Análise de Padrões Comportamentais de Fraudes por Cartão de Crédito*

Carlos Eduardo Costa Kraizfeld[1] Julio Maciel Coelho[2]
Cleia Marcia Gomes Amaral[3]

Resumo

Com a pandemia do Covid-19, os *e-commerce* obtiveram crescimento substancial, pois se mostraram de extrema importância para a população mundial por consequência da restrição do contato físico e, assim, da transmissão do vírus. Este mercado é, em parte, possível e facilitado pelos cartões de crédito, um dos métodos de pagamento mais utilizados pelos brasileiros justamente por sua praticidade. Porém, embora considerado um método seguro, cibercriminosos conseguem, através da clonagem de cartão ou vazamento de dados, fraudar cartões de crédito de terceiros para benefício próprio realizando pagamentos *online*. Com isso é necessário buscar meios de evitar que essas fraudes ocorram. Este trabalho possui o objetivo de selecionar métodos eficientes para mitigação da fraude, adaptando algoritmos de Aprendizado de Máquina conhecidos e fazendo a experimentação dos mesmos usando aprendizagem supervisionada em uma conjuntos de dados de cartão de crédito, avaliando os resultados entre eles e suas eficácias para cada cenário. Este estudo chega à conclusão de que, para o conjunto de dados analisado, os resultados alcançados após a implementação das técnicas selecionadas de balanceamento e codificação, juntamente com a aplicação dos algoritmos escolhidos de Aprendizado de Máquina, demonstraram um desempenho notável. Destaca-se, especialmente, a eficácia do algoritmo *Random Forest*.

Palavras-chave: Aprendizado de máquina, Aprendizagem supervisionada, fraudes de cartão de crédito, comércio eletrônico

*Trabalho de conclusão de curso, Sistemas de Informação, Unidade São Gabriel

[1]Bacharel em Sistemas de Informação , E-mail:carlos.kraizfeld@sga.pucminas.br
Instituto de Ciências Exatas e de Informática da PUC Minas, Brasil.
[2]Bacharel em Sistemas de Informação, E-mail:julio.coelho.1284317@sga.pucminas.br
Instituto de Ciências Exatas e de Informática da PUC Minas, Brasil.
[3]Orientadora, E-mail:cleia@sga.pucminas.br
Instituto de Ciências Exatas e de Informática da PUC Minas, Brasil.

Abstract

With the Covid-19 pandemic, e-commerce experienced substantial growth as it proved to be of utmost importance to the global population due to the restriction of physical contact and, consequently, the transmission of the virus. This market is, in part, made possible and facilitated by credit cards, one of the most widely used payment methods by Brazilians precisely for its convenience. However, despite being considered a secure method, cybercriminals can, through card cloning or data leakage, fraudulently use third-party credit cards for their own benefit when making online payments. Therefore, it is necessary to seek ways to prevent these frauds from occurring. This work aims to select efficient methods for fraud mitigation by adapting well-known Machine Learning algorithms and experimenting with them using supervised learning on a credit card dataset, evaluating the results among them and their effectiveness for each scenario. This study concludes that, for the analyzed dataset, the results obtained after implementing the selected balancing and encoding techniques, along with the application of the chosen Machine Learning algorithms, demonstrated remarkable performance. The effectiveness of the Random Forest algorithm stands out, particularly.

Keywords: Machine Learning, Supervised Learning, credit card fraud, e-commerce

1 INTRODUÇÃO

Na década de 1950, um acontecimento histórico mudaria a forma de pagamento. Durante um almoço, o executivo Frank MacNamara esqueceu de levar a sua carteira, sendo assim, não poderia pagar a conta. Porém, este conversou com o dono do restaurante e o prometeu pagar assim que possível e assinou um "termo de pagamento". Segundo Leblebici (2012), Frank viu uma oportunidade neste contexto e a partir deste dia, foi criado a primeira versão do cartão de crédito.

O intenso processo de globalização ocorrido no século XX trouxe uma necessidade que fora estudada e analisada por grandes participantes industriais. Com a criação do cartão de crédito, estes criaram uma perspectiva nova para o mercado, permitindo a sobrevivência do modelo consumista. Com o passar dos anos, o mundo digital também entrou neste processo criando-se as plataformas de comércio eletrônico (*e-commerce*), que permitem que o usuário faça compras em lojas virtuais utilizando algumas formas de pagamento, entre elas, o cartão de crédito.

Entretanto, com a introdução deste novo método de transação de valores, surgiram também oportunidades para criminosos desenvolverem estratégias inovadoras de fraude e explorarem potenciais vulnerabilidades de segurança. Fraude é um conceito utilizado para descrever ações deliberadas e enganosas que têm como objetivo a obtenção de ganhos financeiros, vantagens indevidas ou a causação de prejuízos a terceiros, sejam eles indivíduos, entidades ou instituições. A fraude no âmbito dos cartões de crédito pode ser conceituada como: “o uso não autorizado de informações confidenciais de um indivíduo para efetuar compras ou remover fundos de suas contas” (SHARMA et al., 2021).

Essa crescente sofisticação das táticas criminosas acabou minando a confiabilidade dos processos de compra *online*, gerando preocupações legítimas entre consumidores e empresas. Nesse ambiente desafiador, a necessidade de implementar medidas de segurança robustas e aprimorar continuamente os sistemas de pagamento se tornou imperativa para garantir transações seguras e proteger a integridade do comércio eletrônico.

Nesse contexto, a necessidade de promover estudos relacionados a essa situação tornou-se fundamental. O objetivo primordial desses estudos é reduzir o impacto causado por fraudadores, elaborando estratégias eficazes para fortalecer a segurança nas transações *online* e, assim, aumentar a confiança dos usuários no comércio eletrônico. Através da pesquisa e do aprimoramento constante das medidas de proteção, será possível estabelecer um ambiente mais seguro e confiável para o comércio eletrônico, beneficiando tanto os consumidores quanto as empresas.

1.1 Objetivos

O aumento das transações por cartão de crédito no *e-commerce* ressalta a urgência de aprimorar a detecção de fraudes para garantir a segurança das transações eletrônicas. A avalia-

ção detalhada de algoritmos de Aprendizado de Máquina é essencial para proporcionar soluções mais eficazes e robustas diante dos desafios específicos associados à detecção de fraudes em transações por cartão de crédito.

1.1.1 Objetivo Geral

O objetivo geral deste trabalho consiste em avaliar algoritmos de Aprendizado de Máquina no contexto de detecção de fraudes em transações por cartão de crédito no *e-commerce*, visando aprimorar a segurança e a confiabilidade das transações eletrônicas.

1.1.2 Objetivos Específicos

- Desenvolver modelos de Aprendizado de Máquina para detecção de fraudes.
- Analisar o desempenho e a eficiência dos algoritmos escolhidos.
- Identificar as vantagens e desvantagens de cada algoritmo para o problema.

1.2 Contribuições Esperadas

Este trabalho pretende contribuir para o desenvolvimento de sistemas de detecção de fraudes mais seguros, confiáveis e eficientes no contexto de *e-commerce*, o que, por sua vez, contribuirá para a maior segurança e confiabilidade dos seus usuários.

