

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleberon dos S. Silva; Mariângela F. F. Molina.

COMPARAÇÃO DE TRÊS ALGORITMOS DE MACHINE LEARNING: ADABOOST, ÁRVORE DE DECISÃO E FLORESTA ALEATÓRIA PARA DETECÇÃO DE FRAUDES DE CARTÃO DE CRÉDITO E SEU IMPACTO EM CLIENTES E INSTITUIÇÕES FINANCEIRAS.

KAUE TSUYOSHI LIMA ONO¹
KLEBERSON DOS SANTOS SILVA²
MARIÂNGELA FERREIRA FUENTES MOLINA³

RESUMO

Este artigo tem como objetivo comparar três algoritmos utilizados na detecção de fraudes em cartões de crédito, analisando o impacto de suas previsões nas instituições financeiras e nos clientes por meio da avaliação das transações. Em um contexto em que o uso de cartões de crédito em compras online se torna cada vez mais comum, a relevância dessa análise torna-se evidente. A análise comparativa engloba três algoritmos: Árvore de Decisão, Floresta Aleatória (ambos utilizando o Índice de Gini) e o algoritmo AdaBoost. Observa-se que, apesar da simplicidade e rapidez de implementação da Árvore de Decisão, ela apresenta resultados razoáveis, porém com uma incidência significativa de erros no modelo de teste. Por outro lado, a Floresta Aleatória consegue mitigar esse problema, obtendo resultados melhores, embora consuma mais recursos computacionais. Já o algoritmo AdaBoost alcança um cenário mais favorável em termos de frequência com que um modelo de aprendizado de máquina identifica corretamente instâncias positivas.

Palavras-chave: Árvore de decisão, Floresta Aleatória, AdaBoost, Índice de Gini.

ABSTRACT

The purpose of this article is to compare three algorithms used in credit card fraud detection, exploring how their predictions impact financial institutions and customers through transaction analysis. In a scenario where the use of credit cards in online purchases is increasingly common, the importance of this analysis is evident. The comparative analysis covers three algorithms: Decision Tree, Random Forest (both using the Gini Index), and the AdaBoost algorithm. It is observed that despite the simplicity and quick implementation of the Decision Tree algorithm, it yields reasonable results but with a significant incidence of errors in the test model. On the other hand, the Random Forest manages to mitigate this problem, achieving better results albeit consuming more computational resources. Meanwhile, the

¹Graduando em Tecnologia em Análise e Desenvolvimento de Sistemas pela Faculdade de Tecnologia de Mogi das Cruzes - FATEC-MC. E-mail: kaue.ono@fatec.sp.gov.br

²Graduando em Tecnologia em Análise e Desenvolvimento de Sistemas pela Faculdade de Tecnologia de Mogi das Cruzes - FATEC-MC.

³Docente, Faculdade de Tecnologia de Mogi das Cruzes – FATEC-MC.

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleber dos S. Silva; Mariângela F. F. Molina.

AdaBoost algorithm achieves a more favorable scenario in terms of the frequency with which a machine learning model correctly identifies positive instances.

Key words: Random Forest, Ada Boost, Decision tree, Gini Index.

INTRODUÇÃO

Para Khine e Khin (2020) a detecção de fraudes em cartões de crédito atualmente apresenta um desafio considerável devido à imensa quantidade de transações cotidianas. Essas fraudes ocorrem quando criminosos buscam acesso aos recursos de crédito alheios, adquirindo bens sem efetuar os devidos pagamentos. Diante desse cenário, o emprego de técnicas avançadas de aprendizado de máquina ganha destaque, como a utilização de Árvores de Decisão com algoritmo CART e índice de Gini, Florestas Aleatórias com índice de Gini e AdaBoost.

O uso do aprendizado de máquina possui o potencial de identificar padrões e comportamentos suspeitos que, de outra forma, seriam praticamente imperceptíveis pelos métodos tradicionais. Isso permite uma análise mais eficiente de uma pequena parcela das transações pelas operadoras de crédito, aprimorando a experiência do usuário e reduzindo os riscos financeiros para as empresas de crédito.

Conforme destacado por Yazici (2020), atualmente, diante do expressivo volume de dados gerados diariamente, alguns indivíduos ainda buscam abordar esse desafio através de algoritmos matemáticos fundamentados em regras. Contudo, observa-se uma crescente demanda por técnicas de aprendizado de máquina e inteligência artificial. Essa tendência é impulsionada, em grande medida, pela necessidade de analisar vastas bases de dados originadas de bilhões de transações, cujas informações têm o potencial de serem valiosas na tentativa de antecipar se uma transação desconhecida é, de fato, uma fraude.

