IDENTIFICAÇÃO DE TRANSAÇÕES FRAUDULENTAS EM CARTÕES DE CRÉDITO UTILIZANDO APRENDIZADO DE MÁQUINA

ALINE BERTOLAZO DOS SANTOS[1], ISADORA DISPOSTI BUENO DOS SANTOS[2], KARINA MITIKO TOMA[3]

[1] Graduando em Engenharia da Computação, IFSP, Câmpus Birigui, a.bertolazo@aluno.ifsp.edu.br
[2] Graduando em Engenharia da Computação, IFSP, Câmpus Birigui, isadora.bueno@aluno.ifsp.edu.br
[3] Mestre em Ciências da Computação, Universidade Federal de São Carlos, karinamt@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.02.02.06-4 Regressão e Correlação

RESUMO: O artigo investiga a identificação de transações fraudulentas em cartões de crédito por meio de técnicas de aprendizado de máquina, com ênfase na aplicação de um modelo de Regressão Logística (RL). A pesquisa ressalta a importância da normalização das variáveis e do balanceamento das classes, uma vez que apenas 0,17% das transações analisadas foram classificadas como fraudulentas, o que pode impactar a eficácia do modelo. Os resultados obtidos indicam uma precisão de 92,50%, *recall* de 82,40% e *F1-Score* de 87,15%, evidenciando um bom equilíbrio entre a detecção de fraudes e a minimização de falsos positivos. Embora algoritmos de aprendizado profundo, como redes neurais, possam oferecer desempenho superior em grandes conjuntos de dados, a escolha da RL foi motivada por sua eficiência e interpretabilidade, características essenciais em ambientes que requerem modelos transparentes. O estudo conclui que, apesar da predominância de métodos mais complexos, abordagens tradicionais como a RL permanecem relevantes e eficazes na detecção de fraudes, especialmente quando aplicadas com técnicas adequadas de pré-processamento, contribuindo para a segurança do sistema financeiro.

PALAVRAS-CHAVE: Regressão Logística; Validação Cruzada, Algoritmo.

IDENTIFYING FRAUDULENT TRANSACTIONS ON CREDIT CARDS USING MACHINE LEARNING

ABSTRACT: The article investigates the identification of fraudulent credit card transactions using machine learning techniques, with an emphasis on the application of a Logistic Regression (LR) model. The research highlights the importance of normalizing variables and balancing classes, since only 0.17% of the transactions analyzed were classified as fraudulent, which can impact the effectiveness of the model. The results obtained indicate a precision of 92.50%, recall of 82.40% and F1-Score of 87.15%, demonstrating a good balance between fraud detection and minimization of false positives. Although deep learning algorithms such as neural networks can offer superior performance on large data sets, the choice of LR was motivated by its efficiency and interpretability, essential characteristics in environments that require transparent models. The study concludes that, despite the predominance of more complex methods, traditional approaches such as LR remain relevant and effective in detecting fraud, especially when applied with appropriate pre-processing techniques, contributing to the security of the financial system.

KEYWORDS: Fraud; Logistic Regression; Cross Validation,Algorithm.

INTRODUÇÃO

De acordo com Alarfaj et al. (2022), algoritmos de aprendizado de máquina e de aprendizado profundo têm sido amplamente explorados na detecção de fraudes, demonstrando resultados promissores em termos de precisão e velocidade. Embora algoritmos de aprendizado profundo

mostram resultados promissores, métodos mais simples, como a Regressão Logística (RL), continuam sendo relevantes devido à sua menor complexidade computacional e facilidade de implementação, o que os torna adequados para ambientes com restrições de recursos ou onde a interpretabilidade do modelo é essencial.

A detecção de fraudes em transações com cartão de crédito é um desafio crítico no setor financeiro, devido ao impacto econômico e à necessidade crescente de garantir a segurança das transações. De acordo com um estudo da ClearSale, empresa especializada em inteligência de dados para prevenção de fraudes, foram analisadas 277,4 milhões de transações em 2023. Dentre elas, foram registradas 3,7 milhões de tentativas de fraude, totalizando aproximadamente R$ 3,5 bilhões em tentativas fraudulentas.