2 REFERENCIAL TEÓRICO

Na era do comércio eletrônico e das transações digitais, o avanço tecnológico proporcionou oportunidades inovadoras, mas também impôs desafios complexos. Um desses desafios cruciais é a ameaça crescente do cibercrime, especialmente nas fraudes relacionadas a transações por cartão de crédito.

2.1 Cibercrime

Durante o século XXI, muitas evoluções e transformações marcaram a sociedade. Com o advento da tecnologia, essas mudanças foram impulsionadas pelo uso da *internet* e as facilidades geradas a partir desta rapidez na comunicação. O mundo virtual se tornou necessário na integração social, tornando o mercado propenso a focalizar seus esforços e investimentos no

mundo digital. Também, com a transição de valiosos recursos, muitos criminosos começaram a se aventurar por este meio, iniciando assim os crimes cibernéticos, ou também conhecidos como cibercrimes.

Os crimes cibernéticos são todos aqueles que utilizam o meio virtual para atacar suas vítimas. Estes podem ser praticados de várias formas com desenvolvimentos diferentes, porém todos visam tirar vantagem do usuário ou do sistema. Segundo Lima (2022), "podemos denominar crimes cibernéticos ou cibercrimes todo ato que envolva computadores ou os meios de tecnologia, utilizados pelos criminosos como objeto de um crime por realizar condutas violadoras de direito privados, que acabam colocando o usuário como vítima de um crime informático.".

Muitos desses crimes capturam dados pessoais e os manuseiam para gerarem uma "falha"de autenticação, onde mesmo com os mecanismos de segurança dos atuais sites, lojas e etc, os fraudadores, por meio dos dados pessoais, conseguem burlar essa defesa.

Estes dados, em sua maioria, são fornecidos pelos próprios usuários que acessam sites inseguros ou desprotegidos e disponibilizam seus próprios dados. "Posto isso, nas fraudes virtuais os usuários são incentivados de forma ingênua a fornecerem informações dos seus dados pessoais em páginas, *e-mails* e mensagens com *links* fraudulentos nas redes sociais. É por esse meio assombroso que o criminoso comete a ação, agindo intencionalmente com o propósito de obter algum benefício material ou até mesmo financeiro", descreve o artigo Lima (2022).

2.2 Fraude por cartão de crédito

Com o crescimento do *e-commerce* e o cartão de crédito estabelecendo-se como um dos métodos de pagamento mais prevalentes para transações *online* no Brasil, torna-se evidente que as fraudes bancárias e de cartão de crédito, frequentemente desencadeadas pelo roubo de dados pessoais, têm se tornado delitos digitais recorrentes no país, segundo Silva e Lima (2020).

Entre as consequências das fraudes efetuadas através do uso de cartão de crédito, destaca-se o *chargeback*. O *chargeback* representa o processo de estorno de um pagamento quando a cobrança é contestada pelo titular do cartão, podendo ocorrer por uma variedade de razões, tais como desacordo comercial, auto-fraude e fraude efetiva. Esse procedimento não apenas coloca uma pressão adicional sobre as instituições financeiras e as empresas, mas também afeta negativamente a experiência do consumidor e a confiança no sistema de pagamento *online*, destacando a importância de medidas eficazes de prevenção e detecção de fraudes.

A fraude efetiva é amplamente compreendida como um ataque cibernético, muitas vezes configurando um estelionato virtual ou a clonagem de cartão, e essa categoria de fraude virtual representa uma das ameaças mais prevalentes e preocupantes na atualidade. Esse tipo de fraude envolve criminosos virtuais habilidosos que buscam explorar vulnerabilidades nos sistemas de pagamento, com o intuito de obter ganhos financeiros ilícitos à custa de consumidores e instituições financeiras. À medida que as tecnologias avançam, a sofisticação das fraudes efetivas continua a evoluir, destacando a necessidade constante de vigilância e medidas de segurança

aprimoradas para proteger a integridade das transações financeiras *online*.

2.3 *E-commerce*

O *e-commerce*, abreviação para comércio eletrônico, é um modelo de negócios que viabiliza a compra e venda de produtos e serviços através da internet. Este engloba diversos formatos, incluindo lojas virtuais, *marketplaces*, transações entre empresas (B2B) e entre empresas e consumidores (B2C), bem como negociações diretas entre consumidores (C2C). O *e-commerce* oferece inúmeras vantagens, como praticidade, ampla variedade de produtos e alcance global, beneficiando tanto consumidores quanto empresas, e se consolidou como um componente fundamental do comércio em escala mundial.

Por essas razões, mesmo diante de adversidades restritivas, como o impacto da pandemia de COVID-19, o comércio eletrônico não apenas se manteve resiliente, mas também floresceu de maneira impressionante, provocando uma transformação significativa no panorama global. Nesse cenário desafiador, o *e-commerce* não apenas sobreviveu, mas se destacou como uma solução essencial para atender às necessidades de consumidores e empresas, impulsionando um novo paradigma nas transações comerciais.

De acordo com o Minjoro (2021), no Brasil, no primeiro semestre de 2021, o setor de comércio eletrônico registrou um faturamento impressionante, atingindo a marca de 53,4 bilhões de reais, ultrapassando o montante total de vendas de todo o ano de 2018. Essa aceleração vertiginosa é um reflexo claro da rápida transformação ocorrida no cenário de compras, à medida que os consumidores buscaram opções *online* para atender às suas necessidades durante os períodos de restrições. Esse fenômeno impulsionou a proliferação de lojas virtuais, ressalta-se que essa tendência beneficiou não apenas as grandes empresas, mas também as micro e pequenas empresas, que compuseram a maioria das lojas físicas fechadas durante a pandemia. Como resultado, elas conquistaram uma parcela significativa do mercado de *e-commerce* no Brasil, demonstrando a resiliência e a adaptabilidade do empresariado local diante de desafios extraordinários.

2.4 *Machine Learning*

Na abordagem da aprendizagem supervisionada, o foco recai sobre a capacidade do sistema em aprender a partir de dados rotulados, permitindo predições precisas em novas instâncias. De acordo com Hastie et al. (2009), o aprendizado supervisionado é uma técnica de aprendizado de máquina que pode ser usada para prever uma ou mais variáveis de saída com base em uma ou mais variáveis de entrada. Essa técnica também pode ser usada para classificar dados, atribuindo cada registro a uma categoria específica e possui muitas aplicações em áreas como finanças, *marketing*, medicina, ciência de dados e muito mais, assim como detecção de

fraude.

Quando se trata de avaliar o desempenho de modelos de aprendizado de máquina, diversas métricas são aplicadas. A acurácia mensura a proporção de previsões corretas em relação ao total de observações, enquanto o *recall* quantifica a capacidade do modelo em identificar corretamente as instâncias da classe positiva. A precisão, por sua vez, mede a proporção de instâncias positivas corretamente identificadas em relação ao total de instâncias identificadas como positivas pelo modelo. Por fim, o *F1-Score* é uma métrica que combina precisão e *recall*, oferecendo uma visão equilibrada do desempenho do modelo, sendo particularmente útil quando há desigualdade entre as classes (HASTIE et al., 2009).

