Predição de Fraudes em Operações Financeiras

Aluno: Gustavo Rodrigues Carvalho RA: 10175838

3. Resumo

Com o avanço da tecnologia em diferentes áreas, uma delas sendo a financeira, gerando assim um maior número de usuários que fazem uso dos meios digitais para realizar suas transações, junto com o aumento de clientes legítimos que usam as ferramentas de forma legal e ordeira, há também aqueles que se utilizam da facilidade e conveniência da facilitação que o meio digital trouxe para as operações financeiras e buscam cometer fraudes, há também um aumento no número de fraudes em operações financeiras. Tendo isso em mente, este trabalho tem como objetivo identificar as características de uma operação fraudulenta, desenvolver modelos de aprendizado de máquina e descobrir qual modelo tem maior precisão ao prever uma operação financeira fraudulenta.

4. Introdução

Contextualização

Segundo o jornal E-Investidor (2023), houve mais de 2,8 mil tentativas de fraudes financeiras por minuto em canais eletrônicos financeiros. No mesmo período, aproximadamente 365 milhões de tentativas de golpes foram registradas, indicando que cerca de 1,73% das transações digitais no sistema financeiro do país tiveram intenções criminosas.

Essa crescente incidência de fraudes pode ser explicada pela maior vulnerabilidade do consumidor no ambiente online. Conforme apontado por Sampaio (2023), *"A facilitação do acesso à internet evidenciou a situação da vulnerabilidade agravada do consumidor dentro do ambiente online. Nessa conjuntura, criminosos deram nova roupagem a crimes já conhecidos e transferiram a sua prática também para o ambiente digital"*.

Justificativa

A mudança do ambiente físico para o digital fez com que a forma como lidamos com fraudes financeiras evoluísse. Antes, a detecção de fraudes dependia predominantemente da capacidade humana. Hoje, devido ao volume massivo de dados e à velocidade com que são gerados, há a necessidade do envolvimento de tecnologia avançada. Lokanan (2024) destaca que essa transição tornou a detecção de fraudes um desafio que exige o uso de inteligência artificial e aprendizado de máquina.

Dado o crescimento contínuo dessas atividades criminosas, torna-se essencial aprimorar métodos automatizados para detectar operações fraudulentas com maior precisão, reduzindo prejuízos financeiros e impactos negativos sobre consumidores e instituições.

Objetivo

O presente estudo tem como objetivo utilizar algoritmos de aprendizado de máquina para classificar operações financeiras fraudulentas. Os resultados obtidos serão comparados com aqueles de estudos anteriores sobre detecção de fraudes financeiras, nos quais foram disponibilizados tanto os modelos aplicados quanto as bases de dados analisadas. Além disso, será realizada uma análise comparativa com indicadores financeiros.

Opção do Projeto

Este projeto se enquadra na **Opção Framework**, empregando técnicas de **Machine Learning** para resolver um problema de **classificação**. O estudo se baseará na aplicação de algoritmos de aprendizado de máquina para prever fraudes em transações financeiras, utilizando um conjunto de dados apropriado para o treinamento e validação dos modelos.

5. Descrição do Problema

Antes de prosseguirmos é necessário introduzir alguns conceitos, dentre eles um interessante de se tomar conhecimento é o de fraude financeira. Segundo Wells(2014) “No sentido mais lato, a fraude pode englobar qualquer crime com fins lucrativos que utilize o engano como principal modus operandi. Das três formas de retirar ilegalmente dinheiro a uma vítima - força, artifício ou furto - todos os crimes que recorrem a artifícios são fraudes” (p. 8).