Este trabalho tem como objetivo o desenvolvimento de uma solução de detecção de fraudes em transações com cartão de crédito, utilizando *Python* e a técnica de RL uma abordagem estatística amplamente empregada para classificação binária. A escolha desta técnica é justificada não apenas pela sua simplicidade e interpretabilidade, mas também por ser uma técnica consolidada na identificação de padrões em dados rotulados, especialmente em cenários onde a explicabilidade do modelo é uma prioridade, caracterizando transações legítimas e fraudulentas (Morettin; Singer, 2022).

Gupta (2023) destaca que a eficácia dos algoritmos de aprendizado de máquina na detecção de fraudes depende significativamente da qualidade dos dados e do processo de engenharia de recursos. Neste contexto, este trabalho enfatiza a importância da preparação e pré-processamento dos dados, incluindo a normalização, tratamento de valores ausentes e desequilíbrio de classes, para aprimorar o desempenho do modelo de RL.

MATERIAL E MÉTODOS

A principal ferramenta utilizada foi o *Python* 3.11, uma linguagem de programação amplamente reconhecida no campo do aprendizado de máquina e da ciência de dados. A escolha do *Python* 3.11 se deu pelas diversas bibliotecas que oferece, como *scikit-learn*, *TensorFlow*, *Keras* e *PyTorch*. Dentre as bibliotecas mencionadas, *scikit-learn* 1.5.2 foi a mais utilizada devido à sua eficiência na implementação de modelos de RL, enquanto as demais blibliotescas foram descartadas por não serem necessárias para esta abordagem específica (GÉRON, 2021)

Além disso, o *Python* possui bibliotecas para manipulação e análise de dados, como *pandas* 2.2.3 e *NumPy* 2.1.2, que foram utilizadas para a limpeza, transformação e exploração dos dados. Para a visualização dos dados, foram empregadas as bibliotecas *Matplotlib* 3.9.2 e *Seaborn* 0.13.2, que permitiram criar gráficos e visualizações essenciais para a análise exploratória dos dados (Chen, 2018).

Para a implementação do algoritmo, foi essencial selecionar uma base de dados apropriada. A base de dados escolhida foi a disponibilizada pelo *Machine Learning Group da Université Libre de Bruxelles*, composta por 284 mil transações realizadas por cartões de crédito em setembro de 2013 (Credit Card Fraud Detection, 2018).

O tratamento e balanceamento dos dados foram etapas essenciais para a eficácia do modelo. A análise exploratória inicial envolveu a visualização das distribuições das variáveis e a identificação de inconsistências ou anomalias. A normalização das variáveis garantiu que todas as características tivessem uma escala comparável. Para lidar com o desequilíbrio entre as classes, foi aplicada a técnica *SMOTE* (*Synthetic Minority Over-sampling Technique*), que gerou amostras sintéticas da classe minoritária, permitindo que o modelo fosse treinado de maneira mais equilibrada, melhorando a capacidade de generalização do modelo preditivo (Taha et al, 2021).

A técnica de RL, conforme descrita por Gupta (2023), é eficaz para resolver problemas de classificação binária, modelando a probabilidade de ocorrência de um evento. A técnica modela a probabilidade de um evento binário (ocorrência ou não ocorrência) e é capaz de lidar bem com dados desbalanceados. A função logística, ou sigmóide, mapeia qualquer valor real em um intervalo entre 0 e 1.

Para a implementação do modelo, foi adotada a fórmula clássica da RL, conforme apresentado por Morettin e Singer (2022). Essa formulação é utilizada para modelar a probabilidade de ocorrência de um evento binário (fraude ou não fraude), sendo expressa pela equação:

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta nXn)}} \quad (1)$$

Onde P(Y=1∣X) é a probabilidade do evento de interesse (por exemplo, uma transação ser fraudulenta), β0 é o intercepto, β1,β2,…,βn são os coeficientes das variáveis preditoras X1,X2,…,n, e e é a base do logaritmo natural (Morettin; Singer, 2022).