Em suma, a aprendizagem supervisionada aliada a algoritmos de *Machine Learning*, juntamente com a análise criteriosa das métricas de avaliação, representa uma abordagem abrangente na construção de modelos preditivos eficientes e precisos.

2.5 Trabalhos Relacionados

Alguns estudos já foram conduzidos com a finalidade de identificar métodos eficazes para mitigar fraudes em transações com cartão de crédito. Por exemplo, o estudo realizado por Nicola et al. (2020) adotou uma abordagem empírica que envolveu a avaliação de diversos classificadores e a aplicação de métodos de balanceamento de dados para a detecção de fraudes. Segundo o mesmo, o algoritmo *Random Forest* foi o classificador que obteve o maior F-score e foi bastante consistente em atingir esse patamar em conjuntos de treino balanceados com sobreamostragem. Além disso, nenhum dos demais classificadores atingiu desempenhos estatisticamente equivalentes ao F-Score ótimo da *Random Forest*. "Os melhores resultados foram obtidos com modelos de *Random Forest* no conjunto desbalanceado e nos balanceados baseados em sobreamostragem ou híbridos. Uma vantagem adicional desse classificador foi sua maior robustez em relação à escolha das configurações de balanceamento e seleção de atributos"(NICOLA et al., 2020).

No mesmo contexto, a pesquisa conduzida por Guimarães (2022) que usou o conjunto de dados desbalanceados "dataset aberto do Kaggle (Credit Card Fraud Detection, 2018)", apresentou um modelo de *Machine Learning* baseado em Regressão Logística e Árvore de Decisão. Neste caso, foi empregado técnicas de balanceamento "ADASYN"e "SMOTE", "o ADASYN concentra-se em gerar amostras próximas as amostras originais que são classificadas erroneamente usando o classificador *k-Nearest Neighbors*, enquanto a implementação do SMOTE não fará nenhuma distinção entre amostras fáceis e difíceis a serem classificadas usando a regra dos vizinhos mais próximos", menciona Guimarães (2022).

O estudo conduzido por Mishra e Ghorpade (2018), intitulado "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques", empregou um conjunto de dados real de transações de cartão de crédito, abrangendo operações realizadas por titulares de cartões europeus em setembro de 2013. O balanceamento foi realizado através

da técnica de *Random Undersampling*, enquanto os algoritmos investigados incluíram *Logistic Regression, Decision Tree, Random Forest*, e *Support Vector Machines*.

Outro estudo relevante e utilizado para contribuir com o trabalho atual foi o artigo escrito por Horta e Loiola (2022), onde também utilizou um conjunto de dados abertos obtidos da plataforma de colaboração entre cientistas de dados chamada Kaggle. Este conjunto de dados foi disponibilizado de forma desbalanceada e então o autor utilizou técnicas de balanceamento para o estudo. De acordo com Horta e Loiola (2022), "Foram aplicadas técnicas de sobre-amostragem da classe minoritária, SMOTE e Borderline-SMOTE". Neste estudo apresentado, foram utilizados os algoritmos *Extra Tree* e *Random Forest* de *Machine Learning*.

Estes estudos enfatizaram a importância de futuras pesquisas, incentivando a exploração de novos conjuntos de dados, variáveis e métodos/técnicas diferentes. Além disso, todos estes recomendaram a realização de experimentos que explorem combinações entre esses elementos, visando aprimorar ainda mais a detecção e prevenção de fraudes no ambiente de transações com cartões de crédito. Essa abordagem de pesquisa contínua é fundamental para manter a eficácia e a relevância das soluções de segurança no setor financeiro.

3 METODOLOGIA

Este trabalho trata-se de um estudo de caso de caráter exploratório e experimental. Como mostrado na Figura 1, inicialmente, foi realizado a coleta de dados relacionados às transações de compras por cartão de crédito da empresa de *e-commerce*. Isso inclui informações sobre as transações legítimas e qualquer histórico de transações fraudulentas. Também, foi realizado uma etapa de pré-processamento nos dados coletados, que envolveu a limpeza dos dados, tratamento de valores ausentes, normalização de dados e codificação de variáveis categóricas, além do balanceamento que visa igualar a classe majoritária à minoritária. Foi certificado que os dados estavam prontos para serem utilizados nos algoritmos de detecção de fraudes.

Em seguida, foram identificados possíveis algoritmos de detecção de fraudes que sejam mais apropriados para transações de compra *online* por cartão de crédito. Como critério de seleção, foram selecionados algoritmos supervisionados.

Depois, ocorreu o treinamento dos algoritmos selecionados utilizando os dados de transações preparados. O conjunto de dados foi dividido em conjuntos de treinamento e teste para avaliar o desempenho dos modelos. Como resultado, foi utilizado a Matriz de Confusão para avaliar o desempenho dos algoritmos.

Posteriormente, foi conduzida uma análise do desempenho dos algoritmos, utilizando as métricas de acurácia, *recall*, precisão e *F1-Score* considerando o contexto de detecção de fraudes em transações de compra *online* por cartão de crédito.

Por fim, foram realizados experimentos com auxílio de bibliotecas de dados. Foram comparados os resultados dos modelos em termos de eficácia na detecção de fraudes específicas desse setor.

Figura 1 – Passo a Passo - Metodologia

Fonte: Figura do autor

4 DESENVOLVIMENTO

Nesta seção será apresentado a etapa de desenvolvimento do projeto, desde a coleta de dados e os resultados obtidos

Foi escolhido empregar o *Google Colab*, uma plataforma baseada na nuvem fornecida pelo *Google*, como ambiente de teste para criação e execução de *notebooks Jupyter*, no desenvolvimento deste projeto foi fundamentada nas suas distintas vantagens. Entre estas, destacam-se a facilidade de colaboração simultânea, o acesso gratuito baseado na nuvem, a integração eficiente com o *Google Drive* para armazenamento centralizado, a presença de bibliotecas populares pré-instaladas e o suporte a múltiplas linguagens, sendo uma delas o *Python*. Essas características proporcionam um ambiente compartilhado e eficaz para o desenvolvimento colaborativo do trabalho.

Concomitantemente, o grupo escolheu o *Python* como linguagem de desenvolvimento, pois apresentam os seguintes pontos positivos:

- Bibliotecas de Aprendizado de Máquina: *Python* oferece uma extensa variedade de bibliotecas especializadas em *Machine Learning*, como *Scikit-Learn*, *TensorFlow*, *Keras*, *PyTorch* e outras. Essas bibliotecas fornecem ferramentas e estruturas que simplificam a implementação de algoritmos de *Machine Learning*. No âmbito deste trabalho, foi empregada a biblioteca *Scikit-Learn* em virtude de seu desempenho destacado e facilidade de manipulação. Dentro dessa biblioteca, foram importados os métodos *LogisticRegression*, *SVM* e *RandomForestClassifier*.

- Comunidade Engajada: *Python* conta com uma comunidade de desenvolvedores ativa e dedicada. Isso significa que é mais fácil encontrar suporte, documentação e recursos *online*. Além disso, há uma abundância de tutoriais, trabalhos e exemplos disponíveis. Em resumo, *Python* é uma escolha amplamente preferida para estudos de *Machine Learning* devido à abundância de bibliotecas, facilidade de uso e suporte da comunidade. Esses fatores fazem dele a linguagem ideal para o desenvolvimento deste projeto.