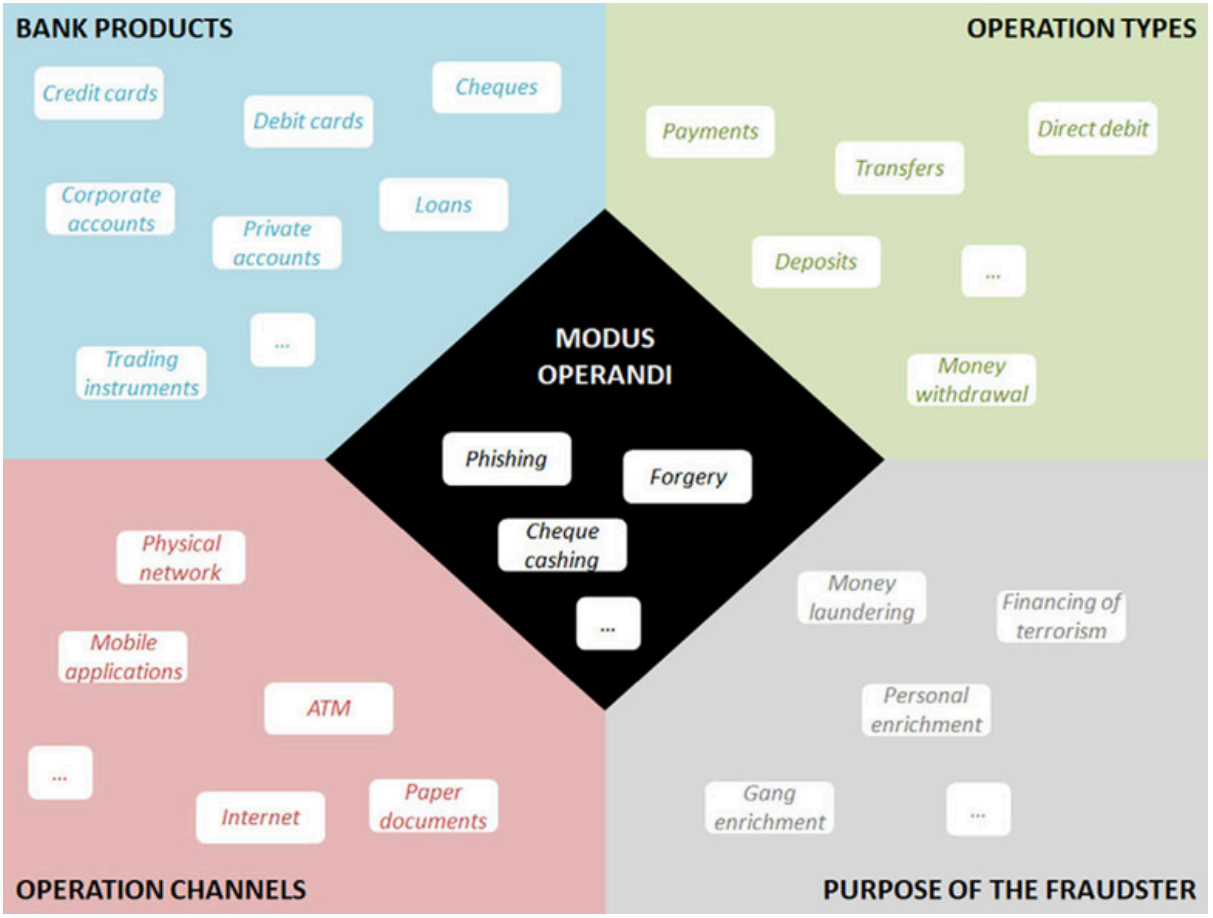

Figura 1. Illustration of fraud dimensions.

Somado ao conceito de fraudes financeiras temos o conceito de operações financeiras segundo (Praxis 2024) as operações financeiras são transações do mercado financeiro, que envolvem a compra e venda de ações e títulos em moedas estrangeiras, transações essas que ocorrem entre investidores, empresas e instituições financeiras (ver Figura 1).

6. Discutir a respeito dos aspectos Éticos do uso da IA e a sua Responsabilidade no desenvolvimento da solução;

Um dos dilemas éticos, que podem ser enfrentados é com relação ao viés da análise, por conta do pequeno número, de casos de fraudes financeiras em relação ao grande número de transações não fraudulentas, o que pode interferir nos resultado das predições, gerando falsos positivos e falsos negativos.

Outro ponto que deve ser levado em consideração, é a sensibilidade dos dados financeiros de pessoas, para isso o dataset selecionado

7. Dataset

O conjunto de dados utilizado neste projeto é uma representação sintética de transações financeiras móveis, gerada pelo **PaySim**, um simulador que utiliza dados agregados de registros financeiros reais de um serviço de dinheiro móvel em um país africano. Esse dataset foi desenvolvido para fins de pesquisa e tem como objetivo permitir estudos de detecção de fraudes em transações eletrônicas.

Origem e Estrutura do Dataset

O PaySim simula transações financeiras móveis com base em um mês de registros financeiros reais, disponibilizados por uma empresa multinacional que opera esse serviço em mais de 14 países. Esse dataset foi reduzido para **1/4 do tamanho original** para facilitar seu uso e análise na plataforma Kaggle.

