

Modelos Supervisionados e Não Supervisionados para Detecção de Fraudes

Alunos: Felipe Silva Alves de Oliveira e Gabriella Alves de Oliveira

Introdução

O crescimento exponencial das transações digitais, impulsionado pelo comércio eletrônico, pelos serviços de internet banking e pelo aumento do uso de cartões de crédito e débito, trouxe consigo um desafio crítico: a intensificação de práticas fraudulentas. Estima-se que fraudes financeiras causem prejuízos de bilhões de dólares anualmente em escala global, afetando tanto instituições bancárias quanto consumidores finais. Esse cenário torna a detecção eficiente de fraudes uma prioridade estratégica para garantir a segurança das operações e reduzir perdas financeiras.

Um dos principais obstáculos para a construção de sistemas automáticos de detecção de fraudes é o **desbalanceamento dos dados**. Em geral, registros de fraude representam uma fração muito pequena do total de transações, o que dificulta o treinamento de modelos capazes de identificar padrões anômalos sem incorrer em alto número de falsos positivos.

Diversas abordagens de aprendizado de máquina têm sido exploradas para mitigar esse desafio. Métodos supervisionados, como árvores de decisão e máquinas de vetores de suporte (SVM), apresentam bom desempenho quando há rótulos disponíveis. Por outro lado, métodos não supervisionados, como o Isolation Forest e autoencoders, são mais adequados em cenários em que as instâncias de fraude são raras ou pouco representativas, permitindo a identificação de anomalias sem necessidade de rótulos explícitos.

Muitos trabalhos relacionados já exploraram esse campo. Pesquisas anteriores indicam que modelos supervisionados tendem a alcançar maior precisão quando a base de dados é balanceada artificialmente (por técnicas como oversampling ou SMOTE), enquanto modelos não supervisionados se destacam pela capacidade de generalização em cenários altamente desbalanceados. Nesse contexto, a combinação ou a comparação sistemática dessas abordagens se mostra um caminho relevante tanto para a academia quanto para aplicações práticas no setor financeiro.

Assim, este projeto se insere na interseção entre teoria e prática do aprendizado de máquina, buscando avaliar criticamente as vantagens e limitações de diferentes métodos para detecção de fraudes financeiras.

Objetivo/Proposta

O objetivo principal deste projeto é **comparar o desempenho de modelos supervisionados e não supervisionados aplicados à detecção de fraudes em transações financeiras**.

De forma mais específica, propomos:

1. **Treinar e avaliar modelos supervisionados** (como Árvores de Decisão e SVM) em conjuntos de dados rotulados, com técnicas de balanceamento para mitigar o impacto da desproporção entre classes.
2. **Aplicar e analisar modelos não supervisionados** (como Isolation Forest e Autoencoders) para detecção de anomalias, considerando cenários em que as fraudes representam uma minoria extrema dos dados.
3. **Comparar as métricas de desempenho** (precisão, recall, F1-score, AUC-ROC, entre outras) entre os dois grupos de modelos, avaliando sua aplicabilidade prática em contextos reais.
4. **Discutir as limitações e potenciais combinações** entre métodos supervisionados e não supervisionados, apontando caminhos para soluções híbridas.

A proposta é contribuir para o entendimento de como diferentes paradigmas de aprendizado de máquina podem ser aplicados a um dos maiores desafios do setor financeiro atual, oferecendo subsídios tanto para aplicações acadêmicas quanto para sistemas reais de prevenção a fraudes.

O projeto contemplará ao menos um tema não abordado em sala de aula (Autoencoders), explorando sua aplicação prática como técnica não supervisionada para detecção de anomalias.

Referências:

- <https://arxiv.org/pdf/1908.11553>
- https://www.researchgate.net/profile/Mohammed-Al-Shabi/publication/335526336_Credit_Card_Fraud_Detection_Using_Autoencoder_Model_in_Unbalanced_Datasets/links/5f2a7901a6fdccc43ac10d4/Credit-Card-Fraud-Detection-Using-Autoencoder-Model-in-Unbalanced-Datasets.pdf
- <https://bright-journal.org/Journal/index.php/JADS/article/view/16>