4.1 Coleta e Pré-processamento de Dados

Os dados necessários para este estudo foram diretamente coletados do *Data Warehouse* (DW) da Empresa, cujo nome será preservado por questões de confidencialidade. O DW está hospedado na plataforma *Amazon Redshift*, sendo uma infraestrutura que reúne e armazena dados de diversas fontes da empresa, permitindo análises avançadas e tomada de decisões. Para a extração dos dados, foi utilizado a linguagem SQL (*Structured Query Language*) como meio de consulta ao banco de dados relacional. Os dados detalhados foram exportados para um arquivo XLSX, que serviu como a conjunto de dados para a análise subsequente por meio de algoritmos de *Machine Learning*. A escolha do formato XLSX permite uma fácil manipulação e processamento dos dados, garantindo a eficiência na análise e detecção de fraudes.

O grupo teve acesso controlado aos dados, seguindo os mais rigorosos protocolos de segurança e ética para garantir a privacidade e a conformidade com regulamentações. Isso incluiu a aplicação de medidas rigorosas de segurança de dados e o cumprimento de regulamentações relevantes, como a Lei Geral de Proteção de Dados (LGPD), se aplicável (BRASIL, 2018). Por este motivo não foi possível a disponibilização dos dados ou do ambiente de testes desenvolvido para o experimento.

Os dados coletados incluem uma ampla gama de informações relevantes para a pesquisa, tais como:

- Dispositivo de compra.
- Bandeira do cartão de crédito utilizado.
- Valor da compra.
- Informações de parcelamento, incluindo a quantidade de parcelas.
- Data e horário da compra.
- Histórico de compra do usuário.
- Segmentação do produto comprado.
- Informações relacionadas a *chargebacks*, incluindo a categoria do *chargeback*, como "fraude"ou "insatisfação do cliente".

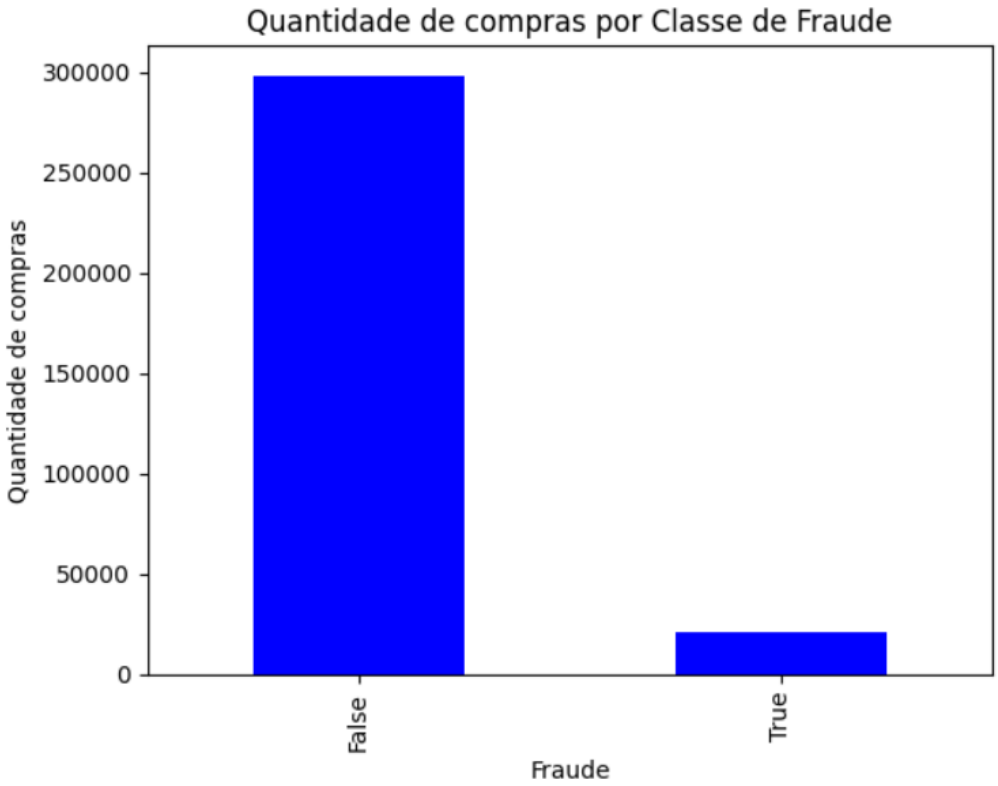
Além disso, para a realização deste estudo, o grupo optou por usar um conjunto de dados específico do período de 2022, uma vez que se trata de um ano já completo. Isso envolve a seleção de amostras mensais proporcionais, coletadas de forma aleatória, totalizando 21.345 compras identificadas como fraudulentas e 298.307 compras legítimas, sendo assim, 319.652 registros. Essa escolha de números se deve pelo fato de ser usado a maior quantidade possíveis de registros fraudulentos para treinamento do modelo. Foi entendido como fraude compras que foram posteriormente identificadas como *chargeback* na categoria de Fraude Deliberada.

Dito isso, o conjunto de dados utilizado permitiu a aplicação do aprendizado supervisionado e assim analisar os padrões de fraude ao longo de um ano (2022) e o ajuste aos modelos de detecção de acordo com as mudanças sazonais. Os dados detalhados desempenharam um papel fundamental na análise de detecção de fraudes, permitindo a aplicação de algoritmos de *Machine Learning* para identificar padrões de comportamento suspeito.

4.1.1 Balanceamento

Como evidenciado na Figura 2 apresentada a seguir, observa-se uma discrepância que resulta no desbalanceamento de dados entre as classes de fraude e não fraude. O desbalanceamento de dados ocorre quando as classes não são representadas de maneira equitativa. Esta situação pode levar à criação de modelos tendenciosos que favorecem a classe majoritária, resultando em desempenho inferior na detecção da classe minoritária (HASTIE et al., 2009).