8. Metodologia.

Coleta e Preparação dos Dados

O dataset utilizado é o **PaySim**, que simula transações financeiras móveis e inclui atividades fraudulentas. Antes de aplicar algoritmos de aprendizado de máquina, realizaremos:

- **Análise Exploratória de Dados (EDA)**
- **Tratamento de dados desbalanceados**,
- **Remoção de colunas inviáveis**
- **Transformação e normalização de dados**

Seleção e Treinamento do Modelo

Serão testados diferentes algoritmos de *Machine Learning* para verificar qual melhor se adapta ao problema de detecção de fraudes:

- **Modelos baseados em árvores de decisão:**
  - *Random Forest*
  - *XGBoost*
  - Isolation Forest (IF)
- **Modelos estatísticos tradicionais:**
  - *Regressão Logística*
- **Modelos baseados em aprendizado não supervisionado:**
  - *K-Means* e *DBSCAN* para identificação de anomalias

Os modelos serão avaliados com **validação cruzada** e métricas adequadas ao problema de classificação desbalanceada, como **AUC-ROC, F1-score, precisão e recall**.

Avaliação de Performance e Ajustes

Com base nos resultados iniciais, ajustaremos os hiperparâmetros dos modelos para otimizar seu desempenho, e utilizaremos métricas financeiras para detectar fraudes como f1-score.

Implementação e Resultados Esperados

O modelo final será implementado e avaliado em um ambiente de teste para verificar sua capacidade de identificar fraudes em novas transações. Os resultados esperados incluem, uma baixa taca de detecção de fraudes, com um baixa taxa de falsos negativos, e positivos e reduzindo o tempo de resposta para possivelmente ser usado em tempo real.

9. Referências citadas no texto.

SEJA PRAXIS. O que é: Operações Financeiras. Disponível em: https://www.sejapraxis.com.br/glossario/o-que-e-operacoes-financeiras/. Acesso em: 23 set. 2024.

E-INvestidor. Brasil sofre 2,8 mil tentativas de fraudes financeiras por minuto; saiba como se proteger. *Estadão*, 27 jun. 2023. Disponível em: https://einvestidor.estadao.com.br/ultimas/brasil-dados-tentativas-fraude-dicas-se-proteger/. Acesso em: 23 set. 2024.

SAMPAIO, Marília de Ávila e Silva. Responsabilidade civil das instituições financeiras nas fraudes eletrônicas. *Tribunal de Justiça do Distrito Federal e dos Territórios*. Disponível em: https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2023/responsabilidade-civil-das-instituicoes-financeiras-nas-fraudes-eletronicas

LOKANAN, M. E. Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks. ***Journal of Applied Security Research***, v. 19, n. 1, p. 20-44, 2024. DOI: 10.1080/19361610.2022.2114744. Disponível em: https://doi.org/10.1080/19361610.2022.2114744. Acesso em: 12 Set 2024.

NESVIJEVSKAIA, Anna; OUILLADE, Sophie; GUILMIN, Pauline; ZUCKER, Jean-Daniel. *Illustration of fraud dimensions*. 2024. Ilustração. Disponível em: https://www.researchgate.net/figure/Illustration-of-fraud-dimensions_fig1_353008528 . Acesso em: 29 set. 2024.

10. Referências bibliográficas

GARETH, J.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. ***An Introduction to Statistical Learning with Applications in R***. 2. ed. Nova York: Springer, 2013. Capítulos 4 e 5.

SILVA, A. S.; PERES, S. M.; BOSCARIOLI, C. ***Introdução a Mineração de Dados – Com Aplicações em R.*** 1. ed. São Paulo: Elsevier, 2016. Capítulos 1, 2 e 3.

HAN, J.; KAMBER, M.; PEI, J. ***Data Mining: Concepts and Techniques***. 3. ed. San Francisco: Morgan Kaufmann, 2012. Capítulos 6, 8, 9, 12 e 13.

WITTEN, I. H.; FRANK, E. ***Data Mining: Practical Machine Learning Tools and Techniques.*** 2. ed. San Francisco: Morgan Kaufmann, 2005. Capítulos 5, 6 e 8.

WELLS, Joseph T. ***Principles of Fraud Examination***. 4. ed. Hoboken: John Wiley & Sons, 2014. Capítulo 1.