Este artigo tem como propósito investigar o papel fundamental desempenhado pela aprendizagem de máquina na detecção de fraudes em cartões de crédito. Vamos abordar as principais técnicas, desafios e benefícios associados à

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
--

implementação dessas abordagens avançadas de análise de dados em um ambiente de pagamento eletrônico. Além disso, discutiremos casos de estudo e exemplos reais que destacam como as soluções de aprendizagem de máquina têm sido bem-sucedidas na proteção de transações financeiras e na redução do impacto das fraudes. Ao mesmo tempo, exploraremos os prós e contras de cada algoritmo, analisando como suas métricas influenciam diversos cenários de negócio e impactam nos resultados desejados pelas empresas, considerando sempre a relação com o cliente.

MATERIAIS E MÉTODOS

Com o intuito de alcançar o objetivo de determinar qual entre essas implementações oferece a melhor relação entre precisão, eficiência e tempo de execução foram utilizados métodos de pesquisa quantitativa e exploratória baseada em experimentos de execução de algoritmos de IA e comparação de gasto de recursos computacionais. E de procedimentos bibliográficos baseados em artigos científicos publicados no portal Google Acadêmico, nos seguintes portais: IEEE, Springer, Arxiv e Mdpi.

A pesquisa se delimita na análise comparativa de três algoritmos de aprendizado de máquina utilizando índice de Gini nos algoritmos: Árvore de Decisão (CART) e Random Forest. Além disso, o algoritmo AdaBoost também foi implementado.

REFERENCIAL TEÓRICO

Uso de IA na Detecção de Fraude de Cartão De Crédito

Para Khine e Khin (2020) Uso de IA na prevenção e identificação de fraudes usam as seguintes técnicas avançadas de mineração de dados, enquanto outros recorrem à mineração de fluxos de dados para identificar atividades fraudulentas em

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleber dos S. Silva; Mariângela F. F. Molina.

transações de cartão de crédito. Fraudes com cartões de crédito podem causar prejuízos bilionários para clientes e instituições bancárias. Diante disso, é essencial que a sociedade disponha de um sistema robusto e eficiente de detecção de fraudes, capaz não só de identificar, mas também de prevenir rapidamente esses atos ilícitos. Assim, os métodos de detecção de fraudes em transações com cartões de crédito demandam constante inovação e vigilância.

Pesquisadores têm realizado comparações detalhadas entre diversas técnicas de mineração de dados e revisões abrangentes da literatura sobre sistemas de detecção de fraudes. Técnicas variadas, como inteligência artificial, aprendizado de máquina, lógica difusa, alinhamento de sequências, redes neurais, regressão logística, bayesiana ingênua, programação genética e árvores de decisão, foram desenvolvidas para identificar diferentes formas de fraude, tanto em aplicações quanto em comportamentos.

Entretanto, a evolução dos métodos de detecção enfrenta desafios significativos, principalmente devido à restrição na troca de informações. Primeiramente, essa limitação é imposta por preocupações com segurança e privacidade. Em segundo lugar, a indisponibilidade de conjuntos de dados e a censura frequente dos resultados dificultam a avaliação dos métodos. Por isso, alguns estudos utilizam dados sintéticos, como os de Brause et al. (1999) e Dorronsoro et al. (1997).

Conforme Yazici (2020), com o grande volume de dados disponível atualmente, há uma crescente demanda por técnicas de aprendizado de máquina e inteligência artificial para abordar esse desafio. A análise de bilhões de transações pode ser potencialmente útil para antecipar se uma transação desconhecida é fraudulenta. A fraude com cartões de crédito é geralmente classificada em dois tipos: fraude de aplicação e fraude comportamental.

No contexto em que (g_t) representa um classificador fraco (um modelo de classificação simples) que produz uma previsão com base em um vetor de entrada (x) , sendo (t) o índice da iteração. Para cada amostra de treinamento, a previsão de

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
--

um classificador fraco é representada por $h(x_n)$. Além disso, embora a análise de fraudes seja complexa, trata-se essencialmente de um desafio de classificação binária, onde o resultado é verdadeiro ou falso. A estratégia de aprendizado, segundo Yazici (2020), utiliza transações passadas, já categorizadas como fraudulentas ou legítimas, para treinar o modelo. Esse modelo, então, é aplicado para prever se uma nova transação será ou não uma fraude.