O conjunto de dados foi dividido de forma estratificada em treinamento (80%) e teste (20%), garantindo que a proporção de transações fraudulentas fosse mantida em ambos os conjuntos. Isso é essencial para garantir uma avaliação justa do modelo e um desempenho representativo na detecção de fraudes.

Na implementação do modelo de RL utilizando a classe *LogisticRegression* da biblioteca *Scikit-learn 1.5.2*, os principais hiperparâmetros ajustados foram *penalty*, responsável por definir o tipo de regularização aplicada, foi configurado com penalização L2 (Ridge), a fim de evitar o sobreajuste. O hiperparâmetro *C*, que controla a força da regularização, foi mantido em seu valor padrão de 1.0, equilibrando a penalização. O *solver* escolhido foi o 'lbfgs', um algoritmo para otimização em problemas de classificação multiclasse. Por fim, o número máximo de iterações para a convergência do modelo foi definido por meio do hiperparâmetro *max_iter*, também mantido no valor padrão de 100, sendo suficiente para a convergência dos dados utilizados no estudo.

Durante o treinamento, o conjunto de dados foi utilizado para ajustar os coeficientes do modelo, minimizando a função de custo por meio do método de otimização de gradiente descendente, com o objetivo de maximizar a verossimilhança das observações.

Para avaliar o desempenho do modelo, aplicou-se a validação cruzada *k-fold* com k=5. Essa técnica divide o conjunto de dados em cinco subconjuntos, treinando o modelo em quatro e validando no quinto, repetindo o processo para todas as combinações possíveis. Isso ajuda a garantir que o modelo se generalize bem para dados não vistos e não esteja super ajustado aos dados de treinamento.

As métricas de avaliação utilizadas neste estudo incluem acurácia, precisão, *recall* e *F1-Score*, cada uma com um papel importante na análise do desempenho do modelo de detecção de fraudes. A acurácia mede a proporção de previsões corretas, mas pode ser enganosa em conjuntos de dados desbalanceados. Já a precisão avalia a proporção de transações corretamente classificadas como fraudulentas, enquanto o *recall* verifica a capacidade do modelo de identificar todas as transações fraudulentas reais, minimizando os falsos negativos (Alarfaj et al, 2022).

O F1-Score é a média harmônica entre precisão e *recall*, sendo especialmente útil em cenários de classes desbalanceadas, pois equilibra a importância de ambos, evitando que um alto valor de precisão mas com baixo *recall*, ou vice-versa, distorça a avaliação. Essas métricas foram fundamentais para avaliar o desempenho do modelo, garantindo que ele fosse capaz de detectar fraudes de maneira eficaz, ao mesmo tempo em que minimizava os erros de classificação (Alarfaj et al, 2022).

Após o treinamento e ajuste de hiperparâmetros, o modelo final foi avaliado no conjunto de testes. A acurácia foi monitorada continuamente, e ajustes adicionais foram realizados conforme necessário. O monitoramento envolveu a análise das métricas de avaliação mencionadas, além da inspeção de exemplos específicos de erros para identificar possíveis melhorias no modelo.

RESULTADOS E DISCUSSÃO

Outras observações relevantes indicaram que as transações fraudulentas tendem a ocorrer em horários específicos e com maiores variações no montante das transações. A variável "*Time*" mostrou-se importante para a análise de padrões temporais, já que algumas fraudes ocorreram com maior frequência em horários fora do padrão habitual de transações. A variável "*Amount*", por sua vez, revelou uma variação significativa entre transações legítimas e fraudulentas, indicando que transações fraudulentas tendem a envolver valores mais altos em comparação com transações legítimas. Essas informações foram fundamentais para o pré-processamento e definição de estratégias de balanceamento e normalização, a fim de melhorar a eficácia do modelo preditivo.

Os resultados obtidos após a aplicação do modelo são visualizados por meio de uma curva de aprendizado, que oferece uma representação gráfica detalhada da acurácia do modelo de RL em duas frentes: no conjunto de treinamento e durante o processo de validação cruzada. A curva de aprendizado é uma ferramenta crucial, pois permite observar a evolução da acurácia do modelo conforme o tamanho do conjunto de treinamento aumenta, destacando como o modelo se adapta e aprende a partir dos dados fornecidos.

