UNIVERSIDADE ESTÁCIO DE SÁ

Relatório Técnico

Sistema de Detecção de Fraudes em Transações Financeiras utilizando Python e Machine Learning

Eduardo Ferreira Rocha
Gustavo Bispo da Silva
Henrique Gomes da Silva

Carapicuíba — SP
2025

Eduardo Ferreira Rocha
Gustavo Bispo da Silva
Henrique Gomes da Silva

Relatório desenvolvido como requisito parcial para avaliação acadêmica, no curso da Universidade Estácio de Sá. O trabalho apresenta o desenvolvimento de um sistema computacional para detecção de fraudes utilizando técnicas de aprendizado de máquina.

Carapicuíba — SP
2025

SUMÁRIO

1. Introdução

A detecção de fraudes é uma área essencial no contexto tecnológico atual, especialmente em sistemas financeiros que envolvem grande volume de transações digitais. Este trabalho apresenta o desenvolvimento completo de um sistema capaz de estimar a probabilidade de fraude utilizando técnicas de machine learning e uma base de dados sintética construída especificamente para fins acadêmicos.

2. Fundamentação Teórica

Modelos de classificação binária são amplamente utilizados em sistemas antifraude. Dentre as técnicas empregadas, destacam-se regressão logística, árvores de decisão, métodos ensemble e aprendizagem online. O modelo utilizado neste projeto baseia-se em técnicas lineares otimizadas com gradiente descendente estocástico.

3. Geração do Dataset

O dataset foi gerado artificialmente com 50.000 transações, incluindo identificadores, valores, comerciantes, países, timestamps e indicação de fraude. O objetivo foi simular um ambiente próximo ao encontrado em empresas reais.

4. Tratamento dos Dados

Foram aplicadas técnicas de limpeza, padronização e conversão de variáveis categóricas. A ferramenta FeatureHasher foi utilizada para transformar textos em vetores numéricos.

5. Engenharia de Atributos

Foram selecionados atributos simples, mas eficazes: valor da transação, comerciante e país. Em sistemas reais, atributos adicionais como padrões temporais e comportamento do usuário poderiam ser aplicados.

6. Modelo de Machine Learning

O SGDClassifier com função log_loss foi utilizado, resultando em um modelo rápido e capaz de retornar probabilidades. O StandardScaler foi empregado para melhorar a convergência e a estabilidade numérica.

7. Avaliação do Modelo

A divisão treino/teste (80/20) permitiu análise adequada das métricas de desempenho. Apesar do desbalanceamento (3% fraudes), o modelo apresentou comportamento

satisfatório dentro da proposta educacional.

8. Funcionamento do Sistema

O sistema final possui interface interativa em que o usuário informa os dados da transação e recebe o cálculo da probabilidade de fraude. A resposta é categorizada entre risco baixo, médio ou alto.

9. Limitações

O conjunto de dados é sintético, o que reduz a aderência a padrões reais. Modelos lineares também possuem limitação ao capturar relações não lineares complexas.

10. Expansões Futuras

Sugere-se a adoção de modelos mais robustos como XGBoost ou redes neurais, integração com APIs, dashboard de monitoramento e análise comportamental mais avançada.

11. Conclusão

O trabalho demonstra todas as etapas necessárias para a construção de um sistema antifraude, desde a concepção da base de dados até a implementação final do modelo preditivo.

12. Referências

Documentações oficiais do scikit-learn, pandas, NumPy, além de artigos acadêmicos sobre detecção de anomalias e classificação binária.