AdaBoost

O AdaBoost é um algoritmo de boosting criado por Yoav Freund e Robert Schapire em 1996 que combina vários classificadores fracos para formar um classificador forte. Cada classificador fraco é treinado de forma sequencial, onde os erros cometidos por classificadores anteriores são corrigidos por classificadores subsequentes, dando maior peso às amostras mal classificadas. AdaBoost é eficaz em aumentar a precisão dos modelos, especialmente em problemas de classificação binária, mas pode ser sensível a ruídos nos dados. Segundo Zhou, et al (2020) O Adaboost é um algoritmo iterativo que se fundamenta na ideia de treinar vários classificadores distintos, conhecidos como classificadores fracos, utilizando o mesmo conjunto de treinamento. Posteriormente, esses classificadores fracos são combinados para criar um classificador final mais robusto, denominado classificador forte.

Deste modo, de acordo com Zhou, et al (2020) a saída do algoritmo AdaBoost consiste em uma soma ponderada, alcançada através da combinação das saídas dos modelos impulsionados individualmente. A formulação matemática do método AdaBoost é apresentada abaixo:

Cada iteração (t), um classificador fraco é selecionado e multiplicado por um coeficiente $\beta(t)$ para calcular o erro de treinamento (ϵ_t) da seguinte forma:

Para Illeberi, Sun e Wang (2021), a fórmula (G_{t-1}) representa um classificador aprimorado na iteração $(t-1)$, enquanto $h_t(x_n)$ denota um classificador fraco considerado para o modelo final.

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
--

Árvore de Decisão

A Árvore de Decisão é um algoritmo simples e intuitivo que foi aprimorado ao longo dos anos por pesquisadores no campo da estatística e da inteligência artificial utilizado para classificação e regressão. Ela divide os dados em subconjuntos com base em testes de valor de atributo, criando uma estrutura semelhante a uma árvore. Cada nó interno representa uma "divisão" baseada em um atributo, e cada nó folha representa uma saída de classificação. O algoritmo é rápido de implementar e interpretar, porém pode ser propenso a sobreajuste se não for corretamente podado. Utilizamos o índice de Gini para calcular a impureza dos nós e selecionar as melhores divisões, a Floresta Aleatória faz uso de árvores de decisão para a classificação de dados, empregando a técnica de ensacamento (É uma meta-algoritmo de conjunto de aprendizagem de máquina projetado para melhorar a escalabilidade e a precisão dos algoritmos de aprendizagem de máquina usados na classificação e regressão estatística), que consiste na agregação. Essa abordagem visa mitigar o problema de sobreajuste (ocorre quando o modelo não consegue generalizar e se ajustar muito ao conjunto de dados de treinamento) associado a árvores de decisão complexas. Abaixo, apresenta-se a explicação inicial do conceito de utilização de árvores de decisão para classificar dados.

Uma árvore de decisão é uma estrutura assemelhada a uma árvore, na qual cada nó interno realiza uma "divisão" com base em um atributo, enquanto cada nó folha representa um resultado de previsão ou classificação. Diversas métricas, como impureza de Gini, entropia e desvio padrão, podem ser empregadas para escolher a melhor divisão, visando o maior ganho de informação. No exemplo a seguir, a impureza de Gini é utilizada para ilustrar como o ganho de informação é calculado em árvores de decisão.

O índice de gini pode ser calculado utilizando a seguinte equação, que pode ser visto na Figura 1:

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
--

Figura 1. Equação de Gine.