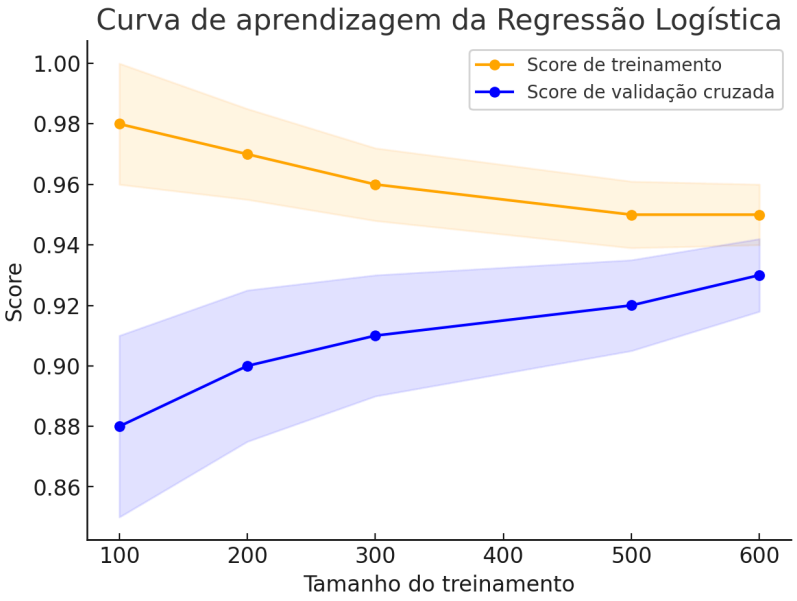

FIGURA 1. Gráfico da curva de aprendizado da Regressão Logística comparada com a Validação Cruzada.

A Figura 1 ilustra a curva de aprendizado da RL, que permite visualizar a acurácia do modelo em função do tamanho do conjunto de dados de treinamento. O eixo X representa o tamanho do treinamento, ou seja, a quantidade de dados utilizada para treinar o modelo em cada etapa. Conforme movemos para a direita, o tamanho dos dados de treinamento aumenta. Já no eixo Y, o score indica a acurácia do modelo, que mede a proporção de previsões corretas feitas tanto para o conjunto de treinamento quanto para o de validação cruzada.

A curva laranja refere-se ao Score de Treinamento, que representa a acurácia do modelo ao ser avaliado nos dados de treinamento. Inicialmente, o modelo apresenta uma acurácia alta, próximo de 100%. No entanto, à medida que o tamanho do conjunto de dados de treinamento aumenta, o score diminui levemente e se estabiliza em torno de 96%. Essa diminuição é normal e reflete que o modelo, ao lidar com mais dados, começa a aprender padrões mais generalizáveis, evitando o sobreajuste (quando o modelo se ajusta demais aos dados de treinamento e perde sua capacidade de generalização).

A curva azul mostra o Score de Validação Cruzada, que avalia o modelo em dados não utilizados diretamente no treinamento. Ela mostra uma tendência de crescimento conforme o tamanho do conjunto de dados de treinamento aumenta, indicando que o modelo melhora sua acurácia em dados não vistos à medida que recebe mais informações para aprender. A curva estabiliza-se em torno de 94%, mostrando que o modelo tem boa capacidade de generalização.

Os resultados obtidos com a aplicação da técnica de RL, após o pré-processamento, mostram métricas de avaliação que indicam um desempenho consistente. O modelo apresentou uma Precisão de 92,50%, que representa a proporção de transações classificadas como fraudulentas de forma correta em relação ao total de previsões positivas. Esta métrica é relevante para avaliar como o modelo trata transações legítimas, minimizando falsos positivos. A métrica de *recall*, por sua vez, alcançou 82,40%, o que indica a capacidade do modelo em identificar corretamente as transações fraudulentas existentes no conjunto de dados. O *recall* é uma métrica importante em casos onde a identificação completa de fraudes é priorizada, mesmo que alguns falsos positivos ocorram. Por fim, o *F1-Score* de 87,15% demonstra um equilíbrio entre a Precisão e o *recall*, sendo particularmente útil em cenários onde existe um desbalanceamento entre fraudes e transações legítimas.

