoder de classe única com algoritmos one class, como o Isolation Forest e o One-Class

SVM, aproveitamos as vantagens de cada método. Os algoritmos one class são projetados especificamente para a detecção de anomalias e podem complementar a detecção do Autoencoder, identificando padrões incomuns nos dados que não se enquadram nas características normais. Dessa forma, a combinação dos dois métodos permite uma detecção mais abrangente de fraudes ou anomalias [27].

A abordagem de combinar um Autoencoder de classe única com outros algoritmos one class tem o potencial de melhorar os resultados na curva ROC, que é uma medida importante para avaliar o desempenho de modelos de detecção de anomalias. A complementaridade das abordagens, onde o Autoencoder captura a estrutura dos dados normais e os algoritmos one class identificam anomalias mais distintas ou em regiões menos representadas, contribui para uma detecção mais eficaz de fraudes e anomalias [28].

Em trabalhos futuros, podem ser exploradas outras combinações de algoritmos one class com diferentes variantes de Autoencoders, bem como técnicas de ajuste fino e otimização dos modelos. Além disso, é importante considerar diferentes métricas de avaliação, além da curva ROC, para uma análise mais completa do desempenho dos modelos.

REFERENCIAS

[1] Brasil fecha o ano com o segundo maior crescimento em comércio eletrônico. INSPER, 15 de dez. de 2021. Disponível em: https://www.insper.edu.br/noticias/brasil-fecha-o-ano-com-o-segundo-maiorcrescimento-em-comercioeletronico/#:~:text=Uma%20das%20mudan%C3%A7as%20mais%20evidentes,das%20vendas%20globa is%20no%20varejo. Acesso em 15 set. 2022

[2] FERREIRA, I.Estudo aponta crescimento de 58% nas tentativas de fraudes contra o e-commerce. Consumidor Moderno. 02 Fev. 2022. Disponível em: https://www.consumidormoderno.com.br/2022/02/02/crescimento-tentativas-fraudes/. Acesso em 01 set. 2022

[3] Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). Gotcha! Network-based Fraud Detection for Social Security Fraud. Management Science, Submitted.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] Cavelty, M. D. (2016). Cyber Security and the Politics of Time. Security Dialogue, 47(1), 5-22.

[5] C. K. Tan, A. C. K. Lee, and P. L. Chong, "A review of credit card fraud detection techniques: Data mining perspective," Journal of Emerging Trends in Computing and Information Sciences, vol. 2, no. 10, pp. 525-534, 2011

[6] BRESLOW, Stuart et al. The new frontier in anti–money laundering. McKinsey & Company, New York, nov. 2017.K. Elissa, “Title of paper if known,” unpublished.

[7] Tudo o que você precisa saber sobre bandeira de cartão, gateway, adquirente e subadquirente(2020) Disponível em: https://pagar.me/blog/subadquirente-e-adquirente-o-que-voce-precisa-saber/. Acesso em 10 de março de 2023.

[8] Relatório de crimes na Internet de 2021 lançado (2021). Disponível em: https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf. Acesso em 28 de outubro de 2022

[9] Ileberi, E., Sun, Y. & Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data 9, 24 (2022). https://doi.org/10.1186/s40537-022-00573-8

[10] T. -Y. Wu and Y. -T. Wang, "Locally Interpretable One-Class Anomaly Detection for Credit Card Fraud Detection," 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, Taiwan, 2021, pp. 25-30, doi: 10.1109/TAAI54685.2021.00014.

[11] NICE Actimize 2023 Fraud Insights Report Reveals Attempted Fraud Transactions Have Increased By 92% Over Previous Year Disponível em: https://www.niceactimize.com/press-releases/nice-actimize-2023-fraud-insights-report-reveals-attempted-fraud-transactions-414/#:~:text=The%20NICE%20Actimize%20Fraud%20Insights,is%20being%20conducted%20almost%20instantly. Acesso em 23 de março de 2023

[12] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58.

[13] Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural computation, 13(7), 1443-1471.

[14] Zhao, Y., & Liu, F. (2019). COPOD: Copula-based outlier detection. Knowledge-Based Systems, 179, 104817.

[15] He, Z., Xu, X., & Huang, J. Z. (2003, August). CBLOF: A cluster-based local outlier factor. In Proceedings of the 2003 ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 682-687).

[16] Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PloS one, 11(4), e0152173.

[17] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413-422). IEEE.

[18] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507.

[19] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research, 11(Dec), 3371-3408.

[20] Tax, D. M., & Duin, R. P. (2004). Support vector data description. Machine learning, 54(1), 45-66.

[21] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A review. arXiv preprint arXiv:1901.03407.

[22] Ruff, L., Vanderme

[23] Breitenbacher, D., Pichler, M., Mayrhofer, R., Engel, R. (2018). Anomaly detection for online fraud prevention using deep autoencoders. In Proceedings of the 26th European Conference on Information Systems (ECIS).

[24] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.

[25] Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. Biometrical Journal, 47(4), 458-472.

[26] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE, 10(3), e0118432.

[27] Chong, E. I., & Ng, V. T. (2017). Autoencoder-based anomaly detection with unlabeled data. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 126-142). Springer, Cham.

[28] Chong, E. I., & Ng, V. T. (2017). Autoencoder-based anomaly detection with unlabeled data. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 126-142). Springer, Cham.