LOKANAN, M. E. Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks. ***Journal of Applied Security Research***, v. 19, n. 1, p. 20-44, 2024. DOI: 10.1080/19361610.2022.2114744. Disponível em: https://doi.org/10.1080/19361610.2022.2114744. Acesso em: 12 Set 2024.

HEWAPATHIRANA, I. Utilizing Prediction Intervals for Unsupervised Detection of Fraudulent Transactions: A Case Study. ***Asian Journal of Engineering and Applied Technology***, v. 11, n. 2, p. 1-10, 2022. DOI: https://doi.org/10.51983/ajeat-2022.11.2.3348. Disponível em: www.trp.org.in. Acesso em: 11 Set 2024.

XU, T.; LIU, J.; CHENG, H. Predicting Fraud in U.S. - Listed Chinese Companies: An Empirical Analysis Based on M-Score and F-Score Models. In: ***Proceedings of the 2023 International Conference on Management Research and Economic Development***. DOI: 10.54254/2754-1169/20/20230177. Acesso em: 12 Set. 2024.

DALAL, S.; SETH, B.; RADULESCU, M.; SECARA, C.; TOLEA, C. Predicting Fraud in Financial Payment Services through Optimized Hyper-Parameter-Tuned XGBoost Model. ***Mathematics***, v. 10, n. 24, p. 4679, 2022. DOI: https://doi.org/10.3390/math10244679. Acesso em: 12 Set. 2024.

GAO, J. X.; ZHOU, Z. R.; AI, J. S.; XIA, B. X.; COGGESHALL, S. Predicting Credit Card Transaction Fraud Using Machine Learning Algorithms**. *Journal of Intelligent Learning Systems and Applications*,** v. 11, p. 33-63, 2019. DOI: https://doi.org/10.4236/jilsa.2019.113003. Recebido em: 6 abr. 2019. Aceito em: 11 ago. 2019. Publicado em: 14 ago. 2019. Disponível em: http://www.scirp.org/journal/jilsa. Acesso em: 16 Set 2024.

NESVIJEVSKAIA, A.; OUILLADE, S.; GUILMIN, P.; ZUCKER, J.-D. The accuracy versus interpretability trade-off in fraud detection model. *Data & Policy*, v. 3, e12, 2021. DOI: 10.1017/dap.2021.3. Disponível em: https://doi.org/10.1017/dap.2021.3. Acesso em: 16 Set. 2024.

SAMPAIO, Marília de Ávila e Silva. Responsabilidade civil das instituições financeiras nas fraudes eletrônicas. *Tribunal de Justiça do Distrito Federal e dos Territórios*. Disponível em: https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2023/responsabilidade-civil-das-instituicoes-financeiras-nas-fraudes-eletronicas. Acesso em: 23 set. 2024.

E-INvestidor. Brasil sofre 2,8 mil tentativas de fraudes financeiras por minuto; saiba como se proteger. *Estadão*, 27 jun. 2023. Disponível em: https://einvestidor.estadao.com.br/ultimas/brasil-dados-tentativas-fraude-dicas-se-proteger/. Acesso em: 23 set. 2024.

SEJA PRAXIS. O que é: Operações Financeiras. Disponível em: https://www.sejapraxis.com.br/glossario/o-que-e-operacoes-financeiras/. Acesso em: 23 set. 2024.

NEHA, S. V. S. T.; YADAV, Yogesh; GOYAL, Yashika. *Introduction to Machine Learning*. **International Journal of Advanced Research in Science, Communication and Technology** (IJARSCT), v. 4, n. 2, p. 100-108, mar. 2024. DOI: 10.48175/IJARSCT-15723. Disponível em: www.ijarsct.co.in. Acesso em: 28 set. 2024.

NESVIJEVSKAIA, Anna; OUILLADE, Sophie; GUILMIN, Pauline; ZUCKER, Jean-Daniel. *Illustration of fraud dimensions*. 2024. Ilustração. Disponível em: https://www.researchgate.net/figure/Illustration-of-fraud-dimensions_fig1_353008528. Acesso em: 29 set. 2024.

WANG, Le; HAN, Meng; LI, Xiaojuan; ZHANG, Ni; CHENG, Haodong. *Review of classification methods on unbalanced data sets*. IEEE Access, v. 9, p. 48152-48171, 2021. DOI: 10.1109/ACCESS.2021.3074243.