Esses resultados foram obtidos a partir de um modelo ajustado com técnicas de normalização e balanceamento de classes. A normalização das variáveis garantiu que os atributos possuíssem a mesma escala, evitando que características com valores numéricos maiores influenciassem de maneira excessiva o modelo. O balanceamento das classes, por sua vez, buscou mitigar o impacto do desbalanceamento natural dos dados de fraude, onde o número de transações legítimas é consideravelmente superior ao de transações fraudulentas. O uso dessas técnicas impactou diretamente

na melhoria da capacidade de generalização do modelo, evidenciado pelos resultados das métricas de avaliação.

Em comparação com a acurácia do modelo treinado sem a aplicação de técnicas de pré-processamento, houve uma diferença significativa nas métricas de precisão e *recall* ao comparar o modelo com e sem pré-processamento. O modelo treinado sem pré-processamento apresentou uma precisão de 97,00%, o que indica que ele acertou mais transações legítimas ao custo de aumentar o número de falsos negativos, resultando em um *recall* de apenas 65,50%. Esse resultado aponta que, embora o modelo seja mais conservador ao classificar fraudes, ele falha ao capturar uma quantidade significativa de transações fraudulentas. Como consequência, o *F1-Score* caiu para 78,00%, evidenciando um desequilíbrio entre as previsões de fraudes corretas e as transações que foram erroneamente classificadas.

Os resultados obtidos neste estudo, como uma precisão de 92,50% e um *recall* de 82,40%, demonstram um desempenho competitivo em relação a outros estudos, como o de Alarfaj et al. (2022), que obteve um *F1-Score* superior utilizando redes neurais convolucionais (CNN). As redes neurais profundas, como as utilizadas por Alarfaj et al., tendem a ter um desempenho superior em conjuntos de dados grandes e altamente desbalanceados, devido à sua capacidade de aprender padrões complexos e não lineares. No entanto, o modelo de RL utilizado neste estudo apresentou resultados robustos, mesmo com um conjunto de dados menor e técnicas de pré-processamento, como o balanceamento de classes, que melhoraram sua capacidade de generalização. Isso sugere que, apesar de os modelos de aprendizado profundo oferecerem melhor desempenho em termos de *F1-Score*, a RL continua a ser uma opção eficaz, especialmente quando a interpretabilidade e a simplicidade computacional são prioritárias.

Em comparação com o estudo de Gupta (2023), que utilizou autoencoders para detectar fraudes em transações de cartões de crédito, os resultados deste estudo mostraram uma precisão de 92,50% com RL, enquanto Gupta obteve uma precisão de 99% com sua abordagem baseada em autoencoders. Embora a precisão dos autoencoders seja superior, é importante notar que este tipo de modelo é mais complexo e pode exigir maior capacidade computacional, além de ser menos interpretável. Em contraste, a RL, embora mais simples, oferece maior transparência no processo de decisão, o que pode ser uma vantagem significativa em contextos onde a aplicabilidade do modelo é crucial. Além disso, os autoencoders são particularmente eficazes em cenários com conjuntos de dados altamente desbalanceados, mas os resultados deste trabalho, com o uso de técnicas de pré-processamento como *SMOTE* e normalização, indicam que modelos mais simples podem alcançar resultados competitivos em termos de precisão, ao mesmo tempo em que mantêm vantagens de simplicidade e interpretabilidade.

CONCLUSÕES

A detecção de fraudes em transações com cartões de crédito é um problema crítico com grandes implicações financeiras. Este estudo utilizou RL combinada com técnicas de normalização e balanceamento de classes para lidar com o desbalanceamento entre transações legítimas e fraudulentas. A escolha dessa técnica foi motivada por sua eficiência e interpretabilidade, características essenciais em cenários que requerem modelos transparentes e explicáveis. Embora algoritmos de aprendizado profundo dominem a área, este estudo mostrou que métodos tradicionais ainda são soluções eficazes quando aplicados corretamente.