Figura 2 – Amostra dos dados

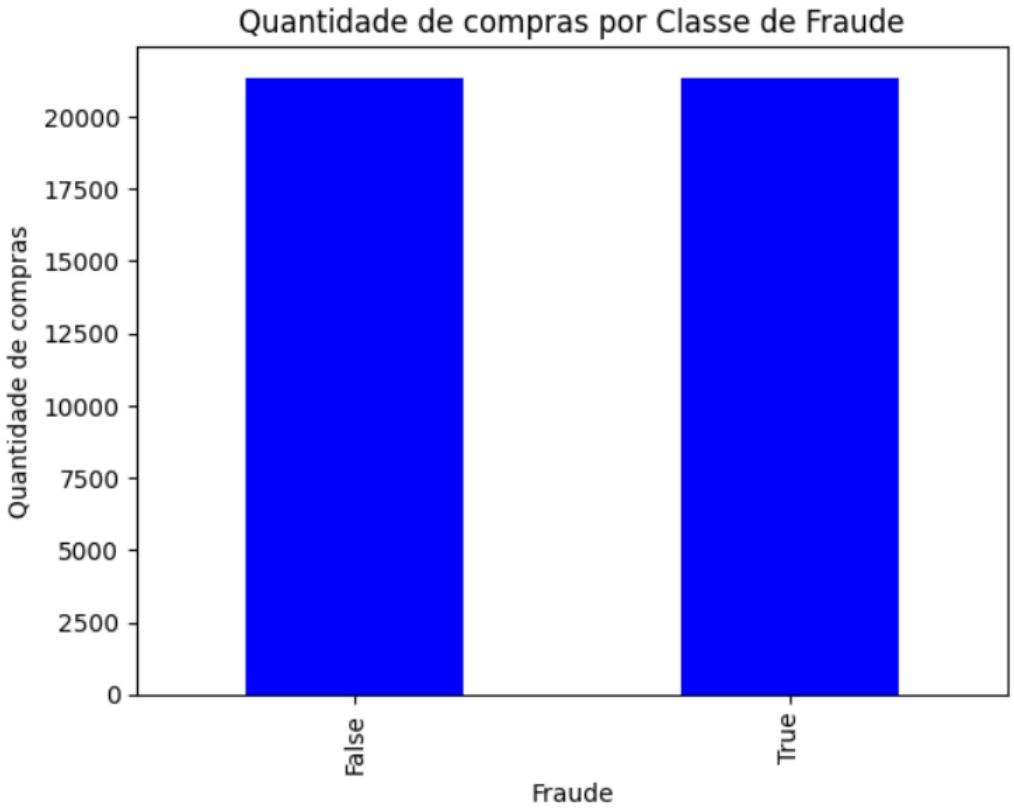


Fonte: Resultados originais da pesquisa (2023)

Devido a isso é necessário realizar o que é chamado de balanceamento de dados. O objetivo do balanceamento de dados é garantir que o modelo seja treinado de maneira justa e eficaz para todas as classes, evitando que seja enviesado em direção às classes mais numerosas. A técnica de balanceamento de dados utilizada foi o *undersampling*. Esta técnica remove exemplos da classe majoritária para equilibrar o número de exemplos em ambas as classes. Na

Figura 3 a seguir, o resultado do balanceamento do conjunto de dados.

Figura 3 – Amostra dos dados balanceados



Fonte: Resultados originais da pesquisa (2023)

4.1.2 Encoding

O conjunto de dados em questão é composta por um total de 18 colunas, distribuídas entre 8 variáveis categóricas e 10 variáveis numéricas. Em virtude dessa distinção, torna-se necessário a aplicação da técnica de pré-processamento denominada *encoding*. O *encoding* constitui uma abordagem essencial no contexto da manipulação de dados, sendo empregado com o propósito de converter informações categoricamente classificadas em representações numéricas. Tal procedimento se mostra indispensável devido à limitação de diversos algoritmos de Aprendizado de Máquina, os quais requerem a apresentação dos dados em formato numérico para seu efetivo processamento e análise. Nesse sentido, a adoção do *encoding* emerge como uma medida crucial para viabilizar a compatibilidade entre as características distintas do conjunto de dados e os requisitos algorítmicos, contribuindo assim para a eficácia do processo de Aprendizado de Máquina.

O método de *encoding* utilizado foi o *Label Enconding* que é uma técnica usada para transformar dados categóricos em rótulos numéricos para algoritmos de aprendizado de má-

quina. Ele atribui um valor inteiro exclusivo a cada categoria no conjunto de dados. Na Figura 4 a seguir é possível verificar as linhas iniciais do conjunto de dados após realizar o *encoding*, os nomes dos campos originais foram mascarados devido a confidencialidade.

Figura 4 – Amostra detalhada dos dados

f1	f2	f3	f4	f5	f6	valor_compra	f7	f8	f9	f10	f11	f12	f13	f14	fraude	f15	f16
0	0	1	5	47	13	17.50	1	1	12	7	1	0	19	2595	0	0.0	0
1	20	1	6	122	21	60.00	1	1	15	7	1	0	18	753	0	0.0	0
8	11	1	5	121	4	108.88	4	2	4	7	1	0	4	1384	0	0.0	0
4	16	1	7	40	4	17.50	1	1	27	7	1	0	26	2435	0	0.0	0
0	17	1	5	121	4	34.74	2	1	6	7	1	0	6	568	0	88.0	2

Fonte: Resultados originais da pesquisa (2023)

4.2 Seleção de Algoritmos

A escolha dos algoritmos a serem utilizados desempenha um papel crucial no desenvolvimento de um sistema de detecção de fraudes por cartão de crédito. No estudo, foram selecionados três algoritmos distintos para abordar essa tarefa: *Logistic Regression*, *Random Forest* e *Support Vector Machine* (SVM). Cada um desses algoritmos oferece abordagens únicas para identificar padrões de fraude nos dados.

Logistic Regression é uma técnica de aprendizado supervisionado comum que é usada para modelar a probabilidade de um evento ocorrer. O modelo de regressão logística surge do desejo de modelar as probabilidades posteriores das K classes por meio de funções lineares em x, ao mesmo tempo garantindo que elas somem para um e permaneçam no intervalo [0, 1] (HASTIE et al., 2009). No contexto de detecção de fraudes, a *Logistic Regression* pode ser aplicada para prever a probabilidade de uma transação ser fraudulenta. Isso é útil quando se deseja quantificar o grau de suspeita em vez de uma simples classificação binária.

Random Forest ou Florestas Aleatórias são uma combinação de preditores de árvores de decisão, de modo que cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores na floresta (BREIMAN, 2001). Na detecção de fraudes, o *Random Forest* é interessante devido à sua capacidade de identificar padrões não-lineares, manejo eficaz de grandes conjuntos de características e incorporação de regularização.

Support Vector Machine (SVM) é um algoritmo utilizado para tarefas de classificação e regressão. SVM é um método poderoso para construir um classificador. Ele visa criar uma fronteira de decisão entre duas classes que permite a previsão de rótulos a partir de um ou mais vetores de características (HUANG et al., 2018). SVM possui uma capacidade de lidar com dados de alta dimensão, eficácia em espaços de alta dimensão, boa generalização, busca pela

margem máxima entre as classes, e baixa sensibilidade a *overfitting*.

A comparação destes algoritmos permite uma abordagem abrangente para a detecção de fraudes por cartão de crédito. Cada algoritmo aborda o problema de maneira diferente, complementando as capacidades uns dos outros. Durante a fase de treinamento e avaliação, a eficácia de cada algoritmo foi testada para determinar qual deles oferece o melhor desempenho na detecção de fraudes com base nos dados disponíveis.

Neste trabalho, a equipe optou por utilizar bibliotecas específicas em Python para processar os dados, implementar os algoritmos escolhidos, gerar gráficos e seus respectivos resultados. Uma biblioteca importante foi a *Scikit-Learn*, devido à sua ampla adoção na comunidade de Aprendizado de Máquina, eficácia comprovada na detecção de anomalias e facilidade de uso. Para gerar os gráfico foi utilizada a biblioteca *Matplotlib* e para o processamento *Pandas* e *Numpy*.

