Aplicação de Modelos de Aprendizado de Máquina para Detecção de Fraudes em Transações de Cartões de Crédito

João Paulo P. Dantas[1], **Ana Régia de M. Neves**[2]

[1]Eixo de Informação e Comunicação – Instituto Federal de Brasília (IFB)
Brasília – DF – Brasil

joaod3v@gmail.com[1], ana.neves@ifb.edu.br[2]

Abstract. *Fraud in credit card transactions in Brazil represents an estimated annual loss of 7 billion dollars. Fraud detection systems must check each transaction accurately and precisely. Current systems based only on rules are no longer able to face this problem. The application of machine learning techniques can learn transaction patterns and detect when one is legitimate or fraudulent with high precision and accuracy, contributing to reduce the scenario of high losses with credit cards. Thus, the objective of this research is, based on the results of the systematic review, to identify which machine learning technique can have the best possible performance in detecting fraud in credit card transactions.*

Resumo. *A fraude em transações de cartões de crédito no Brasil representa perdas anuais estimadas em 7 bilhões de dólares, pois os atuais sistemas de detecção de fraudes, baseados apenas em regras, não apresentam precisão e acurácia suficientes para a identificação das mesmas. As técnicas de aprendizado de máquina aplicadas nos sistemas de detecção de fraudes permitem que as máquinas assimilem padrões das transações e detectem sua legitimidade com a precisão e acurácia e neste trabalho será explorado se essas técnicas são capazes de contribuir para redução dos elevados prejuízos enfrentados pelas instituições que administram os cartões de crédito no Brasil. Assim, o objetivo desta pesquisa é de, com base nos resultados de uma revisão sistemática, identificar qual melhor técnica de aprendizado de máquina poderá ter maior desempenho na detecção de fraudes nesse contexto.*

1. Introdução

A expansão da Internet possibilitou a migração de diversos modelos de negócios para o ambiente virtual transformando o cartão de crédito em um dos principais meios de pagamento em transações comerciais adotados por consumidores no mundo todo. Porém, a facilidade em sua utilização gerou uma janela de oportunidade para que criminosos explorassem as fragilidades desse meio de pagamento praticando diversos tipos de fraudes.

Fraudes em cartões de crédito ocorrem quando o cartão é utilizado por terceiros sem a autorização do titular. Segundo [Can et al. 2020], destacam-se cinco tipos de fraudes, que são: (i) roubo simples, ocorre quando o cartão físico é roubado; (ii) fraude na aquisição, ocorre quando um cartão é adquirido com base em documentos falsos; (iii) fraude de insolvência, ocorre quando o titular continua utilizando o cartão mesmo sem

condições de honrar o dispêndio realizado; (iv) fraude interna, ocorre quando empregados da instituição roubam informações do cartão e o utilizam indevidamente; e (v) fraude de comportamento, ocorre de maneira não presencial quando os fraudadores obtêm os dados do cartão por *phishing* ou aquisição de dados do titular e o utilizam para realizar compras sem autorização do mesmo.

De acordo com [Tingfei et al. 2020], a fraude global em cartões de crédito saltou de 9,84 bilhões de dólares em 2011 para 27,69 bilhões de dólares em 2017, um crescimento de mais de 180% no período. No Brasil, a estimativa de prejuízo das instituições é de 7 bilhões de dólares por ano [de C. B. Paul et al. 2020].

Identificar uma transação fraudulenta em meio à milhares de transações legítimas é um processo que demanda grandes esforços quando realizada manualmente. Segundo [Can et al. 2020], as soluções tradicionais de detecção de fraudes são baseadas em sistemas de regras, em que são estipulados parâmetros de monitoramento de variáveis pré-selecionadas e é emitido um alerta para análise humana quando estes parâmetros são atingidos.

[Bagga et al. 2020] e [Misra et al. 2020] afirmam que devido ao elevado volume das transações eletrônicas, a maneira tradicional de detecção de fraudes demanda muito tempo para análise, não é escalável e possui baixa acurácia. Segundo [Dornadula and Geetha 2019], mesmo com o incremento de novos artifícios tecnológicos, tais como os cartões com chip, os fraudadores adaptam-se e mudam o perfil da fraude.

