

Ciência da Computação / Sistemas de Informação

## **Comparação de Métodos de Classificação de Machine Learning na Identificação de Fraudes Financeiras**

Bryan de Lima Naneti Barbosa - 6º módulo de Ciência da Computação, UFLA, bolsista PIBIC/UFLA. Contato: bryan.barbosa@estudante.ufla.br

Renato Ribeiro de Lima - Professor do Departamento de Estatística, UFLA, Orientador. Contato: rrlima@ufla.br - Orientador(a)

Matheus Saraiva Alcino - Cientista de Dados, Secretaria de Estado de Saúde de Goiás. Contato: osaraivamatheus@gmail.com

### **Resumo**

A identificação de fraudes financeiras tornou-se um desafio significativo devido ao avanço tecnológico, que possibilita novas formas de transações financeiras e, conseqüentemente, novas estratégias para agentes fraudulentos. Este estudo visa explorar e comparar cinco modelos de machine learning (ML) na detecção de fraudes financeiras: Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Random Forest, Regressão Logística e Redes Neurais Artificiais. Foram utilizados dois conjuntos de dados distintos: um totalmente balanceado e outro altamente desbalanceado. A comparação dos modelos foi realizada utilizando várias métricas de desempenho, incluindo matriz de confusão, acurácia, área sob a curva ROC, precision, recall e f1-score. Nos dados balanceados, os modelos Random Forest e XGB apresentaram o melhor desempenho, destacando-se pela sua precisão e capacidade de detecção de fraudes. Em seguida, foi conduzida uma análise comparativa utilizando a base de dados desbalanceada, aplicando os mesmos cinco modelos e métricas. Os resultados indicaram que, mesmo em cenários altamente desbalanceados, Random Forest e XGB mantiveram um desempenho superior, evidenciando sua robustez e eficácia na detecção de fraudes financeiras. Este estudo sugere que a aplicação dos modelos Random Forest e XGB em dados reais pode proporcionar uma ferramenta eficiente para o rastreamento de operações financeiras ilegais e a identificação de fraudadores, independentemente do balanceamento dos dados. Assim, a implementação de tais modelos pode contribuir significativamente para a segurança e integridade das transações financeiras no mercado atual. A adoção de técnicas avançadas de ML para a identificação de fraudes financeiras é essencial para acompanhar as crescentes complexidades das transações e das táticas fraudulentas, assegurando um ambiente financeiro mais seguro e confiável.

Palavras-Chave: Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Métricas de Desempenho.

Instituição de Fomento: CNPq

Link do pitch: <https://youtu.be/UXZUL4fyU2w>