4.3 Treinamento dos Modelos

Após proceder à organização meticulosa do conjunto de dados, implementando estratégias de correção e aprimoramento, o conjunto de dados foi dividido em duas partes distintas: um conjunto de treinamento e um conjunto de teste, na proporção de 80/20 respectivamente. Essa divisão é fundamental para avaliar o desempenho dos modelos em um ambiente controlado.

O conjunto de treinamento foi utilizado para ensinar os modelos a reconhecerem padrões de transações legítimas e fraudulentas. Os algoritmos aprendem com exemplos passados, permitindo-lhes generalizar esses padrões para novas transações.

O conjunto de teste foi reservado para avaliar quão bem os modelos generalizam os padrões aprendidos. Ele contém transações que não foram vistas pelos modelos durante o treinamento e são usados para avaliar a capacidade dos modelos de detecção de fraudes. Em seguida, foi feita a Matriz de Confusão, que consiste em fornecer métricas avaliativas da qualidade dos resultados obtidos.

4.4 Avaliação e Comparação do Desempenho

Matrizes de confusão (MC) são utilizadas para avaliar o relacionamento entre duas ou mais variáveis nominais, isto é, se pertence ou não a uma determinada classe (OLIVA, 2018). Ela compara as previsões de um modelo com os resultados reais para avaliar seu desempenho. A matriz destaca quatro elementos principais:

- Verdadeiros positivos (TP): Casos em que o modelo previu corretamente a classe positiva.
- Falsos positivos (FP): Casos em que o modelo previu incorretamente a classe positiva, quando a verdadeira classe era negativa.

- Verdadeiros negativos (TN): Casos em que o modelo previu corretamente a classe negativa.
- Falsos negativos (FN): Casos em que o modelo previu incorretamente a classe negativa, quando a verdadeira classe era positiva.

Figura 5 – Matriz de Confusão

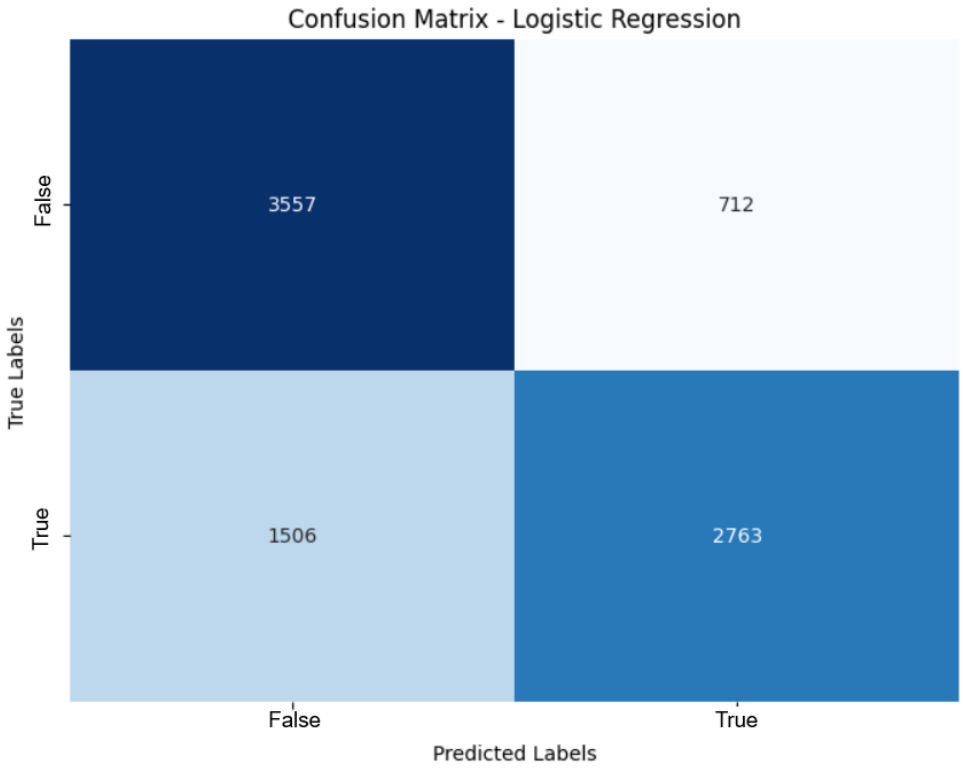
		Valores preditos	
		Fraude	Não Fraude
Valores reais	Fraude	Verdadeiro positivo (TP)	Falso negativo (FN)
	Não Fraude	Falso positivo (FP)	Verdadeiro negativo (TN)

Fonte: Figura do autor

A utilidade da matriz de confusão reside na sua capacidade de fornecer uma visão detalhada do desempenho do modelo, indo além de métricas globais como a acurácia. Com base nos elementos da matriz, é possível calcular métricas específicas, como acurácia, precisão, *recall* e *F1-Score*. Essas métricas foram empregadas no presente trabalho como critérios fundamentais para avaliação dos modelos propostos, permitindo uma análise mais abrangente e refinada. A partir dessas métricas, foi possível comparar o desempenho do presente trabalho com seus trabalhos relacionados, contribuindo para uma compreensão aprofundada da eficácia na detecção de fraudes por cartão de crédito.

A Figura 6 apresenta o resultado da matriz de confusão do algoritmo Regressão Logística. Ao examinar os resultados, observa-se que 3.557 transações legítimas foram corretamente classificadas como não fraudulentas (TN), enquanto 712 transações foram erroneamente categorizadas como fraudulentas (FP). Em contrapartida. o modelo acertou ao identificar 2.763 transações fraudulentas (TP), mas também falhou ao deixar passar 1.506 transações fraudulentas, classificadas de forma equivocada como legítimas (FN). Totalizando assim, 8.538 transações, que correspondem a 20% (conjunto de teste) do total de transações balanceadas.

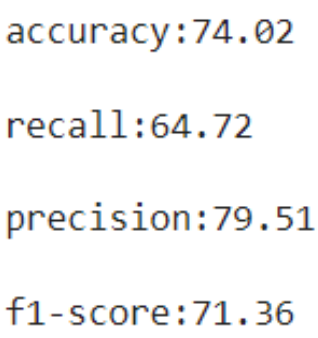
Figura 6 – Matriz de Confusão - Regressão Logística



Fonte: Resultados originais da pesquisa (2023)

Na Figura 7 estão apresentados as métricas de avaliação acurácia, *recall*, precisão e *F1-score* evidenciando uma acurácia de 74,02%, indicando que cerca de três quartos das previsões foram corretamente classificadas. O *recall* de 64,72% destaca a capacidade do modelo em identificar a proporção correspondente de transações fraudulentas em relação ao total de fraudes existentes. A precisão atingiu 79,51%, indicando a confiabilidade nas previsões positivas, enquanto o *F1-Score* de 71,36% proporciona uma avaliação equilibrada entre precisão e *recall*.