$$G = \sum_{k=1}^K p_k (1 - p_k)$$

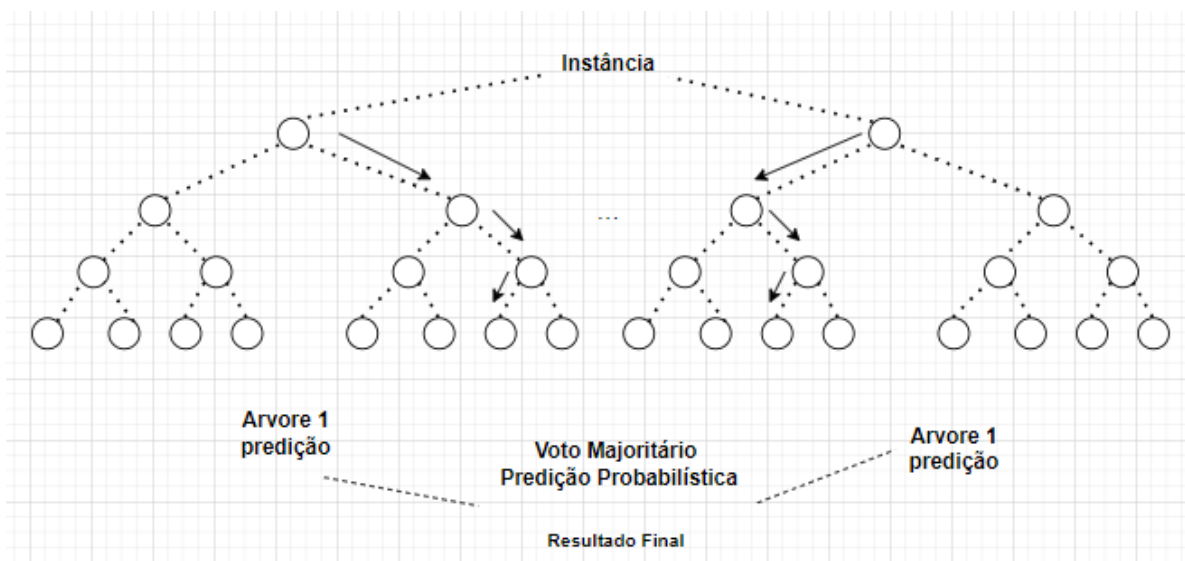
Fonte: Os autores (2023).

A seguir, descrevemos a abordagem de ensacamento, que consiste na seleção aleatória de subconjuntos de dados e atributos para a construção de diversas árvores de decisão. Essas árvores são combinadas para a classificação de dados, visando evitar o problema de sobreajuste inerente às árvores de decisão. Para um conjunto de dados com (d) observações, a abordagem de ensacamento gera (n) subconjuntos de dados ao extrair (d) observações (onde $d \leq d$) com reposição. Cada subconjunto de dados, juntamente com um subconjunto de atributos selecionados aleatoriamente, é utilizado no treinamento de uma árvore de decisão. Assim, são treinadas (n) árvores independentes, cada uma com diferentes subconjuntos de dados e atributos. Finalmente, a votação majoritária ou a média é aplicada às (n), árvores de decisão para obter o resultado final do processo de classificação, como ilustrado na Figura 2.

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono;
Kleberson dos S. Silva;
Mariângela F. F. Molina.

Figura 2. Árvore de decisão/



Fonte: <https://www.mdpi.com/2227-7390/9/21/2683>

Imagem adaptada

Em linhas gerais, um modelo de Floresta Aleatória (RF) é empregado para classificar uma instância de entrada em uma das (r) classes, denotadas como (c_1, \dots, c_r). O procedimento utilizado para construir o modelo RF, composto por (n) árvores de decisão, para um conjunto de dados com (d) observações e (k) atributos, envolve três etapas principais segundo (LIN; JIANG, 2021).

Passo 1. Gere (n) subconjuntos de dados a partir do conjunto original de (d) observações. Cada subconjunto é obtido por amostragem com reposição, selecionando (d) observações, onde ($d \leq d$).

Passo 2. Para cada um dos, (n) subconjuntos de dados, desenvolva uma árvore de decisão, selecionando as melhores divisões nos nós internos da árvore com o maior ganho de informação entre (k) atributos arbitrários, onde ($k < k$). Assim são geradas (n) árvores de decisão no total resultando em (n) classificações, cada uma pertencente a uma das (r) classes ou rótulos (c_1, \dots, c_r).

Passo 3. Agregue os resultados das (n), árvores para determinar a classe predominante como classificação final, representada por $\text{cout} = \arg \max r_i = 1 \text{ freq}(c_i)$, onde $\text{freq}(c_i)$ é a frequência com que (c_i) aparece entre as (n) classificações. É

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.	Kaue T. L. Ono; Kleber dos S. Silva; Mariângela F. F. Molina.
---	---

importante notar que a saída pode ser ajustada para fornecer classificações probabilísticas, gerando as frequências (ou probabilidades) $\text{freq}(c_1), \dots, \text{freq}(c_r)$ para todas as classes (c_1).