Os resultados obtidos, com precisão de 92,50%, *recall* de 82,40% e *F1-Score* de 87,15%, demonstram um bom equilíbrio entre a detecção de fraudes e a minimização de falsos positivos. A comparação com o modelo sem pré-processamento, que obteve resultados significativamente inferiores, reforça a importância das técnicas adequadas de preparação de dados para melhorar a generalização e o desempenho do modelo.

Apesar dos bons resultados, o modelo pode precisar de ajustes em conjuntos de dados diferentes ou padrões de fraude variados. Técnicas de ensemble learning, como boosting e bagging, podem aprimorar o desempenho ao combinar diversos modelos, reduzindo sobreajustes e aumentando a robustez. Boosting ajusta os erros dos modelos anteriores, enquanto bagging melhora a estabilidade ao combinar previsões de diferentes subconjuntos de dados.

Além disso, a inclusão de variáveis contextuais, como dados geográficos e comportamentais, pode melhorar a capacidade do modelo de detectar fraudes mais sofisticadas. Esses dados permitem

identificar padrões atípicos e comportamentos anômalos, aprimorando a detecção em cenários de fraude mais complexos e dinâmicos.

CONTRIBUIÇÕES DOS AUTORES

Isadora Disposti Bueno dos Santos contribuiu para a validação do tema proposto, ao buscar verificar sua relevância para a sociedade. Aline Bertolazo dos Santos desempenhou um papel fundamental no desenvolvimento do algoritmo e na aplicação da técnica estudada, além de confirmar sua eficiência por meio de uma análise detalhada. Aline Bertolazo dos Santos e Isadora Disposti Bueno dos Santos contribuíram significativamente para a redação do documento. Karina Mitiko Toma atuou como orientadora e acompanhou Aline Bertolazo dos Santos e Isadora Disposti Bueno dos Santos em todas as etapas do desenvolvimento da iniciação científica. Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Agradecimentos são dirigidos ao Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Campus Birigui, por oferecer a oportunidade de desenvolver esta iniciação científica. Expressamos também nossa gratidão a todos os professores que, com prontidão, nos auxiliaram ao longo desta jornada.

REFERÊNCIAS

ALARFAJ, Fawaz Khaled et al. Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. IEEE Access, v. 10, p. 39700-39715, 2022. Disponível em: https://doi.org/10.1109/access.2022.3166891. Acesso em: 10 jun. 2024.

CHEN, Daniel. Análise de Dados com Python e Pandas. Novatec Editora Ltda, São Paulo, 2018. E-book p. 366. ISBN 9788575226995.

CLEARSALE. Mapa da Fraude 2023. ClearSale, 2022. Disponível em: https://www.clearsale.com.br. Acesso em: 28 jun. 2024.

KAGGLE inc. Credit Card Fraud Detection. [S. l.], 22 mar. 2018. Disponível em: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. Acesso em: 03 jun. 2024.

GÉRON, Aurélien. Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow: Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes. [S. l.]: Alta Books, 2021. p. 577. ISBN 978-8550815480.

GUPTA, Jai. Credit Card Fraud Detection Using Machine Learning Algorithms. International Journal of Science and Research (IJSR), v. 12, n. 11, p. 1774-1779, 2023. Disponível em: https://doi.org/10.21275/sr231123121203. Acesso em: 1 jun. 2024.

MORETTIN, Pedro Alberto; SINGER, Julio da Motta. Estatística e Ciência de Dados. Rio de Janeiro: LTC | Livros Técnicos e Científicos Editora Ltda, 2022. ISBN 978-85-216-3816-2. Acesso em: 21 jun. 2024.

TAHA, Adil et al. Multilabel Over-Sampling And Under-Sampling With Class Alignment For Imbalanced Multilabel Text Classification. Journal Of Information And Communication Technology, [S. l.], v. 20, n. 3, p. 423–456, 2021. DOI: 10.32890/jict2021.20.3.6.