Figura 7 – Métricas - Regressão Logística

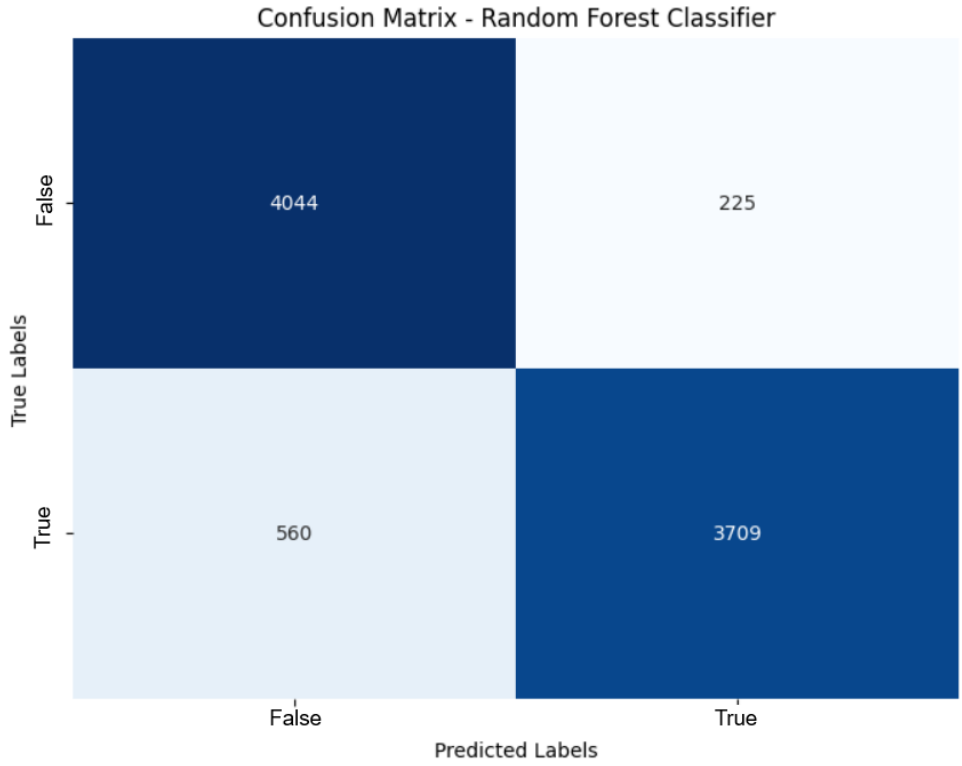


Fonte: Resultados originais da pesquisa (2023)

A avaliação do desempenho do algoritmo *Random Forest*, em comparação com o mo-

delo anterior de Regressão Logística, por meio da Matriz de Confusão (Figura 8) revela melhorias notáveis em seus 4 elementos. O *Random Forest* demonstrou acurácia ao classificar corretamente 3.709 transações fraudulentas como positivas (TP). Além disso, registrou 4.044 transações legítimas corretamente classificadas como não fraudulentas (TN). A redução significativa nos falsos positivos, com apenas 225 transações legítimas erroneamente categorizadas como fraudulentas (FP) e também houve uma queda na quantidade de transações fraudulentas classificadas erroneamente como legítimas (FN) para 560 registros.

Figura 8 – Matriz de Confusão - *Random Forest*



Fonte: Resultados originais da pesquisa (2023)

Como já evidenciado na matriz de confusão, pode-se ver na Figura 9 a significativa melhoria do modelo utilizando o algoritmo de *Random Forest* em relação a Regressão logística. Com acurácia de 90,81%, indicando uma taxa global de previsões corretas, enquanto o *recall* de 86,88% destaca sua eficácia em identificar a maioria das transações fraudulentas. Com uma precisão de 94,28%, o modelo demonstra confiabilidade nas previsões positivas, equilibrada pelo *F1-Score* de 90,43%, sugerindo uma abordagem eficaz na detecção equilibrada de fraudes.

Figura 9 – Métricas - *Random Forest*

accuracy:90.81

recall:86.88

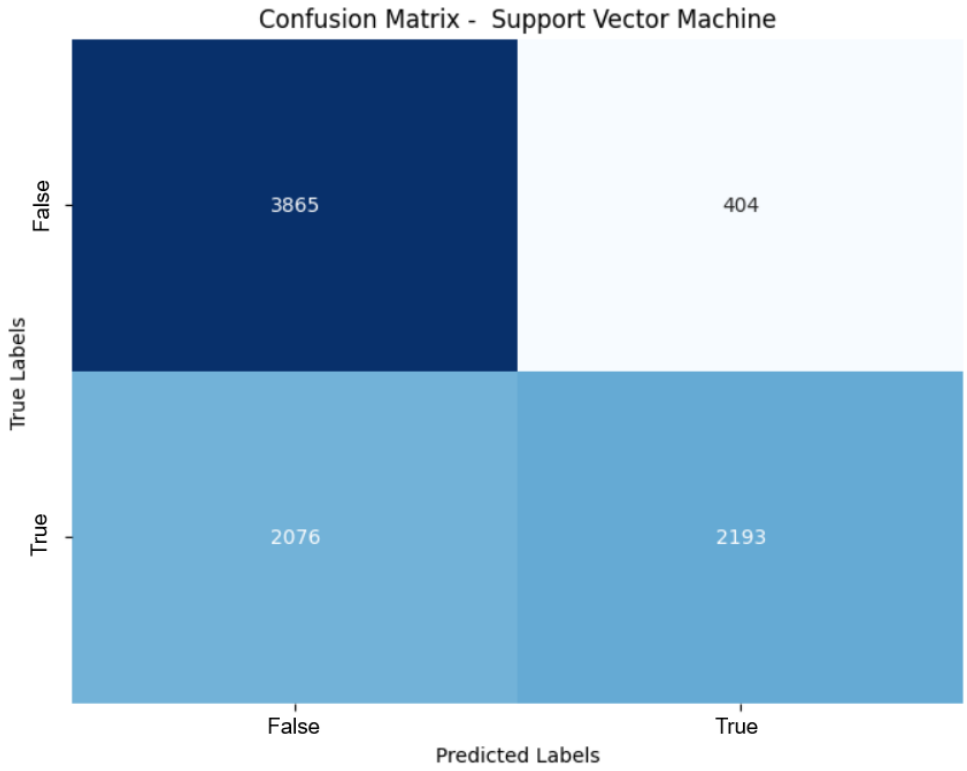
precision:94.28

f1-score:90.43

Fonte: Resultados originais da pesquisa (2023)

O modelo utilizando o algoritmo SVM como mostra a Figura 10, destaca-se negativamente pelo alto número de falsos negativos totalizando 2076 transações. Além disso teve desempenho de 3865 Verdadeiros negativos (TN), com 404 falsos positivos e 2.193 para verdadeiros positivos.

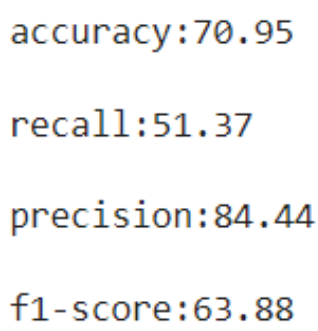
Figura 10 – Matriz de Confusão - SVM



Fonte: Resultados originais da pesquisa (2023)

Quanto a métricas de desempenho o SVM, como evidenciado na Figura 11, demonstrou acurácia de 70,95% além de um *recall* abaixo dos demais com 51,37%, precisão de 84,44% e *F1-Score* de 63,88%.

