RELATÓRIO DE ATIVIDADES ORIENTADAS DE ENSINO

NICOLAS DE OLIVEIRA LOPES BRAGA

TECNOLOGIAS PARA DETECCAO DE FRAUDES BANCÁRIAS

RESUMO: Levantamento de técnicas e tecnologias de IA relacionadas a detecção de fraudes bancárias; Levantamento de datasets para utilização das técnicas elencadas.

Para alcançar os objetivos este estudo explorou a abordagem das fraudes bancárias na literatura, onde em seguida, foram pesquisadas bases de dados que refletissem esse cenário. Uma biblioteca foi empregada no desenvolvimento de vários modelos simultâneos, usando configurações específicas. Para validar os resultados, foram realizados testes, e validações cruzadas foram aplicadas para garantir a robustez das técnicas utilizadas.

O ponto inicial do processo é a seleção de um conjunto de dados relevante e significativo. No entanto, dada a natureza sensível dessas informações, optamos por adotar algumas medidas nos datasets para preserver a privacidade dos dados. A técnica utilizada, muito conhecida, Análise de Componentes Principais (PCA) nos permitiu preservar a representatividade estatística e remover as informações originais. Os dados utilizados na pesquisa podem ser encontrados no Kaggle, estas transações foram feitas por cartão de crédito (disponível em: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud).

Alguns outros processos foram encontrados durante a pesquisa, como alguns sistemas capazes de gerar dados fictícios para compras de cartão de crédito seguindo perfis determinados manualmente. Como estávamos buscando o mais próximo da realidade para analisar os algoritmos optamos pelo dataset que continha dados verdadeiros de transações.

Tratamentos padrões foram realizados na base, como: tratamento de valores ausentes, remoção de duplicatas, tratamento de outliers e balanceamento adequado do conjunto de dados. Sabemos o quão importante esses tratamentos são e que podem afetar diretamente no resultado das nossas análises.

A técnica Smote foi utilizada para o balanceamento da base, esta técnica visa evitar o viés em direção a classe majoritária, melhorando a generalização do modelo. O Smote busca solucionar esse problema gerando exemplos sintéticos da classe minoritária para equilibrar a distribuição de classes.

Durante a pesquisa do estado da arte, notamos inúmeros algoritmos sendo utilizados para análises semelhantes. Neste caso utilizamos uma biblioteca chamada Pycaret, que nos uma implementação eficiente de uma variedade de algoritmos de Machine Learning. Nos fornecendo flexibilidade para explorar modelos como Logistic Regression, Ridge Classifier, Naive Bayes, K Neighbors, Decision Tree e Extreme Gradiente Boosting.

Para garantir a escolha correta dos algoritmos utilizamos algumas métricas de desempenho. Matrizes de confusão, precisão, recall, F1-score, área sob a curva ROC e outras métricas nos forneceram uma visão abrangente sobre como os modelos se comportam em relação a detecção de fraudes.

Em resumo, a abordagem proposta, que combina a seleção cuidadosa do conjunto de dados, um pré-processamento robusto, a utilização eficiente da biblioteca PyCaret e uma análise detalhada dos resultados, visa desenvolver um modelo de detecção de fraudes sólido e eficaz para transações online de cartão de crédito.