Floresta Aleatória

A Floresta Aleatória é um algoritmo desenvolvido por Leo Breiman e Adele Cutler em meados dos anos 2000 que combina múltiplas Árvores de Decisão para melhorar a precisão e reduzir o risco de sobreajuste. Ele utiliza a técnica de ensacamento onde várias árvores são treinadas em diferentes subconjuntos de dados extraídos com reposição. As previsões são agregadas, normalmente por meio de votação majoritária, para obter a classificação final. Este método melhora a capacidade de generalização e estabilidade do modelo, embora exija mais recursos computacionais. De acordo com Xuan, et al (2018) deve-se atentar quanto mais robusta for cada árvore individual e quanto menor for a correlação entre as árvores distintas, melhor será o desempenho da floresta aleatória. A diversidade das árvores resulta de sua aleatoriedade, que incorpora amostras de inicialização (Técnica de reamostragem que envolve extrair repetidamente amostras de dados de origem com repetição geralmente para estimar um parâmetro populacional) e a seleção aleatória de um subconjunto de atributos dos dados

Desta forma diante dos dados desequilibrados, é crucial estabelecer de forma apropriada o custo associado à classificação incorreta, ao mesmo tempo em que se busca implementar soluções fundamentadas em abordagens de aprendizado sensíveis ao custo. No contexto tradicional da detecção de fraudes em transações de cartão, um extenso pré-processamento de dados é necessário, envolvendo etapas como seleção, extração de recursos, amostragem e detecção de valores discrepantes. Assim sendo para DEVI; BISWAS; PURKAYASTHA, (2019) essas fases precedem a aplicação de algoritmos de aprendizado de máquina voltados para a identificação de casos legítimos e fraudulentos.

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleber dos S. Silva; Mariângela F. F. Molina.

DESENVOLVIMENTO DA PESQUISA

A seguir apresenta-se o teste de detecção de fraude de cartão de crédito que o grupo desta pesquisa realizou para implementar os seguintes algoritmos (Árvore de decisão, Floresta Aleatória e AdaBoost). Nele contempla as seguintes informações: O modelo da máquina que foi utilizado, o procedimento e a base de dados que foi aplicado para o treinamento e geração dos resultados. A configuração do computador pode ser obtida na tabela 1.

Tabela 1. Configuração do computador.

Processador	Intel core i5 3230M
Memória	Type - DDR3, Size - 8GBytes, Chanel – Dual
Sistema Operacional	Windows 10 Version 22H2 para x64 - Sistemas baseados (KB5031445)

Fonte: Os autores (2023).

Dados Relacionados à Pesquisa

Na implementação dos algoritmos (Árvore de Decisão, Floresta Aleatória e AdaBoost), utilizamos um conjunto de dados composto por 284.807 transações bancárias coletadas ao longo de dois dias de operação de uma empresa europeia de cartão de crédito. Deste modo foi possível realizar o teste desta natureza da pesquisa.

Devido ao desbalanceamento das classes, empregamos a biblioteca StratifiedShuffleSplit (para garantir que a classe de fraude contenha 10% das transações fraudulentas para seu treinamento).

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.	Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
---	--

Composição do Dataset

A pesquisa utilizou um conjunto de dados composto por 284.807 transações registradas, das quais 10% foram alocadas para o treinamento dos algoritmos. Desse total, 492 transações foram identificadas como fraudulentas, representando 0,17%, enquanto 284.315 transações foram definidas como legítimas, totalizando 99,83%. Essas informações estão apresentadas na Tabela 2.

Tabela 2: Composição do Dataset.

Total Transações (100 %)	Transações Fraudes (0.17%)	Transações Legais (99.83%)
284.807	492	284.315

Fonte: Os autores (2023).

RESULTADOS E DISCUSSÃO

Métricas do modelo implementado e resultados.

Na implementação dos algoritmos, empregou-se uma semente de aleatoriedade fixa em 0, assegurando que os resultados permaneçam consistentes entre diferentes implementações e evitando a variabilidade introduzida pela aleatoriedade. Quanto à profundidade da árvore, foi escolhido um limite máximo de 10 para os modelos Árvore de Decisão e AdaBoost. Essa escolha visa alcançar um

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.	Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
---	--

equilíbrio adequado, mitigando o risco de sobreajuste que poderia comprometer as métricas de acurácia, precisão e recall. Segue abaixo na tabela 3 os resultados.

Tabela 3. Configuração dos algoritmos e tempo de execução.

	Semente Aleatoriedad e	Profundidade Máxima Árvore	Árvores Geradas	Índice de Gine	Tempo de Execução (em segundos)
Árvore de decisão	0	10	1	Sim	5:29
Floresta Aleatória	0	Configuração padrão do DecisionTreeClassifier	50	Sim	15: 55
AdaBoost	0	10	200	Não	23: 05

Fonte: Os autores (2023).