Segundo [de C. B. Paul et al. 2020], novos métodos para validação das transações são necessários para garantir mais segurança, com isso, os métodos estatísticos e de aprendizado de máquina destacam-se no reconhecimento de fraudes de maneira mais precisa, pois processam de forma eficiente elevados volumes de dados. Tais modelos utilizam dados do usuário, da transação e dos metadados do equipamento como insumos para recomendar a aprovação ou não de uma transação. Desta forma, sistemas de detecção de fraude baseados em aprendizado de máquina contribuem para que estas transações possam ocorrer com o menor risco possível, evitando prejuízos para clientes e instituições.

Neste contexto, o objetivo deste trabalho é aplicar técnicas de aprendizado de máquina em uma base de transações de cartões de crédito para detectar transações fraudulentas e identificar quais das técnicas utilizadas se destacam em relação as demais na análise de um grande volume de dados.

As demais seções deste trabalho estão organizadas como segue: a Seção 2 apresenta a revisão sistemática da literatura dos conteúdos de interesse deste trabalho; a Seção 3 apresenta as principais características da base de dados escolhida e uma breve análise estatística; a Seção 4 apresenta o cronograma para execução da pesquisa.

2. Revisão Sistemática da Literatura

Este projeto é fundamentado de acordo com a revisão sistemática da literatura e direcionado pela seguinte questão: “Qual técnica de aprendizado de máquina poderá ter melhor desempenho na detecção de fraudes em transações de cartões de créditos?”. Com base nessa pesquisa, será possível avaliar a possibilidade de implementação de um sistema real de monitoramento de fraudes com base em aprendizado de máquina em uma instituição financeira.

As buscas foram baseadas no título e resumo dos trabalhos, e ocorreram no período de abril até maio de 2021. Os critérios de inclusão e exclusão são descritos na Tabela 1. As fontes onde as buscas foram realizadas e o conjunto de *strings* utilizadas são apresentados na Tabela 2.

Tabela 1. Critérios definidos para a revisão sistemática

Critérios de Inclusão	Critérios de Exclusão
Detecção de fraudes em transações de cartões de crédito com aplicação de técnicas de aprendizagem de máquina	Detecção de fraudes em transações de cartões de crédito sem aplicação de técnicas de aprendizagem de máquina
Identificação de fraudes em transações de cartões de crédito com aplicação de técnicas de aprendizagem de máquina	Identificação de fraudes em transações de cartões de crédito sem aplicação de técnicas de aprendizagem de máquina

Tabela 2. Fontes e *Strings* de busca utilizadas

Base de dados	Palavra-chave	Resultados
IEEE XPLORE	(Machine Learning) AND (Credit Card) AND (Fraud Detection) - Open Access	11
Science Direct	(Machine Learning) (Credit Card) (Fraud Detection) - Open Access	76
Google Scholar	(Aprendizado de Máquina) (Detecção de Fraude) (Cartão de Crédito)	27

2.1. Resultados da Revisão Sistemática da Literatura

Na organização da pesquisa sistemática foi utilizado o aplicativo *Rayyan* [1] e os resultados da revisão são apresentados no Diagrama Prisma, conforme Figura 1. No total foram recuperados cento e quatorze trabalhos das bases *IEEE Xplore*, *Science Direct* e *Google Scholar*. Destes, nove foram considerados para síntese final e três foram considerados como trabalhos correlatos, Seção 2.2, pois demonstraram todo o processo de aplicação das técnicas de aprendizado de máquina.