Figura 11 – Métricas - SVM



Fonte: Resultados originais da pesquisa (2023)

Com base nos resultados apresentados na Tabela 1, o algoritmo *Random Forest* apresentou o melhor desempenho para a detecção de fraudes em transações de cartão de crédito. O modelo conseguiu atingir acurácia de 90,81%, *recall* de 86,88% e *F1-Score* de 90,43%. O algoritmo Regressão Logística também apresenta um bom desempenho, com acurácia de 74,02%, *recall* de 64,72% e *F1-Score* de 71,36%. Já algoritmo SVM apresentou um desempenho abaixo dos demais, com uma acurácia de 70,95%, *recall* de 51,37% e *F1-Score* de 63,88%.

Tabela 1 – Comparação de Resultados

Algoritmo	**Acurácia**	***Recall***	***Precisão***	***F1-score***
Regressão Logística	74,02%	64,72%	79,51%	71,36%
Random Forest	90,81%	86,88%	94,28%	90,43%
SVM	70,95%	51,37%	84,44%	63,88%

Fonte: Fonte: Resultados originais da pesquisa (2023)

4.5 Análise e Discussão dos Resultados

Nos estudos de Nicola et al. (2020) e Mishra e Ghorpade (2018), observa-se que o algoritmo *Random Forest* destacou-se como o mais eficaz na detecção de fraudes em conjuntos de dados distintos. No entanto, é relevante notar que, no primeiro estudo, foram empregadas técnicas de balanceamento diferentes ou aplicadas em conjuntos de dados desbalanceados.

Assim como na presente pesquisa, utilizou-se o *F1-Score* como métrica de avaliação. O resultado considerado ótimo foi de 0,87 (ou 87%) para dados desbalanceados e para conjuntos balanceados por meio de sobreamostragem, enquanto alcançou-se um *F1-Score* de 90,43% em dados balanceados por *undersampling* neste estudo. No estudo de Mishra e Ghorpade (2018), o *recall* foi utilizado, alcançando a marca de 96%, enquanto no presente trabalho obteve-se um desempenho aproximado de 86,88%.

Adicionalmente, o estudo conduzido por Horta e Loiola (2022) explorou outras técnicas de balanceamento não empregadas na presente pesquisa, com o mesmo propósito de detecção de fraudes, mas com o intuito de avaliar essas técnicas em conjuntos de dados desbalanceadas.

Por fim, o trabalho de Guimarães (2022) destacou o algoritmo de Regressão Logística, utilizando dados balanceados com o método ADASYN, como o de melhor desempenho, registrando 91% de acurácia. No entanto, no presente estudo, esse algoritmo apresentou um desempenho inferior, atingindo apenas 74%.

É fundamental ressaltar que cada um dos artigos mencionados utilizou ambientes de testes distintos e, sobretudo, conjunto de dados diferentes. Portanto, a escolha das melhores técnicas deve ser avaliada caso a caso, considerando as particularidades de cada contexto.

5 CONCLUSÃO E TRABALHOS FUTUROS

Este estudo introduziu e analisou a aplicação de três algoritmos distintos de Aprendizado de Máquina no âmbito da detecção de fraudes em transações com cartão de crédito no ambiente de comércio eletrônico. Além disso, foram implementadas técnicas *encoding* e balanceamento de dados, proporcionando contribuições significativas para aprimorar as práticas de segurança e confiabilidade nas transações realizadas no comércio eletrônico.

Com base no que foi desenvolvido neste estudo, como resultado obteve-se o algoritmo *Random Forest* como o melhor. Considerando o conjunto de dados da Empresa, utilizando a técnica de *Label Encoding* e *Underslaping* para o *encoding* e balanceamento dos dados, este algoritmo obteve os melhores resultados com 90,81% de acurácia, 86,88% de *recall*, 94,28% de precisão e 90,43% de *F1-Score*. Esses resultados indicam que o objetivo geral da pesquisa, de avaliação dos algoritmos no contexto de detecção de fraude no conjunto de dados coletado da Empresa, foi alcançado.

Propõe-se, para estudos futuros, a utilização de um conjunto de dados diferente com novas *features* que possam agregar na detecção de fraude. Adicionalmente, sugere-se a investigação de diferentes técnicas de codificação, como o *One-Hot Encoding* ou *Target Encoding*, e a exploração de outras estratégias de balanceamento, incluindo ADASYN, SMOTE e *oversampling*. Além disso, a expansão do escopo para abranger outras técnicas e algoritmos de Aprendizado de Máquina, como o *Isolation Forest*, Redes Neurais, entre outros.

Referências

BRASIL. **Lei Geral de Proteção de Dados (LGPD)**. 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 10 2001.

GUIMARÃES, Mariana Araujo. Detecção de fraude em aplicativos de e-commerce. Universidade Presbiteriana Mackenzie, 2022.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning**. New York: Springer, 2009. 745 p. ISBN 978-0387-95457-0.

HORTA, Guilherme G; LOIOLA, Murilo B. Efeitos da sobre-amostragem na sobreposição e classificação de dados de fraudes em cartões de crédito. 2022.

HUANG, SHUJUN et al. Applications of support vector machine (svm) learning in cancer genomics. **Cancer Genomics & Proteomics**, International Institute of Anticancer Research, v. 15, n. 1, p. 41–51, 2018. ISSN 1109-6535. Disponível em: <https://cgp.iiarjournals.org/content/15/1/41>.

LEBLEBICI, Huseyin. The evolution of alternative business models and the legitimization of universal credit card industry: Exploring the contested terrain where history and strategy meet. In: ____. Emerald Group Publishing Limited, 2012. (Advances in Strategic Management, v. 29), p. 117–151. Disponível em: <https://doi.org/10.1108/S0742-3322(2012)0000029009>.

LIMA, Milena Angela Santos. Cibercrimes: a vulnerabilidade dos usuários. Pontifícia Universidade Católica de Goiás, 2022.

MINJORO, Mariana. A evolução do mercado de e-commerce no brasil e como a pandemia do covid-19 impactou esse processo. 2021.

MISHRA, Ankit; GHORPADE, Chaitanya. Credit card fraud detection on the skewed data using various classification and ensemble techniques. In: IEEE. **2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)**. [S.l.], 2018. p. 1–5.

NICOLA, Victor; LAURETTO, Marcelo; DELGADO, Karina Valdivia. Avaliação empírica de classificadores e métodos de balanceamento para detecção de fraudes em transações com cartões de créditos. In: SBC. **Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional**. [S.l.], 2020. p. 70–81.

OLIVA, Jefferson Tales. **Geração automática de laudos médicos para o diagnóstico de epilepsia por meio do processamento de eletroencefalogramas utilizando aprendizado de máquina**. 2018. Tese (Tese) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, acesso em: 2023-11-29.

SHARMA, Mudita et al. Credit card fraud detection using machine learning algorithms. In: SINGH, Jawar; KUMAR, Sudhir; CHOUDHURY, Umakanta (Ed.). **Innovations in Cyber Physical Systems**. Singapore: Springer Singapore, 2021. p. 547–560. ISBN 978-981-16-4149-7.

SILVA, Jefferson David dos Anjos; LIMA, Maria Vitória Ribas de Oliveira. Os principais cibercrimes praticados no brasil. 2020.