RESULTADOS

Os resultados obtidos com os três algoritmos envolvem três métricas. Acurácia que é uma métrica amplamente utilizada para medir o desempenho em aprendizagem de máquina. Ela representa a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões).

Precisão refere - se ao número de exemplos que o modelo previu como sendo de uma classe e que realmente são divididos pelo número total de exemplos previstos daquela classe.

Recall é uma métrica que mede a frequência com que um modelo de aprendizado de máquina identifica corretamente instâncias positivas verdadeiros

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono;
Kleber dos S. Silva;
Mariângela F. F. Molina.

positivos de todas as amostras positivas reais no conjunto de dados. Os resultados podem ser vistos na Figura 3.

Figura 3. Resultados obtidos após codificação e execução dos algoritmos



Fonte: Os autores (2023).

CONCLUSÃO

A análise das métricas dos modelos de Árvore de Decisão, Floresta Aleatória e AdaBoost revela um desempenho notável na detecção de fraudes de cartão de crédito, todos com uma acurácia acima de 99%.

No entanto, ao examinar métricas mais detalhadas, nuances importantes emergem. A Árvore de Decisão demonstra sólida acurácia, mas seu recall de 0,71 pode resultar em falsos negativos. A Floresta Aleatória combina alta acurácia com um recall de 0,75, destacando sua capacidade superior na detecção de fraudes. Além disso, é vital considerar a escala dos dados e a diversidade das classes a serem analisadas. À medida que o volume de dados cresce e a complexidade das classes aumenta, o modelo AdaBoost se torna a opção preferencial, embora exija mais tempo de processamento. O AdaBoost oferece uma acurácia elevada e um recall de 0,79, tornando-se a escolha ideal para uma identificação superior de fraudes, enquanto a Floresta Aleatória é a melhor opção quando a prioridade do negócio.

Comparação de três algoritmos de Machine learning: Adaboost, Árvore de decisão e Floresta aleatória para detecção de fraudes de cartão de crédito e seu impacto em clientes e instituições financeiras.

Kaue T. L. Ono; Kleberson dos S. Silva; Mariângela F. F. Molina.
--

REFERÊNCIAS BIBLIOGRÁFICAS

DEVI, D.; BISWAS, S. K.; PURKAYASTHA, B. A Cost-sensitive weighted Random Forest Technique for Credit Card Fraud Detection. **IEEE Xplore**, Silchar, p. 01-06, 2019. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8944885>. Acesso em: 02 nov. 2023.

ILEBERI, E.; SUN, Y.; WANG, Z. Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost. **IEEE Explore**, Johannesburg, v. 9, p. 165286-165294, 2021. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9651991>. Acesso em: 15 out. 2023.

KHINE, A. A.; KHIN, H. W. Credit Card Fraud Detection Using Online Boosting with Extremely Fast Decision Tree. **IEEE Xplore**, Houston, p. 1-4, 2020. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9022843/keywords#keywords>. Acesso em: 04 nov. 2023.

LIN, T.; JIANG, J. Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest. **Mdpi**, Los Angeles, p. 03-16, 2021. Disponível em: <https://www.mdpi.com/2227-7390/9/21/2683>. Acesso em: 01 nov. 2023.

ONO, K. T. L.; SILVA, K. dos S.; MOLINA, M. F. F. **Código Fonte**. 2023. Disponível em: https://github.com/softwarekleberson/DeepLearning_FraudeCartao. Acesso em: 06 novembro. 2023.

XUAN, S.; LIU, G.; LI, Z.; ZHENG, L.; WANG, S.; JIANG, C. Random Forest for Credit Card Fraud Detection. **IEEE Xplore**, Shanghai, p. 01-6, 2018. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8361343>. Acesso em: 01 nov. 2023.

YAZICI, Y. APPROACHES TO FRAUD DETECTION ON CREDIT CARD TRANSACTIONS USING ARTIFICIAL INTELLIGENCE METHODS. **Arxiv**, Istanbul, p. 236-244, 2020.

ZHOU, H.; WEI, L.; CHEN, G.; LIN, Peng; LIN, Y. Credit Card Fraud Identification Based on Principal Component Analysis and Improved Adaboost Algorithm. **IEEE Xplore**, Auckland, p. 507-510, 2020. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9051207>. Acesso em: 05 out. 2023.