Dentre os trabalhos analisados cerca de 63% utilizaram as seguintes técnicas de aprendizado de máquina:

- *Support Vector Machine* (SVM), tem por objetivo encontrar o melhor hiperplano que separa de maneira otimizada dois conjuntos de pontos em classes distintas [Makki et al. 2019]. Segundo [Rtayli and Enneya 2020], nesta técnica os atributos de um conjunto de dados são transformados em vetores em um hiperplano, assim, cada vetor é representado como um ponto em planos de coordenadas de k dimensões. Desta forma, o SVM será capaz de executar uma função que busca minimizar a distância entre os pontos comuns a fim de maximizar a distância entre os planos que contêm cada tipo de classificação;

[1] Disponível em https://rayyan.ai

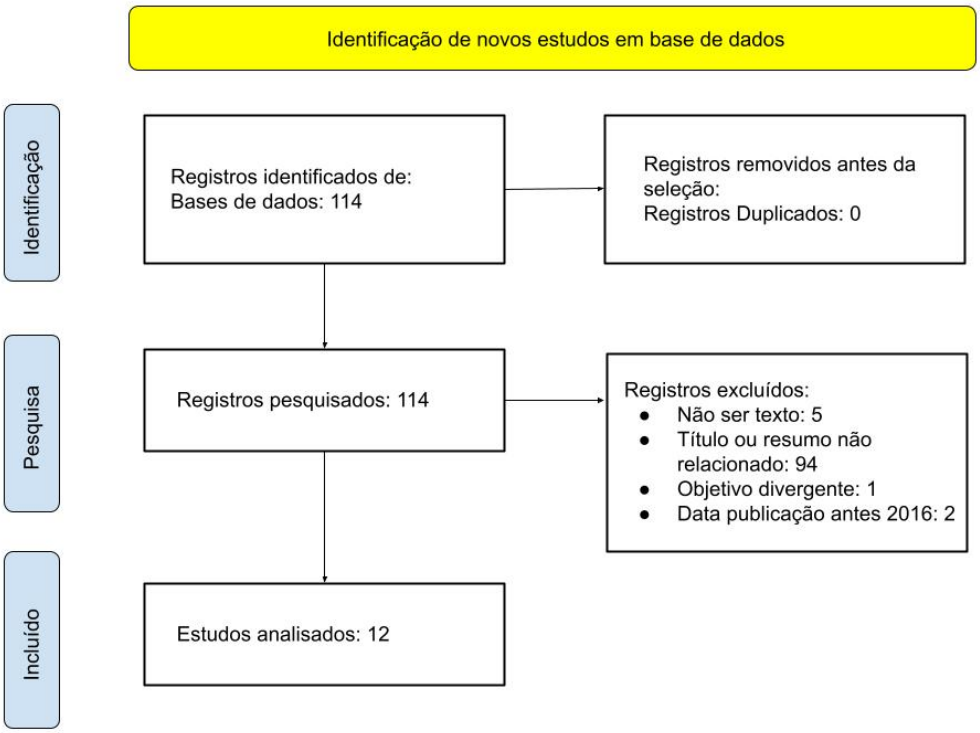

Figura 1. Diagrama Prisma da Revisão Sistemática da Literatura

- *Decision Tree* usa da analogia de uma árvore composta por galhos e folhas para demonstrar um conjunto de regras e seus resultados [Makki et al. 2019]. Nesta técnica, o galho representa um nó em que um atributo é testado por uma regra. O resultado da aplicação da regra é utilizado como insumo pelo novo nó e, assim, sucessivamente, até que esgote todas as possibilidades de regras gerando um resultado final ou, segundo a analogia, uma folha;
- *K-Nearest Neighbors* (KNN), segundo [Makki et al. 2019], cria uma função classificadora por meio de um sistema de votação baseado nos pontos mais próximos de um determinado centro. Suponha que o parâmetro k seja cinco para o número de vizinhos e a forma de calcular a distância entre eles seja a euclidiana. O classificador do KNN encontrará as cinco amostras mais próximas de uma transação. Dessa forma, será possível classificar se uma transação pertence a uma classe ou outra com base na menor distância possível entre seus vizinhos;
- *Logistic Regression* é um modelo de generalização linear, sendo o objetivo encontrar a probabilidade de uma transação ser de uma determinada classe. Cada atributo será multiplicado por um coeficiente e o resultado desses produtos serão somados a probabilidade de erro e a um coeficiente constate, assim será possível estimar o risco da fraude;
- *Multi-Layer Perceptron* é uma estrutura formada por três ou mais camadas compostas por neurônios totalmente interconectados [Can et al. 2020], os quais utilizam de uma função de ativação de dados para realizar testes gerando resultados que são reinseridos no sistema, formando ciclos de retro-propagação, até a realização do aprendizado pela máquina;

- *Naive Bayes*, nela cada atributo é condicionalmente independente dos demais [Makki et al. 2019], o que gera uma dificuldade em definir classes em dados desbalanceados, uma vez que, ela tendencia classes majoritárias;
- e, *Random Forest* cria um conjunto aleatório de árvores de decisão, a partir do número de árvores escolhido pelo usuário [Randhawa et al. 2018]. Posteriormente utiliza de um sistema de votação para determinar qual das árvores do conjunto gerado representa a classificação de uma determinada transação.

[Taha and Malebary 2020] apresentaram em sua revisão da literatura que as técnicas de aprendizado de máquina na detecção de fraudes em cartões de crédito mais populares são as supervisionadas e que as transações devem ser divididas entre, no mínimo, duas classes já previamente rotuladas. O aprendizado supervisionado de máquina necessita que o conjunto de dados seja organizado em um conjunto de treino e em conjunto de teste.

Após a assimilação das características das transações legítimas e das transações fraudulentas no conjunto de testes, será possível a técnica classificar, em níveis satisfatórios, transações nunca vistas do conjunto de testes em classe negativa ou legítima e classe positiva ou fraude. Entretanto, para 75% dos autores, a aplicação de qualquer uma das técnicas sem um tratamento prévio nos dados não será capaz de classificar corretamente uma transação da classe positiva já que o volume de transações dela é muito pequeno, ou seja, o desbalanceamento entre a classe positiva e classe negativa é considerado o maior problema na detecção de fraudes de cartões de crédito.

Para [Dornadula and Geetha 2019], [Can et al. 2020], [Misra et al. 2020], [de C. B. Paul et al. 2020], [Bagga et al. 2020], [Rtayli and Enneya 2020], [Taha and Malebary 2020], [RB and KR 2021] e [Tingfei et al. 2020], o *Synthetic Minority Oversampling Technique* - SMOTE e a redução de dimensionalidade de atributos as técnicas mais usadas no pré-processamento de dados. Segundo [Tingfei et al. 2020], o SMOTE usa o KNN para gerar dados sintéticos a partir dos dados da classe minoritária assimilando seu padrão. Segundo [Rtayli and Enneya 2020], quanto menor o número de atributos não redundantes e que sejam relevantes, melhor será o desempenho do sistema de classificação de fraudes.

Por fim, a última etapa de um processo de aprendizado de máquina é mensuração dos resultados obtido por cada uma das técnica utilizadas, de modo que seja possível comparar o desempenho entre elas e identificar suas vantagens e desvantagens. Quase 70% dos trabalhados recuperados utilizaram as seguintes métricas de avaliação: acurácia, sensibilidade ou revocação, precisão e *F-score*.

Para [de C. B. Paul et al. 2020], a precisão é a "razão das transações fraudulentas em relação ao total de transações que o modelo previu como fraudulentas"; a revocação "é a proporção de casos previstos como positivos do total de casos positivos"; e, o *F-score* é "calculado como média harmônica da precisão e revocação obtidas". Para [Bagga et al. 2020], a acurácia é a razão entre número das predições corretas dividido pelo total de predições.

2.2. Trabalhos Correlatos

Segundo [de C. B. Paul et al. 2020] compararam o *KNN*, *Random Forest* e *Gradient Boosting* após aplicação de técnica de pré-processamento t-SNE e obtiveram que a maior

precisão foi no KNN, porém com baixa revocação e o *Random Forest* apresentou o maior *F-score*.

Já [Bagga et al. 2020] aplicam *oversampling* ADASYN e propõem a utilização do *Pipelining* e *Esemble Learning* para obter melhores resultados na classificação de fraudes comparados a KNN, *Logistic Regression*, *Random Forest*, *Naive Bayes*, *Multilayer Perceptron*, *Adaboost* e *Quadrant Discriminative Analysis*. Os métodos *Pipelining* e *Esemble Learning* obtiveram melhores resultados quando comparados as demais técnicas de aprendizado de máquina, sendo a acurácia a métrica de maior destaque.

Por fim, [RB and KR 2021] comparam o uso de KNN, *Artificial Neural Network* (ANN) e *Support Vector Machine* (SVM). Aplica-se normalização e redutor de escala nos dados após um *undersampling* na classe negativa. Os resultados demonstraram que o método ANN e KNN obtiveram 99% de acurácia, porém, ao custo de perda de sensibilidade e precisão. O SVM é o classificador que se destaca no experimento por seu desempenho melhor em precisão e sensibilidade, sendo a acurácia menor que o demais, mas ainda em nível aceitável de 93%.

3. Material e Métodos

Esta pesquisa é aplicada e utiliza a mesma estrutura dos trabalhos correlatos, organizada em quatro fases, a saber:

- análise exploratória dos dados;
- aplicação de técnicas de pré-processamento, de modo a identificar os atributos mais relevantes da base de dados escolhida e limpeza de registros problemáticos;
- construção do modelo com base nas técnicas selecionadas;
- e, avaliação e comparação de resultados.

3.1. Características da base de dados

A base de dados foi disponibilizada pela empresa Vesta para uma competição do *Kaggle*[2] em Setembro de 2019 e utiliza dados reais de *e-commerce* de transações com cartões de crédito.

Esta base possui dados das transações financeiras e de identificação do usuário que, somados, perfazem o total de quatrocentos e trinta e cinco atributos diferentes, segregados em treino e testes, sendo as principais informações resumidas a seguir:

- transações financeiras treino: total de transações de 590.541 sendo os principais atributos *TransactionId*, *isFraud*, *TransactionDT*, *TransactionAmt* e *ProductCD* ;
- transações metadados treino: total de transações de 144.234, sendo os principais atributos *TransactionId*, *DeviceType* e *DeviceInfo*;
- transações financeiras testes: total de transações de 506.692.
- transações metadados testes: total de transações de 141.908.

[2] `https://www.kaggle.com/c/ieee-fraud-detection/data`

4. Cronograma

Entregáveis	Setembro	Outubro	Novembro	Dezembro	Janeiro
Correção do PTCC					
Análise exploratória					
Pré-processamento					
Construção do modelo					
Análise das métricas					
Redação texto final					
Defesa					

Tabela 3. Cronograma de execução de atividades por Mês.

Referências

Bagga, S., Goyal, A., Gupta, N., and Goyal, A. (2020). Credit card fraud detection using pipeling and ensemble learning. *International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020*, 173:104–112.

Can, B., Yavuz, A. G., Karsligil, E. M., and Guvensan, M. A. (2020). A closer look into the characteristics of fraudulent card transactions. *IEEE Access*, 8:166095–166109.

de C. B. Paul, A., Alves, M. A. Z., and de Oliveira, L. E. S. (2020). Avaliação de métodos de machine learning na detecção de fraude em dados transacionais de cartão de crédito.

Dornadula, V. N. and Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019*, 165:631–641.

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M., and Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*, 7:93010–93022.

Misra, S., Thakur, S., Ghosh, M., and Saha, S. K. (2020). An autoencoder based model for detecting fraudulent credit card transaction. *International Conference on Computational Intelligence and Data Science*, 167:254–262.

Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., and Nandi, A. K. (2018). Credit card fraud detection using adaboost and majority voting. *IEEE Access*, 6:14277–14284.

RB, A. and KR, S. K. (2021). Credit card fraud detection using artificial neural network. *1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020)*, 2(1):35–41.

Rtayli, N. and Enneya, N. (2020). Selection features and support vector machine for credit card risk identification. *13th International Conference Interdisciplinarity in Engineering, INTER-ENG 2019, 3–4 October 2019, Targu Mures, Romania*, 46:941–948.

Taha, A. A. and Malebary, S. J. (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, 8:25579–25587.

Tingfei, H., Guangquan, C., and Kuihua, H. (2020). Using variational auto encoding in credit card fraud detection. *IEEE Access*, 8:149841–149853.